INTERNATIONAL TRANSACTIONS ON ELECTRICAL ENERGY SYSTEMS
Int. Trans. Electr. Energ. Syst. 2015; **25**:2247–2259
Published online 21 June 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/etep.1959

Pattern recognition techniques for power transformer insulation diagnosis—a comparative study part 1: framework, literature, and illustration

Yi Cui, Hui Ma*,† and Tapan Saha

University of Queensland, Brisbane, Australia

SUMMARY

The condition of the insulation system of a power transformer has a significant impact on its overall reliability and serviceability. Transformer oil tests including breakdown voltage, acidity, dielectric dissipation factor, 2-furfuraldehyde, water content, and dissolved gases analysis have been commonly performed in utility companies to provide information regarding the conditions of transformer insulation. Over the past two decades, various pattern recognition techniques are proposed to interpret the oil tests results and make diagnosis on transformer insulation. However, there are still considerable challenging issues to be investigated before the pattern recognition technique can become a "ready-to-use" tool at utility companies. This paper provides a comparative study of pattern recognition techniques for power transformer insulation diagnosis using oil tests results. A general pattern recognition application framework will be outlined in the paper. And a comprehensive literature review on various pattern recognition techniques for transformer insulation diagnosis will be provided in the paper. The important issues for improving the applicability of pattern recognition techniques for transformer insulation diagnosis will also be discussed. A case study will be presented to demonstrate the procedure of applying pattern recognition techniques to practical transformer insulation diagnosis using oil test results. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS: dissolved gas analysis; insulation; oil characteristics; pattern recognition; power transformer

1. INTRODUCTION

Power transformer is one of the most crucial equipment in an electricity grid. Its serviceability has a significant influence on the reliable delivery of electricity. However, a power transformer's insulation can be eventually deteriorated since the transformer is continuously subjected to electrical, mechanical, and thermal stresses. Such insulation deterioration may lead to a disruptive failure of a transformer [1–3]. Therefore, a variety of techniques have been developed for transformer insulation diagnosis, including: (i) oil tests such as breakdown voltage (BDV), acidity, dielectric dissipation factor (DDF), resistivity, 2-furfuraldehyde, water content, and dissolved gas analysis (DGA); (ii) dielectric response measurement consisting of polarization and depolarization current measurement and frequency dielectric spectroscopy; (iii) frequency response analysis; and (iv) partial discharge (PD) measurement [4–8].

Among all of the above techniques, transformer oil tests have been commonly adopted by utilities to reveal various chemical and physical properties of insulating oil. Through the measurement of 2-furfuraldehyde and carbon oxides, oil tests may also provide some indications of the condition of the solid (pressboard and paper) insulation of transformer. In the past 20 years, various interpretation methods have been proposed for analyzing oil tests results (oil characteristics) and

*Correspondence to: Hui Ma, School of Information Technology and Electrical Engineering, University of Queensland, Brisbane 4072, Australia.
†E-mail: huima@itee.uq.edu.au

detecting discharge and overheat faults occurring in a transformer insulation system [9]. However, there still exist some limitations of these conventional schemes. For example, the diagnosis results might be inconsistent by using different industry standards. Sometimes, the conventional schemes cannot produce the diagnosis results for every possible combination of dissolved gases' ratios [10].

To overcome the disadvantages of the conventional interpretation schemes, various pattern recognition techniques such as artificial neural networks (ANN) [11], fuzzy logic [12,13], neural-fuzzy system [14], and wavelet transforms [15] have been extensively investigated. These techniques have achieved some extent of success in transformer insulation diagnosis using oil characteristics. In contrast to several conventional interpretation schemes, pattern recognition techniques utilize not only the oil tests data obtained from the transformer of interest, but also the historical oil test datasets collected from other transformers. By using historic datasets, a pattern recognition algorithm learns the underlying relationship between oil characteristics and the transformers insulation, and it then applies such knowledge to make a diagnosis on the insulation condition of the transformer of interest.

Nevertheless, there are still considerable challenging issues to be investigated before the pattern recognition-based interpretation schemes can become a "ready-to-use" tool for transformer insulation diagnosis in utilities. This paper provides a comparative study of pattern recognition techniques and their applications in power transformer insulation diagnosis using oil characteristics. Starting with a brief review of oil tests and conventional interpretation schemes, this paper outlines the insulation diagnosis problem into a general pattern recognition framework. A comprehensive literature review on various pattern recognition algorithms for transformer insulation diagnosis is provided in this paper. The challenging issues of improving the applicability of pattern recognition techniques will also be discussed. A case study will be presented to demonstrate the procedure of applying pattern recognition techniques to practical transformer insulation diagnosis using oil characteristics. The detailed mathematic formulation, implementation, and statistical performance evaluation of 15 pattern recognition algorithms will be provided in the accompanying paper [16].

2. TRANSFORMER OIL TESTS

2.1. *Oil tests*

Because of continuously operating under various stresses, it is possible that oil molecular bonds partially crack and free particles can be generated. The interactions among these particles or between these particles and external molecules may form by-products including water contents, dissolved gases, acid components, and other types of contaminants in oil.

The transformer oil tests are able to detect the above by-products and subsequently can provide some insights regarding transformer insulation. DGA results can be useful to detect transformer incipient faults such as arcing, PD, and thermal fault. Water contents play a key role in the transformer insulation ageing. The large increase of water contents in transformer oil can decrease the resistivity and electrical strength of transformer oil. In the presence of solid contaminant and water contents, the reverse impacts of acids on the dielectric properties of transformer oil may become significant. The DDF can reveal the contaminant alteration in transformer oil. Moreover, the resistivity test measures the degree of losses of transformer oil, and the BDV indicates the dielectric strength of transformer oil.

