International Journal of Technology 11(2) 388-399 (2020)
Received October 2019 / Revised January 2020 / Accepted April 2020

International Journal of Technology

http://ijtech.eng.ui.ac.id

Data-driven Fault Diagnosis of Power Transformers using Dissolved Gas Analysis (DGA)

Arian Dhini[1*], Akhmad Faqih[2], Benyamin Kusumoputro[2], Isti Surjandari[1], Andrew Kusiak[3]

[1]*Laboratory Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia*
[2]*Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia*
[3]*Department of Industrial and Systems Engineering, College of Engineering, University of Iowa, 4627 Seamans Center for the Engineering Arts and Sciences, Iowa City -Iowa, 52242, USA*

Abstract. A power transformer is a critical piece of equipment in a power plant for distributing electricity, and it experiences thermal and electrical stresses during operation. Dissolved gas analysis (DGA) remains one of the most effective techniques to monitor the health of oil-filled transformers. Some traditional approaches for interpreting DGAs have been introduced. Occasionally, such approaches leave the state of the transformer uncategorized. This study proposed data-driven approaches for a fault diagnosis system based on DGA data using support vector machine (SVM). SVM is known for its robustness, good generalization capability, and unique global optimum solutions, particularly when data is limited. Backpropagation neural networks (BPNN) and extreme learning machine-radial basis function (ELM-RBF), a recent Neural Networks (NN)-based method with extremely fast computation time, were compared to SVM. An advanced technique to overcome the imbalanced data and synthetic minority oversampling technique (SMOTE) was proposed to investigate the effect on classifier performance. The model was trained and tested using IEC TC 10 databases and transformer DGA monitoring data of a thermal power plant in Jakarta. The results indicated that SVM displayed the best performance compared to ELM-RBF and BPNN. It demonstrated extremely high accuracy, while still maintaining fast computation time for all stages in the proposed multistage fault diagnosis system.

Keywords: Condition monitoring; Dissolved gas analysis; Fault diagnosis; Support vector machine; Transformer

1. Introduction

Thermal power plants remain main electricity providers; however, most of them are highly aged. According to Ulum et al. (2017), plant equipment, which degrades over time, leads to electricity production loss. Therefore, the industrial process measurements are imperative to monitor and assure the quality and safety in operations (Kusiak and Song, 2009). Those measurements are the bases to determine the maintenance activities. Such maintenance will raise the reliability and availability of equipment (Pariaman et al., 2017), including the reliability and availability of power transformers, one of the key pieces of equipment in a power distributing system. Despite the fact that faults in a transformer

*Corresponding author's email: arian@ie.ui.ac.id, Tel.: +62-21-78888805, Fax: +62-21-7885656
doi: 10.14716/ijtech.v11i2.3625

occur infrequently, once they do occur, their impact is significant in terms of safety, downtime, and equipment loss (Hernandez and Labib, 2017). DGA has been known to be a popular, sensitive, and reliable technique to monitor the insulation condition of transformers (Saha, 2003; Chakravorti et al., 2013). However, conventional DGA interpretation methods include certain drawbacks in terms of accuracy and uncertainty (Shintemirov et al., 2009; Ghoneim and Taha, 2016). These approaches require the manual interpretation of experts, and some measurements may be unidentifiable when using any interpretation method (Lin et al., 1993; Yang et al., 2009; Abu-Siada and Islam, 2012).

Faults should be diagnosed accurately and in a timely manner, since ignoring them degrades the safety and security of the process, and it could lead to catastrophic failures and loss of material and even life (Gao et al., 2015). Therefore, machine learning methods have been recently applied in studies for transformer fault detection and diagnosis (FDD) using DGA data. Support vector machine (SVM) has become an increasingly popular technique in machine learning. SVM is known for its robustness, good generalization capability, and unique global optimum solution. Shin and Cho (2006) explained that its good generalization capability was due to structural risk minimization, which is employed by SVM, rather than empirical risk minimization, as in NN. SVM also performs well with small samples and high dimension data, while still maintaining short computational times (Lv et al., 2005; Bacha et al., 2012; Sahri and Yusof, 2015; Souza and Ramachandran, 2016). Some studies have compared SVM with other methods, such as kNN, ANN (Shintemirov et al., 2009), the expert system (Lv et al., 2005), fuzzy logic, multilayer perceptron (MLP), and radial basis function neural networks (RBFNN) (Fei et al., 2009; Bacha et al. 2012). These studies found that SVM performed better. Ghoneim and Taha (2016) conducted a fault diagnosis study in the same area, but they only applied limit rules to categorize faults.

In reality, the availability of fault data is usually significantly lower than the normal operating condition (NOC). This situation may lead to a decrease in classifier performance. Fault diagnosis focuses on how to detect a fault, the minority class. Sahri and Yusof (2015) tested some scenarios of input features without considering how to handle imbalanced data. Only a few studies have considered this imbalanced data in FDD. Chawla et al. (2002) proposed the use of the synthetic minority oversampling technique (SMOTE) to overcome the imbalanced data problem by creating new data through an interpolation process. No previous studies have applied SMOTE in the case of transformers for fault diagnosis.

