

Article

A Novel Integrated Method to Diagnose Faults in Power Transformers

Jing Wu ^{1,*}, Kun Li ¹, Jing Sun ¹ and Li Xie ²

¹ School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China; jackli5288@gmail.com (K.L.); jing.sun@buaa.edu.cn (J.S.)

² China Electric Power Research Institute, Beijing 100192, China; xieli@epri.sgcc.com.cn

* Correspondence: wujing06@buaa.edu.cn; Tel.: +86-10-82317304

Received: 10 September 2018; Accepted: 15 October 2018; Published: 5 November 2018



Abstract: In a smart grid, many transformers are equipped for both power transmission and conversion. Because a stable operation of transformers is essential to maintain grid security, studying the fault diagnosis method of transformers can improve both fault detection and fault prevention. In this paper, a data-driven method, which uses a combination of Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), and Support Vector Machines (SVM) to enable a better fault diagnosis of transformers, is proposed and investigated. PCA is used to reduce the dimension of transformer fault state data, and an improved PSO algorithm is used to obtain the optimal parameters for the SVM model. SVM, which is optimized using PSO, is used for the transformer-fault diagnosis. The diagnostic-results of the actual transformers confirm that the new method is effective. We also verified the importance of data richness with respect to the accuracy of the transformer-fault diagnosis.

Keywords: smart grid; transformer-fault diagnosis; principal component analysis; particle swarm optimization; support vector machine

1. Introduction

Several advanced technologies can be used to monitor power-equipment, and the large amount of status data of used equipment helps make the power grid “smarter”. Power transformers are expensive and important components of the smart grid, and hub devices for power transformation and transmission [1–4]. Because various faults, such as discharge and overheating, can occur during the operation of transformers, many characteristics corresponding to the faults can be affected like dissolved gases (H₂, CH₄, C₂H₆, C₂H₄, C₂H₂, CO, etc.), organic compounds (methanol, ethanol and 2-furfural), as well as the current and power of the transformers [5]. The dissolved gas analysis (DGA) is a common tool for monitoring and identifying transformer’s faults. IEEE C57.104. and IEC 60599 provide different methods such as key gases, Doernenburg Ratio, Rogers, three basic gas ratio, Duval triangle, and so on. However, due to the complexity of the working environment and the process structure of the transformers, these methods are not enough to make a right judgement and cannot judge fault fuzzy boundary. According to [6], their accuracy rates are about in 60%, which means the ratio methods cannot account for the diagnostic criteria completely [7]. In addition, the concentrations of cellulose chemical markers in oil, such as methanol, ethanol and 2-furfural, are used as a determination mark for diagnosing transformer insulation failure, which still present many challenges for an accurate interpretation in real transformers [8].

To improve the accuracy of fault diagnosis, artificial intelligence and machine learning algorithm were added to the field of transformer-fault diagnosis (TFD), including fuzzy sets [9], artificial neural networks (ANN) [10], artificial immune networks [11], probabilistic neural networks [12], rough sets [13], and support vector machines (SVM) [14]. These algorithms provide ways to develop new

TFD technologies. However, these algorithms have some disadvantages. For example, it is difficult to determine the selection of parameters of fuzzy sets, artificial immune networks and probabilistic neural networks, ANNs are easier to fall into local minimum, and the fault-tolerant ability and generalization ability of rough set are weak.

SVM is usually used as a classification tool. From early 2-category techniques, multi-class SVM have been developed and are more suitable for TFD. The accuracy of multi-class SVM is determined by the parameters of its kernel function and penalty factor. In order to improve the efficiency of SVM in processing large amounts of input fault data, principal component analysis (PCA) will be used. Moreover, to reduce the influence of human experience and subjective judgment on these parameters, a new Particle Swarm Optimization (PSO) is borrowed to search the optimized parameters. This way, the most suitable SVM parameters within the effective input data reflecting the transformer's fault will be found. SVM integrated with PCA and PSO can improve the speed and accuracy of TFD considerably. This paper is organized as follows. Section 2 introduces the complete TFD procedure implemented by improved SVM; In Section 3, we compare the accuracy of transformer-fault diagnosis using different methods. We then verify the effectiveness of the proposed method, and analyze the effect of data richness on the accuracy of the fault diagnosis. Section 4 summarizes all results.