2.2. *Conventional interpretation schemes*

The conventional interpretation schemes use the concentrations or the ratios of the dissolved gases to diagnose the transformer insulation faults. For example, IEC/IEEE and Rogers's schemes use three gas ratios of C_2H_2/C_2H_4, CH_4/H_2, and C_2H_4/C_2H_6, while Doernenburg method uses four gas ratios of CH_4/H_2, C_2H_2/C_2H_4, C_2H_2/CH_4, and C_2H_6/C_2H_2. The ratios in these schemes are then compared with the threshold values provided by the relevant industry standards to diagnose insulation condition. On the other hand, the Duval triangle method uses a relative portion of three dissolved gases in the forms

DOI: 10.1002/etep

of $\%C_2H_2 = x/(x+y+z)$, $\%C_2H_4 = y/(x+y+z)$, and $\%CH_4 = z/(x+y+z)$, in which x, y, and z denote the concentrations of C_2H_2, C_2H_4, and CH_4, respectively. It can reveal multiple faults (e.g. thermal and discharge faults) that may simultaneously occur in a transformer.

Instead of using the above conventional schemes, this paper investigates pattern recognition techniques and their applications for transformer insulation diagnosis using oil tests data (oil characteristics). After presenting a generic pattern recognition application framework for transformer insulation diagnosis, a comprehensive literature review on various pattern recognition techniques reported in the literature will be provided in the following sections.

3. PATTERN RECOGNITION FRAMEWORK FOR TRANSFORMER DIAGNOSIS

In pattern recognition, a computer algorithm is trained to distinguish the pattern of interest from the background and makes decisions on the category of this pattern.

Figure 1 depicts a generic pattern recognition application framework, which involves three steps: (i) data pre-processing, (ii) feature extraction, and (iii) classification.

3.1. Data pre-processing

The raw data might be inconsistent and noise corrupted. Redundancies may also occur due to the integration of data from different sources. The data pre-processing step applies various techniques to process the raw data including: data cleaning for dealing with noise and removing redundant data; data transformation for converting the raw data into appropriate forms; data reduction for eliminating unnecessary attributes; and data discretization for reducing the number of levels of an attribute of data.

3.2. Feature extraction

Feature extraction aims to find the characteristic attributes (features) from the original data. This will enable pattern recognition algorithms to focus on those most relevant features for fault classification. In transformer insulation diagnosis using oil characteristics, the features may include the concentrations of dissolved gases and oil quality tests results such as acidity, BDV, DDF, 2-furfuraldehyde, resistivity, and water content.

3.3. Classification

In the classification step, the pattern recognition algorithm makes use of a historic oil tests dataset to construct a mathematical model that approximates the relationship between the oil characteristics (e.g. features) and the categories of transformer insulation condition or the types of incipient faults of transformer. Then, the above model is used to make classification on the insulation condition or the incipient fault of the transformer of interest into one of the categories determined in the above historic dataset.

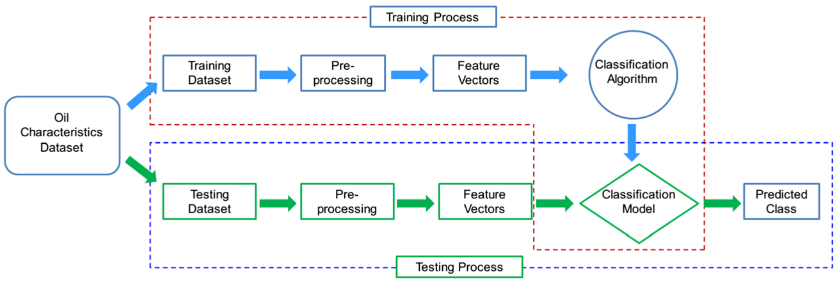

Figure 1. Framework of pattern recognition.

DOI: 10.1002/etep

4. LITERATURE REVIEW ON PATTERN RECOGNITION TECHNIQUES FOR TRANSFORMER INSULATION DIAGNOSIS

This section reviews various pattern recognition techniques for transformer insulation diagnosis. The techniques include ANN, fuzzy logic, expert system, decision-making algorithm, support vector machine (SVM), population-based algorithms, and hybrid algorithms.

4.1. ANN

ANN has been extensively investigated for transformer insulation diagnosis using oil characteristics [17–35]. During the training process, an ANN adjusts the weights between neurons and the thresholds of activation function of each neuron. In this way, the ANN constructs a model describing the dependency between the input features and the fault types. Such model is applied to classify any new oil characteristics data into one of the transformer insulation conditions defined in the training process.