This study aims to develop an accurate and fast transformer fault diagnosis system from DGA data using SVM and SMOTE for data balancing. Other recent classifier methods, extreme learning machine-radial basis function (ELM-RBF), which is known for its extremely fast computation time, is applied. The high computation of ELM-RBF, while still maintains good accuracy, will result in a number advantages, such as fast fault diagnosis and lower hardware investment costs. In addition, the classical approach of the NN-based technique, backpropagation neural networks (BPNN), was also tested. He and Kusiak (2017) employed SVM, Multi-layer perceptron (MLP) and ELM based methods, which performed satisfactorily, to predict wind power. This study contributes to the possibility of lengthening transformer life by proposing fault diagnosis systems using aforementioned data-driven approaches. The proposed models were trained and tested using the fault and NOC DGA data from IEC TC 10 and NOC DGA data from a thermal power plant in Jakarta. Ghoneim and Taha (2016) and Sahri and Yusof (2015) used IEC TC 10 as a data source as well. Only Ghoneim and Taha (2016) and Sahri and Yusof (2015) used real data, whereas Bacha et al. (2012) used simulated data from the Tunisian company of Electricity and Gas.

The remainder of this article is organized as follows. Section 2 describes the study methods used in this research, from the proposed research framework to the methods

applied. The following section, Section 3, elaborates the results and discusses the fault diagnosis based on certain performance measures. Some concluding remarks and areas of future research are provided in Section 4.

2. Methods

2.1. DGA as Condition Monitoring for Transformer

DGA samples are frequently used to analyze the dissolved gas in insulation oil. DGA data of a transformer provides insight into the electrical and thermal stresses of oil-immersed power transformers. Internal transformer faults are electrical and thermal. Owing to electrical and thermal stresses, oil and paper decomposition occurs. Both act as insulation for transformers. These two stresses can cause a breakdown of insulating material and release gaseous decomposition products (Zhang et al., 1996). Paper decomposition produces carbon monoxide (CO) and carbon dioxide (CO_2), both of which act as major factors in cellulose degradation. Griffin (1988) described the fault types related to these gases. Oil decomposition results in hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), and ethane (C_2H_6). CH_4 and C_2H_6 are related to low temperature oil breakdown, whereas C_2H_4 is related to high temperature oil breakdown. C_2H_2 is related to arcing, H_2 is related to corona, whereas CO and CO_2 are related to cellulose insulation breakdown, which are produced from paper decomposition. Based on the type and amount of gas, the fault type can be determined (Duval, 1989). The fault types used in the proposed research framework are based on classification in IEC 60599, which is simplified into five categories: partial discharge, discharges of low energy (D1), discharges of high energy (D2), thermal fault below 700°C and thermal fault above 700°C (Duval and dePablo, 2001).

2.2. Multistage Fault Diagnosis System

An FDD system, also known as a fault diagnosis system, is basically a classification problem, which consists of majority and minority data classes. An SVM is a binary type of classification; therefore, a nested (multistage) classification has been proposed for a fault diagnosis system. Previous studies, which employed SVM, have adopted the multistage fault diagnosis system (Dong et al., 2004; Bacha et al., 2012; Souza and Ramachandran, 2016, Dhini et al., 2018).

