

## Research Article

# A Combined Fault Diagnosis Method for Power Transformer in Big Data Environment

Yan Wang<sup>1</sup> and Liguozhang<sup>2</sup>

<sup>1</sup>Department of Computer, North China Electric Power University, Baoding, China

<sup>2</sup>College of Information Science & Technology, Agricultural University of Hebei, Baoding, China

Correspondence should be addressed to Yan Wang; wangyan1206@126.com

Received 8 December 2016; Accepted 27 March 2017; Published 18 May 2017

Academic Editor: Yaguo Lei

Copyright © 2017 Yan Wang and Liguozhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The fault diagnosis method based on dissolved gas analysis (DGA) is of great significance to detect the potential faults of the transformer and improve the security of the power system. The DGA data of transformer in smart grid have the characteristics of large quantity, multiple types, and low value density. In view of DGA big data's characteristics, the paper first proposes a new combined fault diagnosis method for transformer, in which a variety of fault diagnosis models are used to make a preliminary diagnosis, and then the support vector machine is used to make the second diagnosis. The method adopts the intelligent complementary and blending thought, which overcomes the shortcomings of single diagnosis model in transformer fault diagnosis, and improves the diagnostic accuracy and the scope of application of the model. Then, the training and deployment strategy of the combined diagnosis model is designed based on Storm and Spark platform, which provides a solution for the transformer fault diagnosis in big data environment.

## 1. Introduction

With the continuous development of smart grid, the online monitoring of power transformers has been greatly developed and the monitoring data of transformer has increased exponentially. The online monitoring data of transformer have the characteristics of large quantity, multiple types, and low value density (a large number of data is related to transformer's normal state, while the fault data for fault diagnosis is less), so it is suitable for using the storage and processing technology of big data [1, 2]. The key to fault diagnosis of transformer in big data is how to build the fault diagnosis model and the training and deployment of the fault diagnosis model in the large data processing platform.

It is simple for transformer to diagnose fault based on the dissolved gas analysis (DGA), which has been proved by a large number of research results. Based on DGA characteristic parameters, many kinds of artificial intelligence technology, including fuzzy method [3–5], support vector machine (SVM) [6–9], artificial neural network [10–13], Bayesian network [14, 15], gene expression programming [16–18], expert

system [5, 19], Dempster-Shafer evidential theory [20], and association rule mining [21–23], are adopted to the study of fault diagnosis system for power transformer. Deep learning, which simulates the hierarchical structure of human brain, processing data from lower level to higher level and gradually composing more and more semantic concepts, has been popular in artificial intelligence. Some scholars [24–27] introduce deep autoencoder network and deep belief network to solve the problem of transformer fault diagnosis. All of the above methods have made some fault diagnosis achievements.

However, as mentioned above, the problem is that the samples in the normal state are many, while the fault samples are scarce, which brought difficulties for the promotion of the fault diagnosis method based on DGA. According to the research results, each intelligent diagnosis method has its advantages and its disadvantages. In view of this, to find a fault diagnosis model which can make full use of various intelligent methods' advantages and can also adapt to the characteristics of transformer monitoring big data is the key to solve the problem of transformer fault diagnosis. According to the combination theorem, we can make the

best use of all the information through the combination of various methods, which is helpful to improve the diagnostic performance of the system. Combined diagnosis usually has two methods; one method is to select the best fitting model as the diagnosis model by comparing several diagnostic models and the second method is to calculate the final diagnosis result by selecting the appropriate weights for the diagnosis results of several models. The first method requires a large amount of manual work, so it should be discarded. The second method can make full use of not only the different information of transformer but also the advantages of various diagnostic methods, so we adopt it in this paper.