Zhang *et al*. proposed a two-stage ANN to diagnose transformer insulation condition using DGA data [17]. By including CO_2 as one of input features, the first stage network can distinguish the faults related to paper insulation from other types of transformer inception faults. Tenfold cross validation was employed for finding the optimal number of neurons in the hidden layer of ANN. In [18], Vanegas *et al*. proposed a network with two sets of input features: the first set of features is the IEC gas ratios; and the second set is the concentrations of five dissolved gases. The authors demonstrated that the proposed network could attain better classification accuracy by using five gases concentrations as input features. In [19], Guardado *et al*. compared ANNs' efficiency for transformer insulation diagnosis. The neural network (NN) was trained based on the diagnosis criteria of five conventional interpretation schemes, i.e. Doernenburg, modified Rogers, Rogers, IEC, and CSUS (this diagnosis criterion uses the individual gas concentrations). The authors reported that a network having three layers with few neurons in the hidden layer could be suitable for transformer insulation incipient faults detection with overall accuracy above 87%. In [20], a multinomial logistic regression model and back-propagation (BP) NN were combined to determine the fault types of power transformers. The multinomial logistic regression model was applied to find the compositions of the dissolved gases that were correlated to specific types of faults. The concentrations of these gas compositions were fed into BP-NN for training. With such approach, the diagnosis performance can be improved. Some other types of ANNs for transformer insulation diagnosis reported in the literature are: self-organizing polynomial network, which heuristically formulated the problem into a hierarchical architecture having multiple layers of low-order polynomials [21]; and a reduced multivariate polynomial (RMP)-based network, in which the determination of its parameters only required a predefined RMP's order, and no iterative procedure is needed [22].

To improve the adaptation capability and accelerate the re-training process when new samples (data) become available, the probabilistic NN (PNN) was adopted in [23]. A fuzzy learning vector quantization (FLVQ) network was also proposed in [24]. The FLVQ network used a fuzzy classifier to segment historical DGA data into different categories of gas ratios. Then, for each category of gas ratio, a learning vector quantization network was trained for classifying the types of transformer incipient faults.

For the normal ANNs, they may have some difficulties in determining the network architecture and hidden layer neuron number. To tackle such limitations, the evolutionary algorithm was integrated into ANNs [25,26]. Such hybrid algorithms can calculate the optimal connection weights and bias terms of the networks simultaneously due to its excellent global search capability. The extension theory-based clustering algorithm was also proposed for detecting transformer incipient faults [27,28]. Instead of using the Euclidean distance, the extension distance was used in this method for measuring the similarity between different data points. One significant advantage of this method is that it does not require the tuning of any particular artificial parameters, and thus no learning process is needed.

To provide the data visualization capability, the self-organizing map (SOM)-based network was proposed in [29]. The SOM can also provide the visualization on the evolution of an incipient fault by plotting the DGA trajectories using the data collected throughout many years. Similar to SOM, a set of auto-associative NN were implemented to provide visualization and clustering in transformer insulation diagnosis [30]. The simple *k*-nearest neighbour" algorithm (*k*NN) and adaptive *k*NN algorithm

DOI: 10.1002/etep

were also applied for transformer insulation fault diagnosis [31,32]. Moreover, some researchers adopted Bayesian network for transformer fault diagnosis using DGA measurement results [33].

The radial basis function (RBF) NN has a number of advantages such as simpler network structure and better approximation ability. An integrated self-adaptive training-based RBF for transformer diagnosis was developed in [34]. This RBF network was constructed by making use of fuzzy *c*-means (FCM) clustering and quantum-inspired particle swarm optimization. Recently, the wavelet has also been incorporated into ANNs for analyzing DGA data. For example, a three-layer structured ANN with an evolving wavelet networks [35] and a genetic algorithm tuned wavelet network [36]. In these two wavelet integrated networks, the network parameters and the weighting values were automatically tuned through an evolutionary algorithm-based optimization process.

Though they have been widely applied to transformer insulation diagnosis using oil characteristics, ANNs still suffer some inherited drawbacks. For example, it has no explanation ability and requires a relatively large size of historic dataset for training the networks.

4.2. *Fuzzy logic system and expert system*

The fuzzy logic approach transforms the experience acquired by human experts into decision rules and membership functions. The diagnosis on the insulation system of a transformer can be drawn by mapping the oil characteristics of this transformer to a set of rules. A number of researchers applied fuzzy logic techniques to power transformer fault diagnosis [37–40].

In [41], the authors developed a fuzzy logic system for diagnosing transformer insulation and providing recommendation on maintenance actions. In the system, the fuzzy set concept was adopted to manage the uncertainties in key gas analysis, thresholding, and gas ratio boundaries. In [42], the authors introduced a framework for conducting transformer diagnostics by applying fuzzy information theory. The fuzzy relations were combined into a decision tree to give the diagnosis on transformer incipient faults. In [13], Su *et al*. proposed a fuzzy logic system to deal with the problem, in which multiple faults simultaneously occur in a transformer insulation system. The proposed system can also indicate the severity of each fault. In [43], the authors adopted the acceptable/unacceptable norms of both key gas concentrations and gas ratios in the implementation of the fuzzy logic system. Their method can interpret the boundary cases, in which a transformer has nearly equal probability of having two different types of faults. In [44], Duraisamy *et al*. applied triangular, trapezoidal, and Gaussian membership functions to a fuzzy logic system, which was then integrated with BP network for diagnosing transformer faults. The conventional IEC/IEEE DGA criteria and the gas concentrations were also used as references to build the fuzzy diagnosis system.