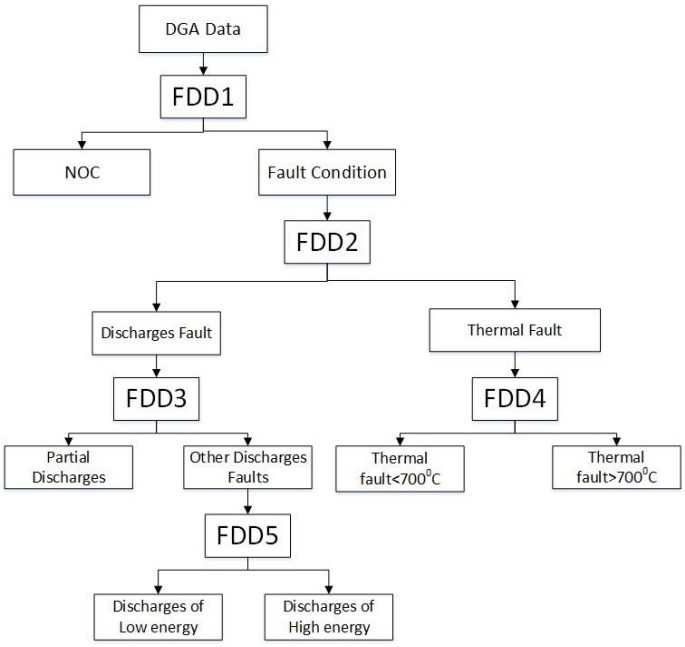

Figure 1 The proposed data-driven Fault Diagnosis system

Figure 1 shows the structure of the proposed data-driven fault diagnosis system, which is the continuation of previous works in Dhini et al. (2018). It is modified from the fault diagnostic model of Bacha et al. (2012). The fault detection and diagnosis 1 (FDD1) stage applies the fault-detection process by classifying the fault data from the NOC. The subsequent stages are the fault diagnosis process. The difference with Bacha et al. (2012) lies in the FDD3 stage, which classifies PD from other discharges faults, which are based on the similarity of certain fault types and discharges of high and low energy. Firstly, Bacha et al. (2012) separated discharges of high energy from discharges of low energy and partial discharges, whereas this study separated partial discharges from other types of discharge faults. Based on data availability, this study also classified thermal fault below 700°C from thermal fault above 700°C.

2.3. DGA Data Description

The proposed fault diagnosis system was tested on 266 DGA data, which consisted of 149 NOC and 117 fault data. Fifty NOC and 117 fault data originated from IEC TC 10 databases, which were collected from Duval and dePablo (2001). In addition, 99 NOC data were obtained from the DGA monitoring of a transformer in a government-owned steam power plant in Jakarta. The fault data consisted of five classes: nine partial discharges (PD), 26 discharges of low energy (D1), 48 discharges of high energy (D2), 16 thermal faults <700°C (T1 and T2) and 18 thermal faults >700°C (T3) data. The training and testing data were constructed using a hold out ratio of 80:20. Therefore, 209 training data and 57 testing data were included. The following seven types of dissolved gases were used as inputs. i.e.: hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO), and carbon dioxide (CO_2).

2.4. Single Minority Oversampling Technique (SMOTE) for Data Balancing

Subsequently, SMOTE was applied on training data for data balancing. Sun (2017) explained the process in SMOTE, where the synthetic data points are generated. First, the difference between the feature vector of the data point under consideration (x_i) and its K nearest neighbours (x_{zi}) should be calculated. Then, the difference is multiplied by a random number (λ) between 0 and 1, which is then added to the data point under consideration (x_i). This process results in the selection of a random point (x_{new}) along the line segment between two specific data points (x_i and x_{zi}). Equation 1 describes the process of synthetic data generation:

$$x_{new} = x_i + \lambda \times (x_{zi} - x_i) \quad (1)$$

There are four types of SMOTE: regular, borderline SMOTE1, borderline SMOTE2, and SVM (Chawla et al., 2002; Han et al., 2005; Tang et al., 2009). This study applied these four types of SMOTE, and the process was conducted only on the training data. The regular algorithm randomly picks all possible x_i available when instantiating a SMOTE object. The borderline versions over-sample only minority classes around the borderline, whereas the SVM version uses the support vectors to create a new sample (Han et al., 2005; API, 2014).

Borderline SMOTE1 and SMOTE2 classify each data point x_i as the following: (1) noise, when all the nearest neighbours come from a different class than the one of x_i; (2) danger, when at least half of the nearest neighbours are from a different class than x_i; or (3) safe, when most of the nearest neighbours are from the same class as x_i. When instantiating a SMOTE object, borderline SMOTE1 and SMOTE2 will use the data points in danger to generate new data points. In borderline SMOTE1, one of the K nearest neighbours x_{zi} will belong to a class different from the one of data point x_i. In contrast, borderline SMOTE2 will consider a neighbor x_{zi}, which can be from any class. Borderline SMOTE2 will not only

generate synthetic data points from each example in danger and its minority nearest neighbours but also does the same from its nearest majority neighbour.

2.5. Support Vector Machine (SVM)

SVM, which was first introduced by Vapnik during the 1990s, is a classifier based on a linear discriminant function, and it has become popular during the past few decades. The objective of an SVM is to find a hyperplane that maximizes the margin between the separated data. Figure 2 illustrates a hyperplane that separates two classes. The closest circles and triangles to the margin are the support vectors. The remaining are the training data. To construct an optimal hyperplane for classification, a non-linear function, namely, a kernel (φ), is selected to map the dimension transformation (Dong et al. 2004). The input vector, x, may be transformed into a higher dimensional characteristic space. In this study, a polynomial kernel as in Equation 2 is applied:

$$K(x,u) = (ax^T u + C)^q, q > 0 \tag{2}$$

An SVM was developed from a linear classification problem. For a non-linear problem, a relaxation parameter, $\xi \geq 0$, is introduced. The problem objective function is then the following:

$$\min \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \tag{3}$$

subject to the following constraints:

$$y_i\left[w_i^T x_i + b\right] \geq 1 - \xi_i, i = 1,2,\dots,N \tag{4}$$

$$\xi_i \geq 0, i = 1,2,\dots,N \tag{5}$$

where C is a factor, which should be a compromise between the classification accuracy and algorithm complexity and is often determined through a cross-validation.