The transformer fault diagnosis method based on factor analysis and gene expression programming (abbreviated to the FA-GEP model in the rest of this paper) has the advantages that transformer running state contained in the transformer oil is fully utilized, the dimension of feature variables is effectively reduced, and it has a higher diagnosis speed and accuracy. However, its accuracy depends on a large number of sample data, especially the fault cases [18]. Even worse, in the initial stage of fault diagnosis system, the fault cases is limited and the cases of each fault type are not balanced, which restricts the further improvement of the algorithm. In the case of no samples, the transformer fault diagnosis model based on cloud model and matter element theory can be built by expanding the traditional fault diagnosis methods such as three-ratio method [28]. This method is helpful to improve the fault diagnosis for the transformer with fewer samples. At the same time, although this method overcomes the inherent defects of the ratio method where the code is not complete, it still established on the basis of experience summarization, which limits the positive rate of fault diagnosis. Support vector machine is a machine learning algorithm; it finds the solution on the basis of statistical learning theory to make the sum of the experience risk and confidence range least. Because the goal is to obtain the optimal solution under the existing information, it has a special advantage in the small sample learning. In view of this, SVM has been promoted in the transformer fault diagnosis to a certain extent [6–8], but when the number of samples increases, the diagnostic accuracy of SVM is not significantly improved.

Therefore, the diagnosis methods are combined to make best use of the advantages and bypass the disadvantages in this paper. We establish a combination fault diagnosis model for power transformer in order to adapt to the characteristics of transformer monitoring data and to improve the fault diagnosis accuracy. In the combined diagnosis method, at first the DGA data are initially by each model and then the various diagnostic results are diagnosed by SVM to obtain a more comprehensive diagnostic result. This method makes full use of the advantages of each diagnostic model and can effectively improve the accuracy and efficiency of fault diagnosis.

Another difficult problem of fault diagnosis is the deployment of fault diagnosis model in big data environment. Storm and Spark are proposed for the real-time processing environment for big data in recent years [29], Storm is mainly for real-time, high concurrent data stream, and Spark's

memory computing model makes it more suitable for the iterative calculation of big data. Based on Storm and Spark platform, this paper designs the deployment and training structure of the fault diagnosis model, which provides a solution to the problem of transformer fault diagnosis in big data environment.

## 2. Combination Diagnosis Based on SVM

*2.1. Combination Diagnosis Model.* Based on the above diagnosis thought, a new fault diagnosis method for power transformer is presented in this paper. Firstly the FA-GEP diagnosis model, the SVM diagnosis model, and cloud matter element diagnosis model are combined into diagnosis model group, and then the new samples including fault diagnosis results of each method and the original DGA data are diagnosed by SVM which is suitable for small sample learning. The transformer fault combination diagnosis model based on SVM is shown in Figure 1.

The multiple weak classifiers are combined to improve the classification accuracy, which is similar to the idea of Boosting. Its specific mathematical formula is shown in formula (1).

$$y'_i = F(\omega_k, f_{ki}) = \sum_{k=1}^m \omega_k f_{ki}, \quad i = 1, 2, \dots, n. \quad (1)$$

Here,  $w_k$  is the combined weight of  $k$ th model and meets  $\sum_k w_k = 1$ . It can be obtained by the Lagrange multiplier method or other mathematical programming method.  $y'_i$  is the combination diagnosis result.  $f_{ki}$  is the diagnosis result of the  $k$ th model in  $i$ th times diagnosis.

The original oil chromatographic data and the diagnostic results of each individual diagnosis model on this training sample are the inputs of SVM combination diagnosis model and the actual fault type of the training sample as the output value of the sample; that is, the output value of the SVM should be obtained. By learning from a large number of samples, the parameters of SVM are optimized to achieve the purpose of correct diagnosis. Therefore, the weight of each diagnostic model in this combination model is implied in the second SVM model, and it can get a higher diagnostic accuracy compared with artificial setting  $w_k$ , because it is a nonlinear combined diagnosis.