The rule-based expert system has also been applied to transformer diagnosis. Such expert system represents the human experts' knowledge into the forms of IF–THEN rules, which is applied for the evaluation of transformer insulation [45]. The rule-based expert system is also integrated with fuzzy logic to deal with uncertainties in the diagnosing process. For example, Flores *et al*. combined an expert system with a Type-2 fuzzy logic system for evaluating the transformer insulation by using oil characteristics [46]. This hybrid system was able to detect whether paper insulation was involved in any insulation faults, and it also allowed other factors as inputs of the pattern recognition algorithm. In [47], Abu-Siada *et al*. incorporated several DGA interpretation schemes into a single expert model to overcome individual expert's limitations. It was reported that such approach can improve the diagnosis performance and also pave the way for standardizing DGA interpretation schemes. The major limitations of fuzzy logic and expert system for transformer insulation diagnosis are that the performances are highly decided by the completeness of the pre-defined knowledge base. Neither fuzzy logic system nor expert system can automatically adjust the system parameters when new knowledge is incorporated.

4.3. *Decision-making algorithms*

Transformer insulation diagnosis is based on a variety of oil characteristics and is formulated as a multiple-attribute decision-making problem. Two commonly adopted methods for solving this problem are evidence reasoning and grey theory.

Tang *et al*. adopted an evidence reasoning algorithm to combine evidences and deal with uncertainties in transformer condition assessment [48]. The algorithm provided the overall evaluation of the transformer condition and ranked several transformers based on their necessities of maintenance. In [49], a fuzzy set theory-based algorithm was first employed, which provided the diagnosis results as a set of possible types of faults with probability to each type of fault. These diagnoses were then aggregated using an evidential reasoning algorithm. Based on an information fusion strategy, a multi-level and multi-aspect expert system was developed in [50].

The grey model can perform pattern recognition using a relatively small size dataset and without involving formal statistic process and inference. A number of approaches combining grey theory and extension theory were applied to predict the trend of dissolved gas in transformer oil [51,52].

4.4. *SVM and population-based algorithms*

There are also a number of other types of pattern recognition algorithms which have been applied for interpreting oil characteristics data. For example, the SVM algorithm and the population based approaches such as particle swarm optimizer (PSO) and genetic programming (GP).

A Parzen windows-based classifier was proposed in [53]. This classifier integrated with a PSO to search for the optimal parameters for Parzen windows-based classifier. In [54], GP and bootstrap were implemented to deal with highly versatile DGA dataset. The bootstrap was used as a pre-processing to make the sample numbers of different fault types equal. Then, a GP was applied to extract features, which were subsequently fed into classifiers for transformer insulation diagnosis.

Over the last two decades, SVMs have been applied to various classification problems. SVM converts the input data from the original space to a higher dimensional space. In [55], the authors implemented a multi-layer SVM classifier and demonstrated its applicability to fault diagnosis of power transformer. In [56], a clonal selection algorithm is adopted to select the optimal input features and appropriate parameters for a SVM algorithm. In [57], the SVM algorithm was integrated with GA to make the forecasting of gases concentrations in transformer oil. This hybrid algorithm can prevent from over-fitting or under-fitting with the proper selection of SVM parameters using GA. Particle swarm optimization (PSO) can also be integrated into SVM, in which PSO searches the optimal parameters for SVM [58–60]. Some modifications have been made by adopting time-varying acceleration coefficients for improving the PSO convergence in searching the optimal parameters in SVM [61]. Moreover, the artificial immune network classification algorithm was also applied to transformer diagnosis [62].

4.5. *Other hybrid algorithms*

Attempts have been made to combine ANNs and fuzzy logics or expert systems in transformer insulation diagnosis. Such hybrid algorithm takes the advantages of the learning capability of ANNs, the knowledge formation of expert system, and the uncertainty representation of fuzzy logics. Xu *et al*. proposed a consultative mechanism to combine fuzzy logic and ANN [63]. A combined expert system and ANN was also proposed in [64]. In the above two hybrid algorithms, the knowledge base was derived from IEEE/IEC interpretation schemes and also included the experts' experiences. Miranda *et al*. proposed to combine a NN into a fuzzy system for the extraction of rules [65]. Some other hybrid approaches proposed in the literature include: integration of conventional DGA interpretation schemes with ANNs and fuzzy logics [66], integrated neural-fuzzy approach [14], association rule mining [67], hybrid of fuzzy approach, and evidential reasoning-based decision-making approach [68]. Other types of hybrid approaches have also been proposed in the literature. For example, in [69], the Dempster–Shafer evidential theory was combined with BP-NN and fuzzy logic. In [70], neuro-fuzzy scheme and PNN were integrated. Table I summarizes the data configuration (total samples number, training dataset size, testing dataset size, number of input features, and the types of transformer insulation faults in terms of fault classes) as well as the classification accuracies of different pattern recognition algorithms presented in the literature. This table does not intend to supply an exhaustive survey but aims to provide a clear picture of the setup and performances of some state-of-the-art pattern recognition algorithms for transformer insulation diagnosis.

DOI: 10.1002/etep

Table I. Dataset configurations and results of some representative pattern recognition algorithms in the literature.