2. TFD Model Based on SVM Integrated with PCA and PSO

TFD model based on SVM integrated with PCA and PSO is shown in Figure 1. It includes two main parts. One is that a set of transformer fault data (Data set) such as the densities of the dissolved gases is preprocessed by PCA. The other is that the parameters of SVM model are searched and optimized by PSO.

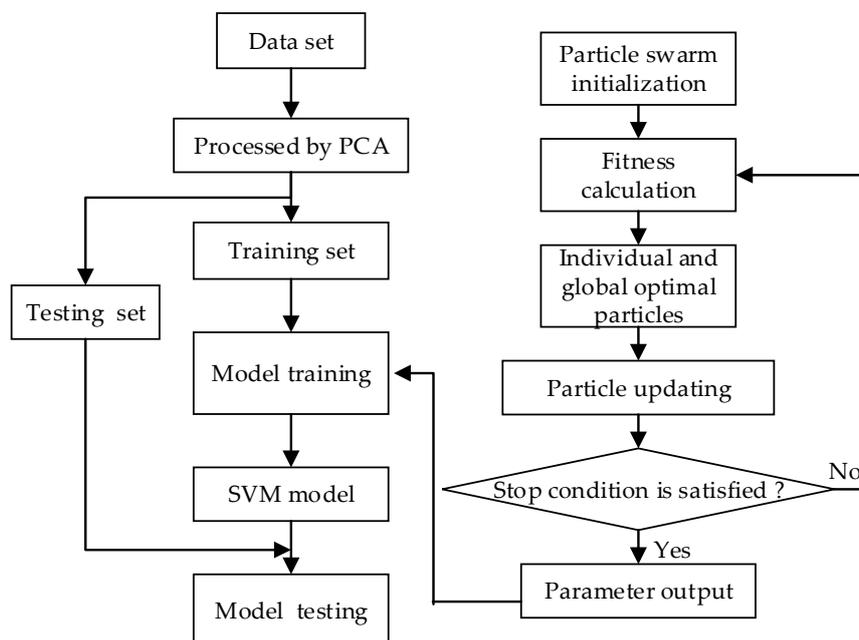


Figure 1. Transformer-fault diagnosis (TFD) model based on support vector machines (SVM) integrated with principal component analysis (PCA) and particle swarm optimization (PSO).

2.1. Data Set Preprocessed by PCA

TFD is a complicated task. In order to improve the operating efficiency of the SVM when there are many transformer fault data, the data needs to be pre-processed before they are used to train the SVM model. PCA aims to reduce the dimensions of fault data and replaces them with fewer uncorrelated and unoverlapped data (called principal components). The number of principal components is selected by variance contribution rate indicating how much information is included.

Suppose the data set X has n groups and each group has p fault data and they construct an original data observation matrix:

$$X_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \quad (1)$$

To solve the principal components, it needs to find $i (i \leq p)$ linear functions: $y_{ji} = A_i^T X_j$, where $j = 1, \dots, n$, $X_j = (x_{j1}, x_{j2}, \dots, x_{jp})^T$, and $A_i = (a_{1i}, a_{2i}, \dots, a_{pi})^T$ is unknown. The amount of information of x is proportional to its variance. Letting $Y_{ji} = (y_{1i}, y_{2i}, \dots, y_{ni})^T$ and to avoid $\text{var}(Y_{ji}) \rightarrow \infty$, $\|A_i\|^2 = 1$ Therefore, to obtain the maximum variance, the following equations of conditional extremes are formed:

$$\begin{cases} \max \text{var}(Y_{ji}) = \max A_i^T \Sigma A_i \\ A_i^T A_i = 1 \end{cases} \quad (2)$$

where Σ represents the covariance matrix.