*2.2. SVM and Its Kernel Function Selection.* On the basis of linear classifier, support vector machine is formed by introducing structural risk minimization principle, optimization theory, and kernel function. The classification principle can be summarized as follows: it searches for classified hyperplane, which makes the two types of samples in the training sample separated, and the distance from the plane as far as possible; for the linearly nonseparable problem, the low dimensional space is mapped onto the high dimension space by kernel function, which can transform the linear nonseparable problem of the original low dimensional space into a linear separable problem in high dimension space.

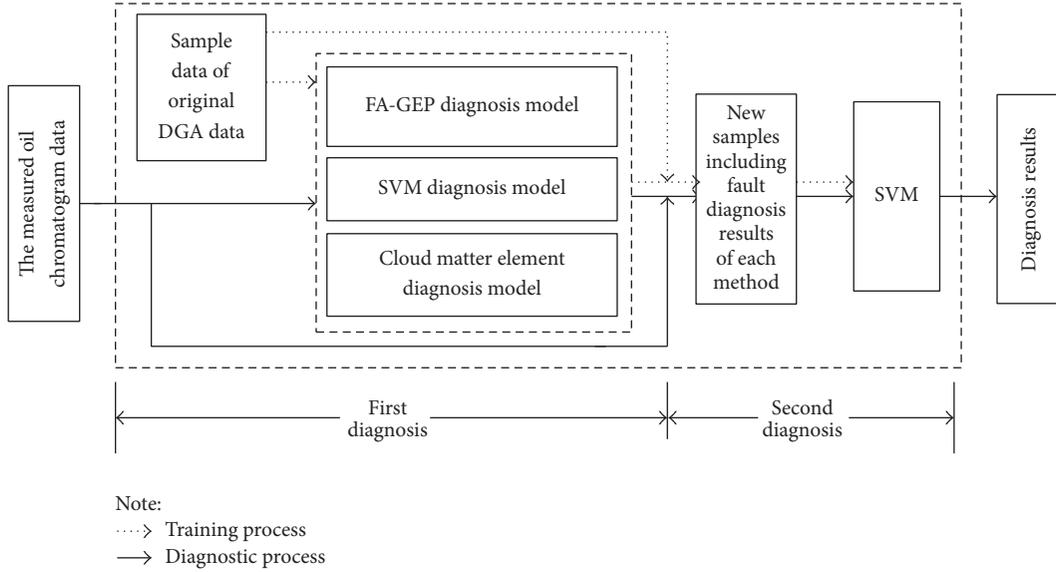


FIGURE 1: Combination diagnosis models for transformer.

The selection of SVM's kernel function has great influence on the accuracy of diagnosis model, and different kernel function can be used to construct the learning machine for different types of nonlinear decision surfaces in the input space.

Common kernel functions include linear kernel function:  $K(x, y) = x \cdot y$ , Gauss RBF kernel function:  $K(x, y) = \exp(-\gamma \|x - y\|^2)$ , and polynomial kernel function:  $K(x, y) = [(x \cdot y) + 1]^q$ .

The study found that, compared to other kernel functions, the RBF core has a relatively good performance [30]. Therefore, this paper will use the RBF kernel function for SVM modeling, as shown in formula (2).

$$K(x, y) = \exp(-\gamma \|x - y\|^2). \quad (2)$$

### 3. Training and Deployment of Combined Fault Diagnosis Model

The implementation of fault diagnosis model is divided into two parts, training and deployment. The former is done by the Spark platform, which is good at iterative computing and machine learning, while the Storm platform is responsible for transformer's fault diagnosis based on the training model, as shown in Figure 2.

On the Storm platform, the trained fault diagnosis model is deployed to the parallel fault diagnosis of abnormal data. The essence of the deployment model is to design a set of data processing logic structure according to the business requirements. It is called Topology. When the monitoring parameters from the transformer monitoring module flow into the system, first feature extraction Bolt is performed to extract features, then three different classifier Bolts are triggered to carry out the concurrent diagnosis for DGA extract features, and at last the diagnostic results are sent to the combined diagnosis Bolt to get the final diagnosis result.