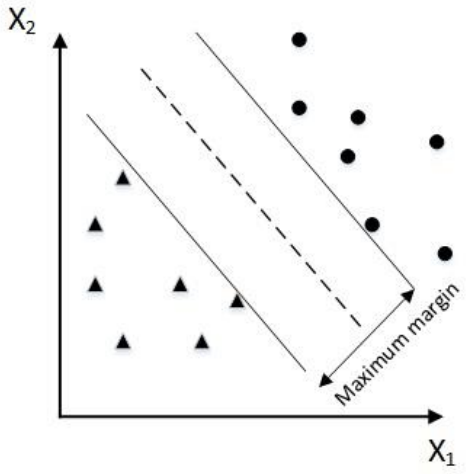

Figure 2 Classification by maximized hyperplane in SVM

The optimal classification hyperplane is constructed to solve the quadratic programming problem and to find only the minimum solution, which is solved using the minimum point of Lagrange's function as follows:

$$\mathcal{L}(w,\xi,\mathrm{b},\alpha) = \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i\left[1 - \xi_i - y_i\left(w_i^T x_i + b\right)\right] \tag{6}$$

where w and b are primal variables, and α_i is a Lagrange multiplier, in which $\alpha_i \geq 0$. According to the Karush-Kuhn-Tucker (KKT) theorem, the classification threshold value b

may be obtained using the corresponding data, which satisfies the following condition: $0<\alpha_i<C$. The subsequent Equation 7 is for decision making, based on the optimal classification hyperplane.

$$f(x) = sgn(wx + b) \tag{7}$$

One-versus-the-rest or one-against-all (OAA) is the earliest and most common extension of an SVM for a multiclass problem. A total of M binary SVM classifiers will be created, where *M* is also the number of classes. This study applies the OAA approach to solve the multiclass problems using an SVM.

2.6. Extreme Learning Machine-Radial Basis Function (ELM-RBF)

The seminal work by Broomhead and Lowe (1988) proved that RBF networks corresponded with the radial functions in a single-layer network. The network architecture consists of a three-layered feedforward NN. The first layer (input layer) distributes the input signal linearly. The second layer (hidden layer) is non-linear and employs radial basis functions, and the third layer (output layer) linearly combines the radial basis function outputs. The commonly applied radial basis function is Gaussian. A bias neuron is added to improve the accuracy of the model. Traditionally, the tap weights between the hidden and output layers are adjusted during learning using the training data. One of the common techniques employed to optimize the weights is a gradient descent algorithm, which converges extremely slowly if the learning rate is too low. Thus, an ELM was proposed to overcome this weakness. In a single-hidden-layer feedforward NN, ELM randomly chooses the input weight, and therefore the hidden biases of the neurons analytically determine the output weights (Huang and Siew, 2005). Moreover, Huang and Siew (2005) claimed that the use of an ELM resulted in better generalization and an extremely high learning speed.

The ELM-RBF randomly generates the kernel centers and the impact widths of the RBF kernels, and it analytically calculates the output weights. As determined by Huang and Siew (2005), (x_i,y_i), in which $x_i = [x_{i1},x_{i2},\ldots,x_{in}]^T \in R^n$ and $t_i = [t_{i1},t_{i2},\ldots,t_{im}]^T \in R^m$, RBFs with Ñ kernels can be mathematically modeled as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i \emptyset_i(x_j) = o_j, \quad j = 1, \ldots, N \tag{8}$$

where $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{ik}]^T$ is the weight vector connecting kernel *i* and the output neurons, and $\emptyset_i$ (x) is the output of kernel *i*. In addition, $\mu_i = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}]^T$ is the center of kernel *i*, σ_i is the impact width, and $\emptyset$ is the radially symmetric kernel function; in addition, it is assumed that $\emptyset_i$ (x) is a non-linear bounded integral and is always nearly continuous:

$$\sum_{i=1}^{\tilde{N}} \beta_i \emptyset_i(\mu_i, \sigma_i, x_j) = t_j, j = 1, \ldots, N. \tag{9}$$

$$H\beta = Y \tag{10}$$

In short, as in Equation 10, where *H* is the hidden-layer output matrix of the RBF network. Column *i* of *H* is the output of kernel *i* with respect to inputs $x_1, x_2, \ldots, x_N$. However, because Ñ ≠ N (the number of kernels is less than the number of training samples, Ñ << N), *H* is a non-square matrix, and β_i = (*i* = 1, 2, …, Ñ) may not exist, such that $H\beta = T$. Therefore, the unique smallest-norm least-squares solution $\hat{\beta}$ of the above linear system is

$$\hat{\beta} = H^+T \tag{11}$$

where H^+ denotes the Moore-Penrose generalized inverse.