Ref	Model	Samples	Training	Testing	# features	# classes	Accuracy (%)
[17]	Two-step ANN (BP)	40	N/A	N/A	5	4	86–95
[18]	ANN	26	N/A	N/A	5/8/3	3	73–96
[19]	ANN- BP	150	117	33	3/5	8	87–100
[21]	Polynomial networks (BP)	711	N/A	N/A	3/5	8	87–97
[22]	Multivariate polynomial neural network	167	156	11	6	N/A	100
[23]	PNN	503	497	4	4	8	100
[24]	Fuzzy LVQ	711	N/A	N/A	3	8	97
[25]	Genetic-based neural networks	630	N/A	N/A	3/5	5	91–95
[26]	Evolving neural nets	820	N/A	N/A	3/5	5	90–93
[28]	Extension theory and ANN	22	N/A	N/A	5	9	96–100
[30]	Auto associative neural networks	352	N/A	N/A	3	5	100
[41]	Expert system	101	N/A	N/A	9	3	93
[71]	Evolutionary fuzzy logic	711	N/A	N/A	3	8	92
[72]	Adaptive fuzzy logic	561	N/A	280	N/A	5	94
[43]	Fuzzy logic	20	N/A	N/A	9	9	100
[51]	Grey prediction	46	N/A	N/A	10	10	97
[53]	Parzen–Windows and PSO	168	N/A	N/A	3/3/8	4	80
[55]	SVM	75	50	25	5	4	100
[64]	ANN and ES	210	150	60	24	6	93–96
[65]	Knowledge extraction and ANN	318	N/A	88	3	5	99
[73]	Clustering and extension theory	21	N/A	N/A	8	9	88
[74]	SVM and genetic algorithm	142	N/A	N/A	5	4	94

5. KEY ISSUES OF DEVELOPING THE "READY-TO-USE" PATTERN RECOGNITION ALGORITHMS FOR TRANSFORMER DIAGNOSIS

A power transformer is constructed with the complex combination of different materials, and its operation is extensively complicated; it is not a trivial task to make assessment on transformer insulation system using oil characteristics. Pattern recognition-based interpretation scheme is still not a ready-to-use tool for utilities. There are several key issues that need to be further investigated.

The first issue is the evaluation of the performance of pattern recognition algorithms in diagnosing transformer insulation faults. It can be observed from Table I that there was a lack of a common framework for defining the process of training, cross validation, testing, and evaluation to ensure the generalizability and applicability of various algorithms as well as for providing the statistical comparisons amongst different algorithms.

The second issue is how to build up a statistically satisfied training database. It is well known that most pattern recognition algorithms require a large size of historic dataset for training. However, the occurrence of transformer fault is an event with relatively small probability. Therefore, the historic dataset may only consist of few records of certain fault types. If an algorithm is trained by such database, it cannot make reliable insulation diagnosis. In addition, the process of collecting oil samples, conducting dissolved gas measurement, and interpreting the measured data may vary among different utilities. This may compromise the generalization capability of pattern recognition algorithms, i.e. some algorithms trained on a "local" dataset cannot readily applied "globally".

The third issue is the data quality issue. Because of the complex geometry of transformer insulation system, limitation of the measurement system, and possible presence of multiple faults in transformer insulation system, there may exist inconsistence in reaching diagnosis of transformer insulation.

The next section will present an illustrative case study to address procedures of dataset preparation, training, cross validation, and testing for applying pattern recognition algorithm for transformer insulation diagnosis. The statistical comparison of different pattern recognition algorithms will be addressed in the accompanying paper (i.e. Part 2 of this paper) through extensive case studies [16].

DOI: 10.1002/etep

The research directions for solving the above second and third issues will also be discussed in the accompanying paper [16].

6. AN ILLUSTRATIVE CASE STUDY—BY GENERALIZED REGRESSION NEURAL NETWORK (GRNN)

In this case study, the GRNN is applied to make transformer insulation diagnosis. GRNN has been widely used in applications involving classification and predication. It can learn the underlying functions between input and output (e.g. class) from samples without having the prior knowledge of any specific function form between input and output [75]. Thus, GRNN has the advantages of having simple structure and less computation time.

In the case study, a database consisting of oil characteristics with the already-known insulation conditions of the corresponding transformers is used for training, validation, and testing GRNN. Table II presents the configuration of such a database used in this case study. The original data in this database was obtained from a utility company. In Table II, the features (input of algorithms) are the concentrations of dissolved gases (i.e. hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2)) and oil test results of water content, acidity, DDF, 2-furfuraldehyde, resistivity, and BDV. The insulation condition is categorized into four different classes (categories) by combining the diagnosis results obtained from FCM clustering algorithm, Duval triangle method, and utility experts' assessment. The readers may refer to the authors' previous publication for more details of constructing this training database [76].

The procedure of training, cross validation, and testing of GRNN algorithm (other pattern recognition algorithms as well) using the database in Table II is as follows:

(1) Normalizing all data into the range of [0, 1].
(2) Randomly splitting database into a training sub-dataset and testing sub-dataset by 70% and 30% (this ratio can be adjusted based on the user's requirements).
(3) Performing tenfold cross validation on training sub-dataset to obtain the optimal parameters of the algorithm.
(4) Training the algorithm with the optimal parameters obtained in 3).
(5) Testing the trained algorithm on the test sub-dataset (e.g. unseen data by the algorithm).
(6) Obtaining the classification accuracy.

The above procedure will be repeated 50 times to obtain the statistical performance evaluation on the algorithm. The averaged classification accuracy on each class and the overall classification accuracy over 50 runs of the algorithm are adopted as the performance evaluation criteria.

Figure 2 depicts the sample distribution of the database as described in Table II. It is obvious that this database has unequal distributed samples in different classes, e.g. 80 samples belong to "Excellent" class while only 21 samples belong to "Fair" class. This will introduces significant difficulties for algorithms to make correct classification on the minority class. Some re-sampling or data balancing techniques can be applied to address this issue [10].