In the Spark platform, the sample data of historical cases are used to initialize the model of fault diagnosis. When a diagnosis on the Storm platform is complete, the input feature vector and the output result will form a new sample. The new samples will be added to the sample set and will be converted into RDD (Resilient Distributed Datasets) on the Spark platform, which can be used for the parallel update learning of the fault diagnosis classifier.

Set  $TS_i = \{x_1, \dots, x_m, S_i\}$ , in which  $TS_i$  represents the  $i$ th sample of the training sample data,  $x_m$  is the value of the  $m$ th attribute variable of the sample, and  $S_i$  indicates the actual state of transformer. The specific training process is as follows.

*Step 1.* Establish the diagnostic model group: FA-GEP diagnostic model and SVM model are generated by training the first sample data; matter element model is established based on IEC three-ratio method.

*Step 2.* Input the training data  $TS_i$  into FA-GEP diagnosis model, SVM diagnosis model, and the cloud matter element diagnosis model separately, and then get the initial diagnosis results  $F_{1i} = (f_{i1}, f_{i2}, \dots, f_{in})$ . Here,  $n$  is the number of models in the diagnosis model group,  $f_{in}$  is the diagnostic result of the  $n$ th diagnosis model on the  $i$ th sample data, and the value range of  $f_{in}$  is the same as  $S_i$ .

*Step 3.* The original oil chromatographic data, its corresponding initial diagnosis results, and the actual state of power transformer are organized into the second training samples  $TS'_i = \{f_{i1}, \dots, f_{in}, x_1, \dots, x_m, S_i\}$ .

*Step 4.* As shown in Figure 2, on the basis of the second training samples  $TS'_i$  train the secondary SVM model and adjust the relevant parameters to obtain the corresponding output value that is tested by practice. In this way, we can train the system to realize the nonlinear combination diagnosis, in which the combination weight is implicit in the training model.

TABLE 1: Comparison of fault diagnosis results.

	Combination model	FA-GEP	Cloud matter element	SVM	BP
Number of correct diagnoses	85	82	78	80	81
Positive rate	94.4%	91.1%	86.7%	88.9%	90%