3. Results and Discussion

3.1. Accuracy Comparison of SMOTE types

Table 1 summarizes the average results of the accuracy classification in each stage. The numbers shown are the average of ten iterations for each SMOTE type. The SMOTE program was run on Python 2.7. The training and testing data using any SMOTE types resulted in similar accuracy from FDD1 to FDD5. However, borderline SMOTE1 had the lowest standard deviation of training data and the highest overall accuracy (81.6%). However, the p-value of the one-way ANOVA test was 0.996, which means the differences among the various SMOTE approaches were non-significant, Borderline SMOTE1 still exhibited better performance compared to the others. Therefore, training data resulting from borderline SMOTE 1 were used in the next stage of fault classification.

3.2. Comparison of Classification Algorithms: SVM, ELM-RBF and BPNN

After data balancing with SMOTE, the fault classification algorithm SVM was conducted. The kernel type used was a polynomial (order 3) as in Equation 2, and the optimization problem was solved using quadratic programming (QP). For comparison, ELM-RBF and BPNN were employed. For clustering data in the hidden layer, RBF-type approaches were introduced, that is, a self-organizing map (SOM) and k-means. For ELM-RBF (SOM), the number of clusters selected ranged from 15 to 30, whereas for ELM-RBF (Kmeans), 5 to 10 hidden neurons were applied. The BPNN was run using the following setting criteria: (1) a maximum iteration of 100; (2) 34 hidden neurons; (3) alpha = 0.1; (4) momentum = 0.3; and (5) a sigmoid activation function. The classification programs were run on MATLAB R2017b. The accuracy at each stage, overall accuracy, and standard deviations are presented in Table 2. Both ELM-RBF and BPNN were run ten times, and the results were averaged.

Table 1 Comparison of accuracy and standard deviation using different SMOTE types

	SMOTE Types							
	Regular		Borderline SMOTE1		Borderline SMOTE2		SVM	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
FDD1	86.60%	82.46%	86.12%	82.46%	86.12%	82.46%	87.56%	80.70%
FDD2	94.51%	73.08%	93.41%	80.77%	94.51%	73.08%	95.60%	76.92%
FDD3	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
FDD4	92.31%	75.00%	92.31%	75.00%	92.31%	75.00%	92.31%	87.50%
FDD5	86.21%	68.75%	86.21%	62.50%	81.03%	68.75%	86.21%	62.50%
Std Dev	0.06	0.12	0.06	0.14	0.07	0.12	0.06	0.14
OverallAcc		80.80%		81.60%		80.80%		80.80%

Table 2 Accuracy Results of SVM, ELM-RBF and BPNN

	No SMOTE		SMOTE - BorderlineSMOTE1							
	SVM		SVM		ELM-RBF (SOM)		ELM-RBF (Kmeans)		BPNN	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
FDD1	87.56%	82.46%	86.12%	82.46%	82.44%	82.11%	80.38%	81.58%	87.66%	84.56%
FDD2	95.60%	76.96%	93.41%	80.77%	81.43%	80.00%	78.35%	77.31%	91.10%	74.69%
FDD3	100.00%	94.44%	100.00%	100.00%	98.62%	89.44%	93.08%	83.33%	100.00%	98.33%
FDD4	92.31%	87.05%	92.31%	75.00%	70.00%	50.00%	66.54%	48.75%	88.46%	75.00%
FDD5	97.93%	68.75%	86.21%	62.50%	82.20%	60.00%	59.48%	55.63%	82.93%	81.25%
StDev	0.05	0.10	0.06	0.14	0.10	0.17	0.13	0.16	0.06	0.10
TestAcc		81.58%		81.60%		77.84%		75.52%		83.46%

The first two columns use original data for training (without SMOTE), while the rests show the accuracy of all methods using SMOTE for training data. Only minor differences between the accuracy of SVMs with or without SMOTE (p-value = 0.6761 from the Mann-Whitney test) were detected. The effect of SMOTE was more significant for large differences in the number of instances between the two classes, such as 1:100 or 1:1000. In this study, five years of observation, which were originally 3,988 instances, revealed that only 99 NOC data remained as a result of the IEC and TDCG combination requirements. However, for the first three stages, FDD1 to FDD3, SVM with SMOTE presented better results than SVM without SMOTE.