Here Lagrange multiplier method is used to solve (2). The Lagrangian objective function is expressed as:

$$Q(A_i) = A_i^T \Sigma A_i - \lambda_j (A_i^T \Sigma A_i - 1) \quad (3)$$

where the Lagrange multiplier λ_j is the characteristic root of Σ and A_i is the corresponding eigenvector. Because $A_i \neq 0$ and $A_i^T \Sigma A_i = \text{var}(A_i^T X_j) > 0$, Σ is positive definite and all characteristic roots are positive. Assuming that:

$$\lambda_{j1} \geq \lambda_{j2} \geq \dots \geq \lambda_{jp} \geq 0 \quad (4)$$

In the practical applications, only p' principal components will be selected, which satisfies $\sum_{k=1}^{p'} \lambda_{jk} / \sum_{i=1}^p \lambda_{ji} \geq 0.85$. The k -th principal component for the j -th group is $y_{jk} = A_k^T X_j$. All the principal components form a vector $Y_j = (y_{j1}, y_{j2}, \dots, y_{jp'})^T$.

2.2. Support Vector Machine

Suppose the j -th group of principal component Y_j reflects the fault type z_j . We divide n groups of fault data into two sets. One set is the training set including l groups and the other set is the testing set including $(n-l)$ groups. The training set is used to solve the parameters of SVM.

TFD is usually a multi-class problem to classify the categories in $d (d \geq 2)$. The one-versus-one (OVO) method is adopted to extend 2-category SVM to multi-class SVM in this paper. This means it need to build SVM classifiers for any two different fault types F_1 and $F_2 (F_1, F_2 \in \{z_j\})$, and there are a total of $d(d-1)/2$ classifiers. Assume a hyperplane function $\omega^T \varphi(Y) + b = 0$ can accurately separate F_1 and F_2 whose category labels are marked in -1 and 1 . Here ω is the normal vector of the hyperplane, b is the offset, and $\varphi(y)$ is nonlinear transformation function. For the optimal classification hyperplane, the following conditions should be satisfied:

$$\begin{cases} \omega^T \varphi(Y_j) + b \geq 1, H_j = 1 \\ \omega^T \varphi(Y_j) + b \leq -1, H_j = -1 \end{cases} \quad (5)$$

and $H_j \in \{F_1, F_2\}$ is the classification of Y_j . In this case, Y_j is mapped into a high-dimensional space.

The maximum margin between the plane and the nearest data is $1/\|\omega\|$. The greater it is, the better the classification confidence is. To increase the misclassification tolerance of SVM, a non-negative variable e_j is introduced. Then the problem can be described as:

$$\begin{cases} \min \left(\frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^l e_j \right) \\ H_j(\omega^T \varphi(Y_j) + b) \geq 1 - e_j \\ e_j \geq 0 \end{cases} \quad (6)$$

where C is a constant named penalty factor and controls the punishment degree for misclassified data. Lagrange multiplier method is also used to solve (6). The corresponding Lagrangian function is:

$$L(\omega, \sigma, e_j, \alpha_j, \beta_j) = \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^l e_j - \sum_{j=1}^l \alpha_j [H_j(\omega^T \varphi(Y_j) + b) - 1 + e_j] - \sum_{j=1}^l \beta_j e_j \quad (7)$$

where $\alpha_j > 0$ and $\beta_j > 0$ are the Lagrangian multipliers. After $\alpha_j (j = 1, \dots, l)$, ω and b are solved, the final SVM classification function is:

$$f(Y'_r) = \text{sgn} \left(\sum_{j=1}^l \alpha_j H_j K(Y'_r, Y_j) + b \right) \quad (8)$$

where $Y'_r (r = l + 1, \dots, n)$ is the r -th group data in testing set, $K(Y'_r, Y_j) = \varphi(Y'_r) \cdot \varphi(Y_j)$ is the kernel function and we choose Gaussian radial basis function:

$$K(Y'_r, Y_j) = \exp \left(-\frac{\|Y'_r - Y_j\|^2}{\sigma^2} \right) \quad (9)$$

where σ is the parameter of kernel function.