After establishing the training sub-dataset, tenfold cross validation is applied to obtain the optimal parameters of GRNN. The procedure of the tenfold cross validation is as follows: (i) the original training sub-dataset with N samples are divided into 10 groups of size $N/10$; (ii) the algorithm is trained on nine groups and tested on one group; and (iii) it is performed for 10 times, and the mean accuracy is taken as the classification accuracy of the algorithm. The above procedure was repeated for each possible value of algorithm's parameters, and the parameters with the highest mean accuracy in the

Table II. Configuration of database using oil characteristics.

Sample	Feature	Class (condition of transformer insulation)			
181	11	Excellent 80	Good 50	Fair 21	Poor 30

DOI: 10.1002/etep

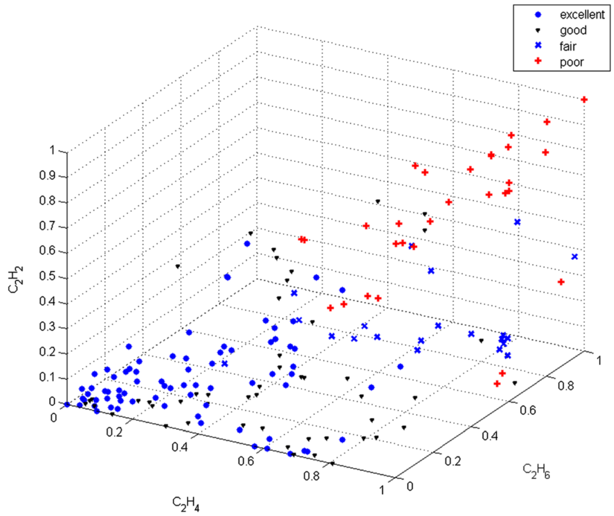

Figure 2. Sample distribution of the database in Table II.

Table III. Classification accuracy of GRNN algorithm for transformer insulation diagnosis using oil characteristics.

Excellent	Good	Fair	Poor	Overall
96%	90%	75%	83%	90%

above step (iii) are taken as the best parameters. After finding the optimal values of the above parameters, the algorithm is trained on the whole training sub-dataset. Subsequently, the trained algorithm is applied to evaluate the conditions of transformers in the testing sub-dataset. At this stage, the algorithm automatically interprets the oil characteristics data of the transformers in the testing sub-dataset and classifies these transformers into particular categories of insulation conditions such as Excellent, Good, Fair, and Poor condition. The above optimal parameter acquisition, training, and testing procedure is repeated for 50 times (runs). At each run, the classification accuracy is computed according to N'_{total}/N_{total}, where N'_{total} denotes the number of transformers that are correctly classified, and N_{total} denotes the total number of transformers in the testing sub-dataset. The overall accuracy is the average over 50 runs. Table III presents the classification accuracy (averaged over 50 runs) of the GRNN algorithm.

It can be observed from Table III that the GRNN algorithm can correctly recognize faults occurring in most transformers in the testing dataset. The overall classification accuracy is 90% averaged over 50 runs. However, the fault classification accuracy of transformers with "Fair" insulation condition may not be satisfied (about 75%). This is because the original database (Table II) is imbalanced, in which the transformers with "Fair" condition are outnumbered by the transformers with other three conditions (Excellent, Good, and Poor). Trained by such an imbalanced database, the algorithm can lead to misclassification on the class of transformers with "Fair" condition. To deal with such problem, some preprocessing methods can be integrated with the algorithm to facilitate it achieving consistent desirable classification accuracy [10,76].

7. CONCLUSIONS

This paper studied the pattern recognition techniques and their application for power transformer insulation diagnosis using oil characteristics. A general pattern recognition application framework was presented in the paper. A comprehensive literature review on the state-of-the-art pattern recognition techniques for transformer insulation diagnosis was also provided. Moreover, a case study was

DOI: 10.1002/etep

presented to illustrate the process of oil characteristics database preparation, training/testing datasets formation, cross validation, testing, and evaluation in applying pattern recognition algorithm for diagnosing transformer insulation.

8. LIST OF SYMBOLS AND ABBREVIATIONS

8.1. Symbols

H_2	Hydrogen
CH_4	Mathane
C_2H_6	Ethane
C_2H_4	Ethylene
C_2H_2	Acetylene
N	sample number in training sub-dataset
N'_{total}	number of correctly classified samples in testing sub-dataset
N_{total}	sample number in testing sub-dataset

8.2. Abbreviations

BDV	Breakdown Voltage
DDF	Dielectric Dissipation Factor
DGA	Dissolved Gas Analysis
PD	Partial Discharge
ANN	Artificial Neural Network
SVM	Support Vector Machine
NN	Neural Network
BP	Back Propagation
RMP	Reduced Multivariate Polynominal
PNN	Probabilistic Neural Network
FLVQ	Fuzzy Learning Vector Quantization
SOM	Self Organizing Map
*k*NN	*k*-Nearest Neighbor
RBF	Radial Basis Function
FCM	Fuzzy *c*-Means
PSO	Particle Swarm Optimizer
GP	Genetic Programming
GRNN	Generalized Regression Neural Network

ACKNOWLEDGEMENT

We gratefully acknowledge the Australian Research Council, Powerlink Queensland, Energex, Ergon Energy, and TransGrid for providing supports for this work (Grant No. LP110100324).