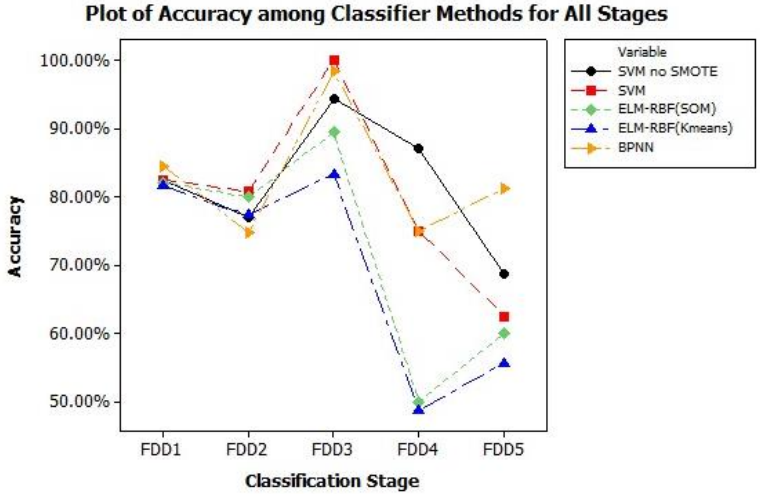

Figure 3 Accuracy for each method in each stage

For data with SMOTE, the overall accuracy of BPNN outperformed SVM and ELM-RBF. Its standard deviation for testing was the same as SVM without SMOTE, which was the smallest: 0.1. However, in a more detailed analysis, in the second and the third stages, SVM displayed better performance than BPNN. Figure 3 shows the accuracy for each method graphically to clarify the results presented in Table 2. In FDD1, when fault data is detected from normal data, the performance among all classifier methods are similar (approximately 80% accurate). With a decreasing number of training and testing data, the accuracy among methods exhibited larger variances, such as in FDD4 and FDD5. ELM-RBF resulted in the lowest accuracy in FDD4 and FDD5. This may be due to the limited data training in these stages. Consequently, they had relatively lower accuracy than the other stages at only approximately 60% to 80% for training, and only 50% to 60% for testing data. In contrast, SVM still performed well in the last stages, proving its advantages in cases of limited data. Despite the fact that BPNN also displayed similar accuracy to SVM in the last stages, it required significantly longer computation times. The computation time comparison is depicted in Table 4.

The accuracy of the testing data, which shows the generalization capability of the developed model, resulted in a slightly lower performance. The highest accuracy was 100% for FDD3, whereas the lowest accuracies were only approximately 50% for the FDD 4 and FDD5 stages. The testing data accuracy for ELM-RBF was also lower than that of the SVM, which may be due to data limitation. When data is limited, SVM will have a better generalization capability. Figure 3 supports this relationship. Despite the similar accuracy at the beginning of the FDD process, along with the lower number of data, the SVM and BPNN consistently exhibited better performance. These results support previous studies that demonstrated the high performance of SVM and that demonstrated BPNN as the most

popular NN-based method (Guardado et al., 2001; Dong et al., 2004; Bacha et al., 2012; Souahlia et al., 2012; Sahri and Yusof, 2015; Souza and Ramachandran, 2016). However, ANOVA test results showed that the accuracy differences among the various classifier methods in all stages were non-significant (p-value = 0.496).

The accuracy in this study was higher than previous studies that used the same data (IEC TC 10). Ghoneim and Taha (2016) displayed only 71.71% accuracy using limit rules method, whereas Sahri and Yusof (2015) displayed 51.15% and 51.92% accuracy using SVM and GA-SVM, respectively. Bacha et al. (2012) had 73.73% accuracy using SVM with similar key gas inputs; however, they used experimental data from the Tunisian Company of Electricity and Gas. The results of this study may not be completely satisfactory; however, its 80% accuracy represents considerable improvement.

The other performance metric is recall or sensitivity. Recall measures the ratio of true positive (TP) to the total number of TP and false negative (FN), or as follow: TP/(TP+FN). By convention, the positive class is the minority, which is the main concern. FN describes a condition where a fault datum is classified as a normal datum. In contrast, a false positive (FP) occurs when a normal datum is detected as a fault datum. In the FDD problem, the cost of a FN is higher than that of a FP, and the recall or sensitivity is more suitable for the performance measurement. In this study, the computations were conducted only for SVM and BPNN, and recall was only calculated in this first stage.

Table 3 Accuracy versus recall of FDD1

	SVM		BPNN	
FDD1	Training Data	Testing Data	Training Data	Testing Data
Accuracy	86.12%	82.46%	88.04%	84.21%
Recall/Sensitivity	81.32%	76.92%	87.91%	80.77%

Table 3 only depicts a comparison of the recall and accuracy for the FDD1 stage, and the BPNN results are derived from running a program only once. Therefore, the exact number of FNs can be calculated. In the testing data, the recall is slightly lower than that of the training data. There is no major difference between accuracy and recall. This indicates that during the classification process, not many FNs were found or misclassified as positive data (fault) into negative data (NOC). It means that the performance of both classifiers were satisfactory.