2.3. Parameter Optimization in SVM Using Improved PSO

As mentioned before, when using SVM for fault diagnosis, we first need to determine the parameter σ in kernel function (9) and penalty factor C in (6). σ affects the optimal classification performance and generalization ability of the SVM. C is required to balance the learning machine's complexity and empirical risk when determining the minimization of the objective function. Therefore, σ and C should be optimized. We use an improved PSO algorithm for optimization.

Assuming that in a 2-dimensional search space, there is a swarm including S particles, $q_s = (q_{s1}, q_{s2}) (s = 1, \dots, S)$. Each particle represents a potential solution and corresponds to a point in the 2-dimensional search space. Its velocity is $v_s = (v_{s1}, v_{s2})^T$ and optimal position is $P_s = (P_{s1}, P_{s2})^T$. The optimal position within the S -particle population represents the global extremum, and it is set to $P_g = (P_{g1}, P_{g2})^T$. The position-updating method for the particle's velocity is expressed as:

$$v_{sd}(t+1) = wv(t) \cdot v_{sd}(t) + c_1(t)r_1(t)(P_{sd}(t) - q_{sd}(t)) + c_2(t)r_2(t)(P_{gd}(t) - q_{sd}(t)), d = 1, 2 \quad (10)$$

$$q_{sd}(t+1) = q_{sd}(t) + v_{sd}(t+1) \quad (11)$$

where $c_1(t)$ and $c_2(t)$ are acceleration constants, $r_1(t)$ and $r_2(t)$ obey the (0,1)-uniform distribution, $wv(t)$ is the speed update inertia weight representing the effect of the previous generation's particles on the next generation particles' velocity during the particle updating process.

Generally, the algorithm has relatively strong global optimization capability when $wv(t)$ is large, and a relatively strong local optimization capability when $wv(t)$ is small. However, the linear weight-adjustment method is single, and thus limits the optimization of the search ability. Aiming to

change to single adjustment mode and better adapt to the complex environment, we present a new scheme for the stochastic inertia weight:

$$\begin{cases} wv(t) = \alpha_1 + \varepsilon/2.0, k \geq 0.05 \\ wv(t) = \alpha_2 + \varepsilon/2.0, k < 0.05 \end{cases} \quad (12)$$

$$k = \frac{\text{fit}(t) - \text{fit}(t-10)}{\text{fit}(t-10)} \quad (13)$$

where $\text{fit}(t)$ represents the optimal fitness value of the t -th generation and $\text{fit}(t-10)$ is the optimal fitness value of the $(t-10)$ -th generation, α_1 and α_2 are set to 0.5 and 0.4, respectively, reflecting the search ability in different situations, and ε is a random value between 0 and 1. The acceleration constants $c_1(t)$ and c_2 are modified in:

$$\begin{cases} c_1(t) = c_{11} + (c_{12} - c_{11}) \frac{t}{T} \\ c_2(t) = c_{21} + (c_{22} - c_{21}) \frac{t}{T} \end{cases} \quad (14)$$

where $c_1(t)$ decreases linearly from the initial value c_{11} to the final value c_{12} , while $c_2(t)$ increases linearly from c_{21} to c_{22} .

3. Verification and Discussion

Based on the above mentioned SVM-diagnosis model, optimized using PSO, a code is made in MATLAB in which SVM algorithm is implemented directly by MATLAB toolkit [15]. Some real TFD examples are analyzed.