REFERENCES

1. Wang M, Vandermaar AJ, Srivastava KD. Review of condition assessment of power transformers in service. *IEEE Electrical Insulation Magazine* 2002; **18**:12–25.
2. Jahromi AN, Piercy R, Cress S, Service JRR, Fan W. An approach to power transformer asset management using health index. *IEEE Electrical Insulation Magazine* 2009; **25**:20–34.
3. Saha TK. Review of modern diagnostic techniques for assessing insulation condition in aged transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* 2003; **10**:903–917.
4. Saha TK, Purkait P. Investigation of polarization and depolarization current measurements for the assessment of oil-paper insulation of aged transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* 2004; **11**:144–154.
5. Duval M. New techniques for dissolved gas-in-oil analysis. *IEEE Electrical Insulation Magazine* 2003; **19**:6–15.

6. Wang Z, Li J, Sofian DM. Interpretation of transformer FRA responses-part I: influence of winding structure. *IEEE Transactions on Power Delivery* 2009; **24**:703–710.
7. Ekanayake C, Gubanski SM, Graczkowski A, Walczak K. Frequency response of oil impregnated pressboard and paper samples for estimating moisture in transformer insulation. *IEEE Transactions on Power Delivery* 2006; **21**:1309–1317.
8. Strachan SM, Rudd S, McArthur SDJ, Judd MD, Meijer S, Gulski E. Knowledge-based diagnosis of partial discharges in power transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* 2008; **15**:259–268.
9. Duval M, DePablo A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electrical Insulation Magazine* 2001; **17**:31–41.
10. Ma H, Ekanayake C, Saha TK. Power transformer fault diagnosis under measurement originated uncertainties. *IEEE Transactions on Dielectrics and Electrical Insulation* 2012; **19**:1982–1990.
11. Da Silva IN, Imamura MM, de Souza AN. Application of neural networks to the analysis of dissolved gases in insulating oil used in transformers. Proceedings of IEEE International Conference on Systems, Man and Cybernetics, Nashville, TN, USA, 2000; 2643–2648.
12. Xu Z. Multivalued logic and fuzzy logic - their relationship, minimization, and application to fault diagnosis. *IEEE Transactions on Computers* 1984; **C-33**:679–681.
13. Su Q, Mi C, Lai LL, Austin P. Fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer. *IEEE Transactions on Power Systems* 2000; **15**:593–598.
14. Naresh R, Sharma V, Vashisth M. An integrated neural fuzzy approach for fault diagnosis of transformers. *IEEE Transactions on Power Delivery* 2008; **23**:2017–2024.
15. Chen W, Pan C, Yun Y, Liu Y. Wavelet networks in power transformers diagnosis using dissolved gas analysis. *IEEE Transactions on Power Delivery* 2009; **24**:187–194.
16. Cui Y, Ma H, Saha TK. Pattern recognition techniques for power transformer insulation diagnosis - a comparative study part 2: implementation, case study, and statistical analysis. *International Transactions on Electrical Energy Systems* 2014. Accepted on 26 May, 2014.
17. Zhang Y, Ding X, Liu Y, Griffin PJ. An artificial neural network approach to transformer fault diagnosis. *IEEE Transactions on Power Delivery* 1996; **11**:1836–1841.
18. Vanegas O, Mizuno Y, Naito K, Kamiya T. Diagnosis of oil-insulated power apparatus by using neural network simulation. *IEEE Transactions on Dielectrics and Electrical Insulation* 1997; **4**:290–299.
19. Guardado JL, Naredo JL, Moreno P, Fuerte CR. A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis. *IEEE Transactions on Power Delivery* 2001; **16**:643–647.
20. Yang MT, Hu LS. Intelligent fault types diagnostic system for dissolved gas analysis of oil-immersed power transformer. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; **20**:2317–2324.
21. Yang HT, Huang YC. Intelligent decision support for diagnosis of incipient transformer faults using self-organizing polynomial networks. *IEEE Transactions on Power Systems* 1998; **13**:946–952.
22. Wu H, Li X, Wu D. RMP neural network based dissolved gas analyzer for fault diagnostic of oil-filled electrical equipment. *IEEE Transactions on Dielectrics and Electrical Insulation* 2011; **18**:495–498.
23. Lin WM, Lin CH, Tasy MX. Transformer-fault diagnosis by integrating field data and standard codes with training enhancible adaptive probabilistic network. *IEE Proceedings-Generation Transmission and Distribution* 2005; **152**:335–341.
24. Yang H, Liao C, Chou J. Fuzzy learning vector quantization networks for power transformer condition assessment. *IEEE Transactions on Dielectrics and Electrical Insulation* 2001; **8**:143–149.
25. Huang YC. Condition assessment of power transformers using genetic-based neural networks. *IEE Proceedings-Science Measurement and Technology* 2003; **150**:19–24.
26. Huang YC. Evolving neural nets for fault diagnosis of power transformers. *IEEE Transactions on Power Delivery* 2003; **18**:843–848.
27. Wang MH. Extension neural network for power transformer incipient fault diagnosis. *IEE Proceedings-Generation Transmission and Distribution* 2003; **150**:679–685.
28. Wang MH. A novel extension method for transformer fault diagnosis. *IEEE Transactions on Power Delivery* 2003; **18**:164–169.
29. Thang KF, Aggarwal RK, McGrail AJ, Esp DG. Analysis of power transformer dissolved gas data using the self-organizing map. *IEEE Transactions on Power Delivery* 2003; **18**:1241–1248.
30. Miranda V, Castro A, Lima S. Diagnosing faults in power transformers with autoassociative neural networks and mean shift. *IEEE Transactions on Power Delivery* 2012; **27**:1350–1357.
31. Sun S, Huang R. An adaptive k-nearest neighbor algorithm. Proceedings of Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2010; 91–94.
32. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 1967; **13**:21–27.
33. Carita A, Leite LC, Junior A, Godoy RB, Sauer L. Bayesian networks applied to failure diagnosis in power transformer. *IEEE Latin America Transactions* 2013; **11**:1075–1082.
34. Meng K, Dong ZY, Wang DH, Wong KP. A self-adaptive rbf neural network classifier for transformer fault analysis. *IEEE Transactions on Power Systems* 2010; **25**:1350–1360.
35. Huang YC, Huang CM. Evolving wavelet networks for power transformer condition monitoring. *IEEE Transactions on Power Delivery* 2002; **17**:412–416.