Table 4 Computation time for each stage

	Computation time (s)		
	SVM	ELM-RBF	BPNN
FDD1	2.560	0.045	6.945
FDD2	1.215	0.026	5.150
FDD3	0.920	0.013	4.496
FDD4	0.733	0.007	2.409
FDD5	0.773	0.011	3.666
Total	6.201	0.102	22.666

Table 4 shows that on average, the ELM-RBF finishes the computations faster than the SVM, (by only 0.102 s in total), whereas SVM is the second fastest (6.201 s) and BPNN is the slowest (22.666 s). Such results are supported by previous studies. Despite BPNN demonstrating good accuracy, it has some drawbacks, as stated by Kusumoputro et al. (2016). These include high computational costs, which is shown by longer computation

time, and the risk of being trapped in local minima. The slowest computation time of BPNN was due to the gradient descent algorithm applied in the backpropagation, whereas SVM used more simple calculations, similar to regression principles. In contrast, ELM-RBF has an extremely high learning speed, as claimed in Huang and Siew (2005), which can be a hundred times faster than SVM.

ELM-RBF may be preferable over SVM when big data are applied, in which case the difference in accuracy, along with a higher number of data instances, may become insignificant. The overall results revealed that in terms of accuracy, BPNN was slightly higher than SVM. However, the computation time of BPNN was almost three times slower than SVM. The SVM approach with and without SMOTE resulted in higher accuracy and consistency compared to the ELM-RBF approach; however, BPNN achieved slightly higher accuracy than SVM. The SVM still performed significantly better for a limited amount of data because it displayed better generalization, whereas BPNN exhibited the best accuracy.

4. Conclusions

Based on the two performance measures of accuracy and computation time, it can be concluded that the SVM outperforms other NN-based methods, ELM-RBF and BPNN, for the proposed multistage fault diagnosis of power transformers using DGA data. SVM performs satisfactorily with high accuracy and fast computation time. It has proven to be an effective method in terms of classification with a better generalization. Despite its poorer results compared to SVM, ELM-RBF also performed the best in terms of computation time. In terms of accuracy, BPNN exhibited the best accuracy. In the future, research to improve accuracy, data quantity, and data quality should be enhanced by adding more training data, selecting optimum features, and combining other types of input data, such as gas ratio. More comparisons with other data-driven methods, such as decision trees, kNN, or even deep learning, and applying optimization in parameter selections, represent additional options.

References

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 Transactions on Dielectrics and Electrical Insulation*, Volume 19(3), pp. 1007–1012

API, 2014. Imbalanced-Learn Documentation: User Guide. Available Online at: https://imbalanced-learn.org/stable/over_sampling.html#smote-variants. Accessed on April 18, 2020

Bacha, K., Souahlia, S., Gossa, M., 2012. Power Transformer Fault Diagnosis based on Dissolved Gas Analysis by Support Vector Machine. *Electric power systems research*, Volume 83(1), pp. 73–79

Broomhead, D.S., Lowe, D., 1988. Radial Basis Functions, Multi-variable Functional Interpolation and Adaptive Networks. *Royal Signals and Radar Establishment*. Memorandum 4148

Chakravorti, S., Dey, D., Chatterjee, B., 2013. Recent Trends in the Condition Monitoring of Transformers. *Power Systems*. Springer-Verlag, London, United Kingdom

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligent Research*, Volume 16, pp. 321–357

Dhini, A., Surjandari, I., Faqih, A., Kusumoputro. 2018. Intelligent Fault Diagnosis for Power Transformer Based on DGA Data Using Support Vector Machine (SVM). *In*: The 3rd International Conference on System Reliability and Safety, pp. 294–298

Dong, M., Xu, D., Li, M., Yan, Z., 2004. Fault Diagnosis Model for Power Transformer based on Statistical Learning Theory and Dissolved Gas Analysis. *In*: Conference Record of the 2004 IEEE International Symposium on Electrical Insulation, pp. 85–88

Duval, M., 1989. Dissolved Gas Analysis: It Can Save Your Transformer. *IEEE Electrical Insulation Magazine*, Volume 5(6), pp. 22–27

Duval, M., dePablo, A., 2001. Interpretation of Gas-in-oil Analysis using New IEC Publication 60599 and IEC TC 10 Databases. *IEEE Electrical Insulation Magazine*, Volume 17(2), pp. 31–41

Fei, S.-w., Liu, C.-l., Miao, Y.-b., 2009. Support Vector Machine with Genetic Algorithm for Forecasting of Key-gas Ratios in Oil-immersed Transformer. *Expert Systems with Applications*, Volume 36(3) Part 2, pp. 6326–6331

Gao, Z., Ding, S.X., Cecati, C., 2015. Real-time Fault Diagnosis and Fault-tolerant Control. *IEEE Transactions on Industrial Electronics*, Volume 62, pp. 3752–3756

Ghoneim, S.S., Taha, I.B., 2016. A New Approach of DGA Interpretation Technique for Transformer Fault Diagnosis. *International Journal of Electrical Power & Energy Systems*, Volume 81, pp. 265–274

Griffin, P.J., 1988. Criteria for the Interpretation of Data for Dissolved Gases in Oil from Transformers (A Review). *Electrical Insulating Oils*, ASTM International, pp. 89–107