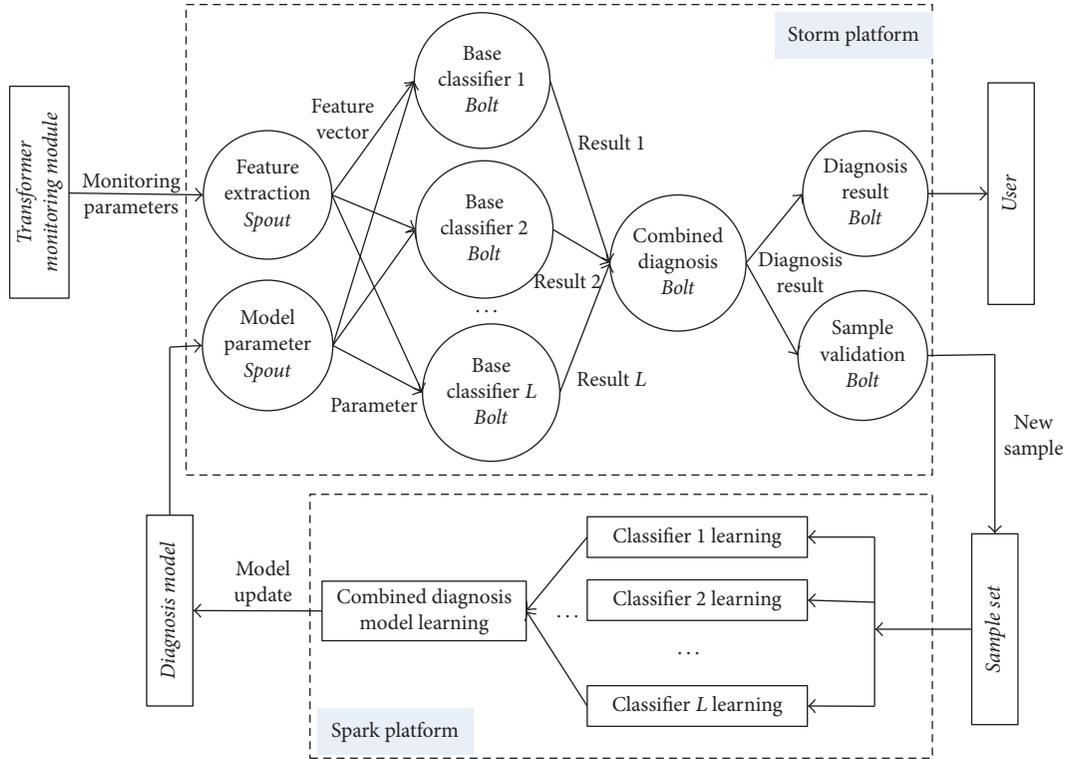


FIGURE 2: Training and deployment of combined fault diagnosis model.

## 4. Experiment and Case Analysis

**4.1. Experiment Analysis.** From the recent relevant literatures and the transformer oil chromatographic detection records investigated from multiple power supply companies, we collected 395 DGA samples with a clear conclusion, including the normal data of 306 and the fault data of 89. Taking the content of  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ , and  $C_2H_2$  in oil as attribute information, the transformer is divided into five states including normal (S1), low energy discharge (S2), high energy discharge (S3), low temperature overheating (S4) (less than  $700^\circ C$ ), and high temperature overheating (S5) ( $>700^\circ C$ ). 305 samples were randomly selected as training samples and the remaining 90 samples as test samples. After training, the proposed model is used to diagnose the test samples, and the diagnostic accuracy and the comparison with other models are shown in Table 1. As can be seen from Table 1, the diagnostic accuracy of the combined model is higher than that of other methods.

### 4.2. Case Analysis

(1) *Case 1.* The 4# power transformer SPS-360000/500 of a plant was put into operation in 1994 and the oil chromatographic analysis results in 1996 are shown in Table 2 [31].

TABLE 2: Oil chromatographic data of some transformer ( $\mu L/L$ ).

$H_2$	$CH_4$	$C_2H_6$	$C_2H_4$	$C_2H_2$
43.1	19.4	8.4	22.7	0

The analysis believes that the total hydrocarbon content in transformer oil exceeds the standard, and the code of IEC three ratios is 001, so it is identified as the low temperature overheating fault. A comprehensive examination of the transformer finds that the connecting Bolt tablets have the overheated traces, and tablet has pitting.

The conclusion of the combined diagnosis model is the high temperature overheating, which is in accord with the actual situation. The diagnostic results of the three single diagnostic models are high temperature overheating (FA-GEP model), high temperature overheating (SVM model), and high energy discharge (matter element model).

(2) *Case 2.* The oil chromatographic analysis results of Beijing power plant's 2# main transformer power transformer (220 kV) in 1990-10-4 are shown in Table 3 [32]. The code of IEC three ratios is 101, so it is identified as the low energy discharge fault.

TABLE 3: Oil chromatographic data of 2# transformer ( $\mu\text{L/L}$ ).

$\text{H}_2$	$\text{CH}_4$	$\text{C}_2\text{H}_6$	$\text{C}_2\text{H}_4$	$\text{C}_2\text{H}_2$
93	17	6.2	17	23.4

The conclusion of the combined diagnosis model is high energy discharge. The diagnostic results of the three single diagnostic models are high energy discharge (FA-GEP model), low energy discharge (SVM model), and high energy discharge (matter element model). After the actual test, it is found that there are obvious signs of arc discharge, which is consistent with the diagnostic result.