3.1. TFD Example 1

We analyze the dissolved-gas data for the existing 157 groups of transformers under normal and other fault conditions. The dissolved-gas data were detected from 6 types of real transformer faults: low-energy discharge fault (LE-D), high-energy discharge fault (HE-D), high temperature overheat fault (HT), medium temperature overheating fault (MT), medium and low temperature overheating failure (ML-T), and low temperature overheating fault (LT). 112 groups of data were selected as training samples, and the remaining 45 groups were used for testing. The distribution of the various faults and normal state samples are shown in Table 1.

Table 1. Statistics of samples for training and testing, corresponding to various types of real faults.

Fault Type	Training Sample	Test Sample	Total
Normal	17	7	24
LE-D	23	10	33
HE-D	20	8	28
HT	23	10	33
MT	7	2	9
ML-T	13	5	18
LT	9	3	12
Total	112	45	157

In this analysis, the particle swarm number is 20, the maximum iteration number is 200, and the search intervals for parameters C and σ are [0.01, 1000] and [0.01, 1000], respectively. Furthermore, $C = 15.8823$ and $\sigma = 50.1658$.

The fault diagnostic results of both the training set and the testing set are shown in Figure 2. Only one group of samples shows diagnosis errors among 112 groups of training samples (LE-D fault is diagnosed as HT fault), while 3 out of 45 samples yield diagnosis errors which are marked by circles. The accuracy reaches 93.33%. Diagnostic errors are either normal (diagnosed as HE-D), or LE-D

fault (diagnosed as HE-D), or MT fault (diagnosed as HE-D). The results indicate that this method is relatively accurate and capable of realizing the aim of TFD.

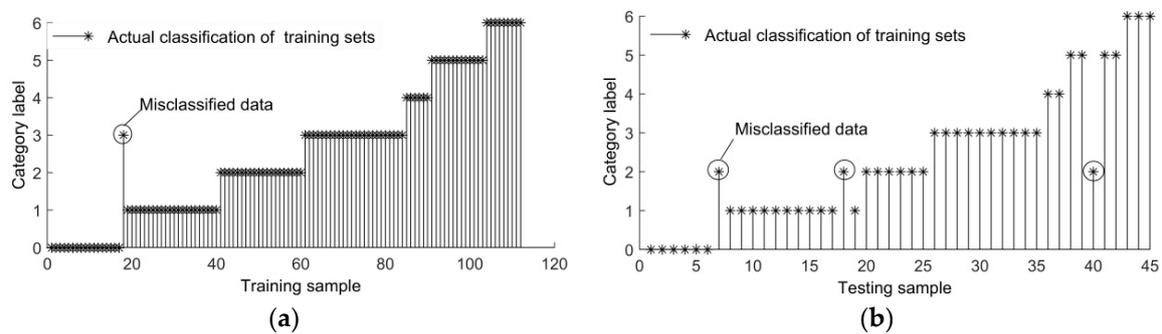


Figure 2. Results of the transformer-failure diagnosis: (a) training sets; (b) testing sets.

We adopt the three-ratio method, Duval triangle method, back propagation neural network (BPNN), and SVM methods to diagnose the testing data set for comparison. The same set of data was used for all methods. During the test, BPNN selected a network structure with 13 hidden nodes.

Table 2 shows the fault-diagnosis accuracy for different methods, when testing the same sample of transformer. The Duval triangle method shows the lowest accuracy. The three ratio method's accuracy is better than The Duval triangle method, however, worse than other methods. Both of three-ratio and Duval triangle methods are obtained from typical accidents, and they will fail when dealing with some complicated faults. The accuracy of the neural-network algorithm (BPNN) is 60% and it will be improved if there are a lot of data. Compared with the BPNN and IEC methods, the SVM method shows a relatively good diagnosis. When the SVM parameters are optimized, the accuracy of the fault diagnosis improves substantially.

Table 2. Accuracy rate for the different diagnostic methods of transformer.