Guardado, J.L., Naredo, J.L., Moreno, P., Fuerte, C.R., 2001. A Comparative Study of Neural Network Efficiency in Power Transformers Diagnosis using Dissolved Gas Analysis. *IEEE Transactions on Power Delivery*, Volume 16(4), pp. 643–647

Han, H., Wang, W.-Y. Mao, B.-H., 2005. Borderline-SMOTE: A New Over-sampling Method in Imbalanced Data Sets Learning. *In*: International Conference on Intelligent Computing, pp. 878–887

He, Y., Kusiak, A., 2017. Performance Assessment of Wind Turbines: Data-Derived Quantitative Metrics. *IEE Transactions on Sustainable Energy*, Volume 9(1), pp. 65–73

Hernandez, M.D.P.C., Labib, A., 2017. Selecting a Condition Monitoring System for Enhancing Effectiveness of Power Transformer Maintenance. *Journal of Quality in Maintenance Engineering*, Volume 23(4), pp. 400–414

Huang, G.-B., Siew, C.-K. 2005. Extreme Learning Machine with Randomly Assigned RBF Kernels. *International Journal of Information Technology*, Volume 11(1), pp. 16–24

Kusiak, A. Song, Z. 2009. Sensor Fault Detection in Power Plants. *Journal of Energy Engineering*, Volume 135(4), pp. 127–137

Kusumoputro, B., Sutarya, D., Faqih, A., 2016. Performance Analysis of an Automatic Green Pellet Nuclear Fuel Quality Classification using Modified Radial Basis Function Neural Networks. *International Journal of Technology*. Volume 7(4), pp. 709–719

Lin, C.E., Ling, J.-M., Huang, C.-L., 1993. An Expert System for Transformer Fault Diagnosis using Dissolved Gas Analysis. *IEEE transactions on Power Delivery*, Volume 8(1), pp. 231–238

Lv, G., Cheng, H., Zhai, H., Dong, L., 2005. Fault Diagnosis of Power Transformer based on Multi-layer SVM Classifier. *Electric Power Systems Research*, Volume 75(1), pp. 9–15

Pariaman, H., Garniwa, I., Surjandari, I., Sugiarto, B., 2017. Availability Analysis of the Integrated Maintenance Technique based on Reliability, Risk, and Condition in Power Plants. *International Journal of Technology*, Volume 8(3), pp. 497–507

Saha, T.K., 2003. Review of Modern Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers. *IEEE Transactions on Dielectrics and Electrical Insulation*, Volume 10(5), pp. 903–917

Sahri, Z., Yusof, R., 2015. Fault Diagnosis of Power Transformer using Optimally Selected DGA Features and SVM. *In*: IEEE 10th Asian Control Conference (ASCC)

Shin, H., Cho, S., 2006. Response Modeling with Support Vector Machines. *Expert Systems with Applications*, Volume 30(4), pp. 746–760

Shintemirov, A., Tang, W., Wu, Q., 2009. Power Transformer Fault Classification based on Dissolved Gas Analysis by Implementing Bootstrap and Genetic Programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Volume 39(1), pp. 69–79

Souahlia, S., Bacha, K., Chaari, A., 2012. MLP Neural Network-based Decision for Power Transformers Fault Diagnosis using an Improved Combination of Rogers and Doernenburg Ratios DGA. *International Journal of Electrical Power & Energy Systems*, Volume 43(1), pp. 1346–1353

Souza, F.R., Ramachandran, B., 2016. Dissolved Gas Analysis to Identify Faults and Improve Reliability in Transformers using Support Vector Machines. *In*: IEEE Power Systems Conference (PSC), Clemson University

Sun, S., 2017. *Imbalanced Binary Classification for Detecting Transcription Factor Binding Sites in H1 Human Embryonic Stem Cells*, UCLA Electronic Theses and Dissertations

Tang, Y., Zhang, Y.-Q., Chawla, N.V., Krasser, S., 2009. SVMs Modeling for Highly Imbalanced Classification. *IEEE Transactions on Systems, Man, Cybernetics, Part B*, Volume 39(1), pp. 281–288

Ulum, B., Nurrohman, Ambarita, E., Gaos, Y.S., 2017. Energy and Exergy Analysis of Mount Salak Geothermal Power Plant Unit 1-2-3. *International Journal of Technology*, Volume 8(7), pp. 1217–1228

Yang, Z., Tang, W., Shintemirov, A., Wu, Q., 2009. Association Rule Mining-based Dissolved Gas Analysis for Fault Diagnosis of Power Transformers. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Volume 39(6), pp. 597–610

Zhang, Y., Ding, X., Liu, Y., Griffin, P., 1996. An Artificial Neural Network Approach to Transformer Fault Diagnosis. *IEEE Transactions on Power Delivery*, Volume 11(4), pp. 1836–1841