## 5. Conclusions

Aiming at the characteristics of large data of transformer monitoring, this paper puts forward a new fault diagnosis method for transformer, in which a variety of fault diagnosis models are used to make a preliminary diagnosis, and then the support vector machine is used to make the second diagnosis. On the basis of intelligent complementary thought, the FA-GEP diagnosis model applicable for a large number of samples, SVM diagnostic model suitable for small sample, and the cloud matter element model with no sample to be created are combined as the diagnosis model, which effectively improves the diagnostic accuracy and the scope of application. Then, based on Storm and Spark platform, the training and deployment architecture of fault diagnosis model is designed in big data processing environment, which provides a solution for the fault diagnosis of transformer in big data environment.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

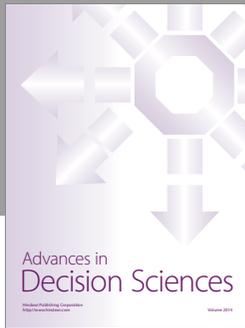
## Acknowledgments

This work was supported by the National Natural Science Fund (51677072) and the Fundamental Research Funds for the Central Universities (2014MS132).

## References

- [1] W. Alves, D. Martins, U. Bezerra, and A. Klautau, "A Hybrid approach for big data outlier detection from electric power SCADA system," *IEEE Latin America Transactions*, vol. 15, no. 1, pp. 57–64, 2017.
- [2] J. Lang, S. Pascoe, J. Thompson, J. Woyak, K. Rahimi, and R. Broadwater, "Smart grid big data: automating analysis of distribution systems," in *Proceedings of the 60th Annual IEEE Rural Electric Power Conference, (REPC '16)*, pp. 96–101, USA, May 2016.
- [3] A. Abu-Siada and S. Hmood, "A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis," *International Journal of Electrical Power and Energy Systems*, vol. 67, pp. 401–408, 2015.
- [4] R. Liao, H. Zheng, S. Grzybowski, L. Yang, Y. Zhang, and Y. Liao, "An integrated decision-making model for condition assessment of power transformers using fuzzy approach and evidential reasoning," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 1111–1118, 2011.
- [5] P. Yuyun, *Research on Intelligent Fault Diagnosis System for transformer based on DGA Technology [Ph.D. thesis]*, Wuhan University, China, 2004.
- [6] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," *Electric Power Systems Research*, vol. 83, no. 1, pp. 73–79, 2012.
- [7] L. V. Ganyun, C. Haozhong, Z. Haibao et al., "Fault diagnosis of power transformer based on multi-layer SVM classifier," *Electric Power Systems Research*, vol. 74, no. 1, pp. 1–7, 2005.
- [8] Z. Jianbai, *The application research of Support Vector Machines in transformer fault diagnosis [MA. Thesis]*, North China Electric Power University, 2007.
- [9] C. Wei, W. Tang, and Q. Wu, "Dissolved gas analysis method based on novel feature prioritisation and support vector machine," *IET Electric Power Applications*, vol. 8, no. 8, pp. 320–328, 2014.
- [10] H. A. Illias, X. R. Chai, A. H. A. 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, Article ID e0129363, 2015.
- [11] G. Jun and H. Junjia, "Application of quantum genetic ANNs in transformer dissolved gas-in-oil analysis," in *Proceedings of the CSEE*, vol. 30, pp. 121–127, 2010.
- [12] H. Malik, A. K. Yadav, S. Mishra, and T. Mehto, "Application of neuro-fuzzy scheme to investigate the winding insulation paper deterioration in oil-immersed power transformer," *International Journal of Electrical Power & Energy Systems*, vol. 53, pp. 256–271, 2013.
- [13] D. Bhalla, R. K. Bansal, and H. O. Gupta, "Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 1196–1203, 2012.
- [14] W. H. Tang, Z. Lu, and Q. H. Wu, "A bayesian network approach to power system asset management for transformer dissolved gas analysis," in *Proceedings of 3rd International Conference on Deregulation and Restructuring and Power Technologies, (DRPT '08)*, pp. 1460–1466, April 2008.
- [15] W. H. Tang, J. Y. Goulermas, Q. H. Wu, Z. J. Richardson, and J. Fitch, "A probabilistic classifier for transformer dissolved gas analysis with a particle swarm optimizer," *IEEE Transactions on Power Delivery*, vol. 23, no. 2, pp. 751–759, 2008.
- [16] A. Shintemirov, W. Tang, and Q. H. Wu, "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*, vol. 39, no. 1, pp. 69–79, 2009.
- [17] A. Abu-Siada and S. Islam, "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*, vol. 19, no. 3, pp. 1007–1012, 2012.
- [18] Z. Wu, Z. Dong, G. Yu, and Y. Zhu, "Transformer fault diagnosis based on factor analysis and gene expression programming," *Journal of North China Electric Power University*, vol. 39, no. 4, pp. 47–51, 2012.