Method	Three-Ratio	Duval Triangle	BPNN	SVM	This Paper
Accuracy rate	51.111%	42.222%	60.000%	75.556%	93.333%

3.2. TFD Example 2

This section uses SVM optimized by PSO to analyze the fault and normal states from the 132 groups of data detected from real transformers. The data were from the oil-dissolved gas and SCADA. We also verified the impact of data richness on the results. The dissolved gases in the oil include C_2H_2 , C_2H_4 , C_2H_6 , CH_4 , CO , CO_2 , H_2 and total hydrocarbon. The SCADA data include maximum current, minimum current, average current, maximum active power, minimum active power, average active power, maximum reactive power, minimum reactive power, and average reactive power. SVM optimized by PSO is used to diagnose the faults for three kinds of data: using only the dissolved gas data in oil, using only the SCADA data, and using all data. We used 112 groups as the training set and 20 groups as the testing set, and then judged the effect of data types on fault diagnosis.

In this experiment, the number for the particle swarm is 20, the maximum iteration number is 200, and the search interval of parameters C and σ are $[0.01, 1000]$ and $[0.01, 1000]$, respectively. The optimized parameter values and accuracy rate of different data types are shown in Table 3.

Table 3. Fault-diagnostic results of transformer of different methods.

Types	Data	C	σ	Accuracy Rate
1	Dissolved gas data in oil only	5.023	0.709931	80%
2	SCADA data only	26.5631	7.3787	65%
3	Both above data	36.6918	0.074581	95%

For TFD results using only the dissolved gas in the oil, four out of 20 groups of normal data are misdiagnosed as fault state. However, the fault state is diagnosed correctly, and so the accuracy can reach 80% in general. The TFD results of using only SCADA data, as shown in Table 3, has many errors in the non-fault diagnosis. 13 out of 20 groups were correctly evaluated, and the accuracy is as high as 65%. For the TFD results using the dissolved gas in oil and the SCADA data as input data of SVM optimized by PSO, only 1 out of 20 groups of test data is erroneous (the non-fault condition is misdiagnosed as fault state). Hence, the accuracy of the fault diagnosis is 95%.

Comparing the diagnostic results for the above three data types, type 1 and type 2 can be used to diagnose most of the transformer states, but there errors are different, when diagnosing non-fault states. This is because each data type can reflect the fault states of the transformer to some extent, including part of the transformers' state changes—but not completely. By combining these two kinds of data, more state information can be considered. The SVM optimized by PSO can learn the rich information from the data, thus improving the accuracy of the transformer-fault diagnosis.

4. Conclusions

A new fault-diagnosis technology for transformers, SVM optimized using PSO, was proposed in this paper, and it fully combines the advantages of the SVM and PSO. We used the dissolved-gas data in oil as the characteristic quantity and compared it with traditional methods as three-ratio, Duval triangle, and the artificial intelligence and machine learning algorithms as BPNN and SVM. The results showed that the proposed technology greatly improves the accuracy of the fault diagnosis. We also analyzed the effect of different data types of transformer on the TFD results. The data types included the dissolved-gas data in the transformer oil only, the SCADA data only, and both of them. The combination of the dissolved-gas data and SCADA data can improve the accuracy of the new technology substantially. This also testifies the importance of having sufficient data to perform an accurate transformer fault diagnosis.

The proposed method can be applied to prognostics and health management system in smart substations. It is noted that, if the amount of fault data is very big, it will take a long time to perform the fault diagnosis and it is inappropriate for on-line diagnosis; if the amount of fault data is not enough, its accuracy will be limited. The balance between the data amount and accuracy in its real applications needs further discussion.

Author Contributions: Conceptualization, K.L., J.S., J.W. and L.X.; Methodology, K.L., J.W., J.S. and L.X.; Software, K.L.; Validation, K.L. and J.W.; Data resources, L.X.; Writing—original draft preparation, K.L.; Supervision, J.W.; Project administration, J.S. and J.W.; All authors have read and approved the final manuscript.