- [19] N. Bakar, A. Abu-Siada, and S. Islam, "A review of dissolved gas analysis measurement and interpretation techniques," *IEEE Electrical Insulation Magazine*, vol. 30, no. 3, pp. 39–49, 2014.
- [20] D. Bhalla, R. K. Bansal, and H. O. Gupta, "Integrating AI based DGA fault diagnosis using dempster-shafer theory," *International Journal of Electrical Power & Energy Systems*, vol. 48, no. 1, pp. 31–38, 2013.
- [21] Z. Yang, W. H. Tang, A. Shintemirov, and Q. H. Wu, "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*, vol. 39, no. 6, pp. 597–610, 2009.
- [22] D.-E. A. Mansour, "Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 5, pp. 2507–2512, 2015.
- [23] W. Zuo, H. Yuan, Y. Shang, Y. Liu, and T. Chen, "Calculation of a health index of oil-paper transformers insulation with binary logistic regression," *Mathematical Problems in Engineering*, vol. 2016, Article ID 6069784, 2016.
- [24] X. Shi, Y. Zhu, X. Ning, L. Wang, G. Sun, and G. Chen, "Transformer fault diagnosis based on deep auto-encoder network," *Electric Power Automation Equipment*, vol. 36, no. 5, pp. 122–126, 2016.
- [25] X. Shi, Y. Zhu, Sa C., Wang L., and Sun G., "Power transformer fault classifying model based on deep belief network," *Power System Protection and Control*, vol. 44, no. 1, pp. 71–76, 2016.
- [26] P. Tamilselvan and P. F. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering & System Safety*, vol. 115, pp. 124–135, 2013.
- [27] Q. M. Cui, K. Liang, H. Y. Gao et al., "Research of the transformer fault diagnosis expert system based on esta and deep learning neural network programmed in MATLAB," in *Proceedings of the 2016 International Conference on Civil, Transportation and Environment*, pp. 772–778, 2016.
- [28] Z.-L. Wu, N. Yuan, Z. Gong, and Y.-L. Zhu, "Improved matter-element model based on cloud model and its application in power transformers fault diagnosis," in *Proceedings of the International Conference on Electronics, Communications and Control, ICECC 2011*, pp. 3758–3762, September 2011.
- [29] "Spark:Choose your real-time weapon [EB/OL]," 2014 <http://www.infoworld.com/article/2854894/application-development/spark-and-storm-for-real-time-computation.html>.
- [30] L. Maoliu and C. Chunyu, "A performance comparison of SVMs based on Fourier kernel and RBF kernel," *Journal of Chongqing University of Posts and Telecommunications (Natural Science)*, vol. 17, no. 6, pp. 647–650, 2005.
- [31] M. Shunan, *Research on Transformer Remote Faults Diagnosis Based on Support Vector Machine [MA. Thesis]*, Beijing Jiaotong University, 2010.
- [32] W. Youyuan, *Study on Prediction Models of Power Transformer Fault Based on Genetic Algorithm [MA. Thesis]*, Chongqing University, 2003.



# Hindawi

Submit your manuscripts at  
<https://www.hindawi.com>