Funding: This work was supported in part by National Natural Science Foundation of China, grant number 51207004.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhao, Z.Y.; Tang, C.; Zhou, Q. Identification of Power Transformer Winding Mechanical Fault Types Based on Online IFRA by Support Vector Machine. *Energies* **2017**, *10*, 2022. [[CrossRef](#)]
2. Kari, T.; Gao, W.S.; Zhao, D.B.; Zhang, Z.W.; Mo, W.X. An Integrated Method of ANFIS and Dempster-Shafer Theory for Fault Diagnosis of Power Transformer. *IEEE Trans. Dielectr. Electr.* **2018**, *25*, 360–371. [[CrossRef](#)]
3. Wang, Y.Y.; Gong, S.L.; Grzybowski, S. Reliability evaluation method for oil-paper insulation in power transformers. *Energies* **2011**, *4*, 1362–1375. [[CrossRef](#)]
4. Tenbohlen, S.; Coenen, S.; Djamali, M.; Müller, A.; Samimi, M.H.; Siegel, M. Diagnostic measurements for power transformers. *Energies* **2016**, *9*, 347. [[CrossRef](#)]
5. Zheng, H.B.; Liao, R.J.; Grzybowski, S.; Yang, L.J. Fault diagnosis of power transformers using multi-class least square support vector machines classifiers with particle swarm optimization. *IET Electr. Power Appl.* **2011**, *5*, 691–696. [[CrossRef](#)]

6. Zhao, A.X.; Zhang, C.T. DGA fault diagnosis based on the counter propagation neural network optimized by parallel genetic algorithm. In Proceedings of the 2013 IEEE International Conference of IEEE Region 10 (TENCON 2013), Xi'an, China, 22–25 October 2013; pp. 1–5.
7. Ghoneim, S.S.M.; Taha, I.B.M.; Elkalashy, N.I. Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis. *IEEE Trans. Dielectr. Electr. Insul.* **2016**, *23*, 1838–1845. [[CrossRef](#)]
8. Jalbert, J.; Lessard, M.C.; Ryadi, M. Cellulose chemical markers in transformer oil insulation part 1: Temperature correction factors. *IEEE Trans. Dielectr. Electr. Insul.* **2013**, *20*, 2287–2291. [[CrossRef](#)]
9. Arispe, J.C.G.; Mombello, E.E. Power transformer diagnosis using FRA and Fuzzy Sets. *IEEE Latin Am. Trans.* **2015**, *13*, 2991–2997. [[CrossRef](#)]
10. Seifeddine, S.; Khmais, B.; Abdelkader, C. Power transformer fault diagnosis based on dissolved gas analysis by artificial neural network. In Proceedings of the 2012 First International Conference on Renewable Energies and Vehicular Technology, Hammamet, Tunisia, 26–28 March 2012; pp. 230–236.
11. Xiong, H.; Sun, C.X. Artificial immune network classification algorithm for fault diagnosis of power transformer. *IEEE Trans. Power Deliv.* **2006**, *22*, 930–935.
12. Li, S.; Li, X.Y.; Wang, W.X. Fault Diagnosis of Transformer Based on Probabilistic Neural Network. In Proceedings of the Fourth International Conference on Intelligent Computation Technology and Automation, Shenzhen, China, 28–29 March 2011; Volume 1, pp. 128–131.
13. Su, P.; Yuan, J.S.; An, X.L.; Li, Z.; Shen, T. Diagnosis method of transformer faults based on rough set theory. *Electr. Power Sci. Eng.* **2008**, *24*, 56–59.
14. Bigdeli, M.; Vakilian, M.; Rahimpour, E. Transformer winding faults classification based on transfer function analysis by support vector machine. *IET Electr. Power Appl.* **2012**, *6*, 268–276. [[CrossRef](#)]
15. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).