DOI: 10.1002/etep

36. Huang YC. A new data mining approach to dissolved gas analysis of oil-insulated power apparatus. *IEEE Transactions on Power Delivery* 2003; **18**:1257–1261.
37. Huang YC, Sun HC. Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; **20**:974–981.
38. Dhote NK, Helonde JB. Fuzzy algorithm for power transformer diagnostics. *Advances in Fuzzy Systems* 2013; **B**: 421621–421627.
39. David IG, Rajaram M, Gnanasaravanan A. Power transformer faults identification using fuzzy based dissolved gas analysis method. *Journal of Electrical Engineering* 2013; **13**:5.
40. Dhote NK, Helonde JB. Improvement in transformer diagnosis by DGA using fuzzy logic. *Journal of Electrical Engineering & Technology* 2014; **9**:615–621.
41. Lin CE, Ling JM, Huang CL. An expert system for transformer fault-diagnosis using dissolved-gas analysis. *IEEE Transactions on Power Delivery* 1993; **8**:231–238.
42. Tomsovic K, Tapper M, Ingvarsson T. A fuzzy information approach to integrating different transformer diagnostic methods. *IEEE Transactions on Power Delivery* 1993; **8**:1638–1646.
43. Islam SM, Wu T, Ledwich G. Novel fuzzy logic approach to transformer fault diagnosis. *IEEE Transactions on Dielectrics and Electrical Insulation* 2000; **7**:177–186.
44. Duraisamy V, Devarajan N, Somasundareswari D, Vasanth A, Sivanandam SN. Neuro fuzzy schemes for fault detection in power transformer. *Applied Soft Computing* 2007; **7**:534–539.
45. Kirtley JL, Hagman WH, Lesieutre BC, *et al*. Monitoring the health of power transformers. *IEEE Computer Applications in Power* 1996; **9**:18–23.
46. Flores WC, Mombello EE, Jardini JA, Rattá G, Corvo AM. Expert system for the assessment of power transformer insulation condition based on type-2 fuzzy logic systems. *Expert Systems with Applications* 2011; **38**:8119–8127.
47. Abu-Siada A, Hmood S, Islam S. A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; **20**:2343–2349.
48. Tang WH, Spurgeon K, Wu QH, Richardson ZJ. An evidential reasoning approach to transformer condition assessments. *IEEE Transactions on Power Delivery* 2004; **19**:1696–1703.
49. Spurgeon K, Tang WH, Wu QH, Richardson ZJ, Moss G. Dissolved gas analysis using evidential reasoning. *IEE Proceedings-Science Measurement and Technology* 2005; **152**:110–117.
50. Wang XL, Li QM, Li CR, Yang R, Su Q. Reliability assessment of the fault diagnosis methodologies for transformers and a new diagnostic scheme based on fault info integration. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; **20**:2292–2298.
51. Wang MH. Grey-extension method for incipient fault forecasting of oil-immersed power transformer. *Electric Power Components and Systems* 2004; **32**:959–975.
52. Wang MH, Hung CP. Novel grey model for the prediction of trend of dissolved gases in oil-filled power apparatus. *Electric Power Systems Research* 2003; **67**:53–58.
53. Tang WH, Goulermas JY, Wu QH, Richardson ZJ, Fitch J. A probabilistic classifier for transformer dissolved gas analysis with a particle swarm optimizer. *IEEE Transactions on Power Delivery* 2008; **23**:751–759.
54. Shintemirov A, Tang W, Wu QH. 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* 2009; **39**:69–79.
55. Lv G, Cheng H, Zhai H, Dong L. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electric Power Systems Research* 2005; **75**:9–15.
56. Cho M, Lee T, Gau S, Shih C. Power transformer fault diagnosis using support vector machines and artificial neural networks with clonal selection algorithms optimization. In *Knowlegde-Based Intelligent Information and Engineering Systems*. Springer Berlin Heidelberg: Berlin, 2006; 179–186.
57. Fei S, Sun Y. Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm. *Electric Power Systems Research* 2008; **78**:507–514.
58. Liao R, Zheng H, Grzybowski S, Yang L, Tang C, Zhang Y. Fuzzy information granulated particle swarm optimisation-support vector machine regression for the trend forecasting of dissolved gases in oil-filled transformers. *IET Electric Power Applications* 2011; **5**:230–237.
59. Lee T, Cho M, Fang F. Features selection of SVM and ANN using particle swarm optimization for power transformers incipient fault symptom diagnosis. *International Journal of Computational Intelligence Research* 2007; **3**:60–65.
60. Lee T, Cho M, Shieh C, Lee H, Fang F. Particle swarm optimization-based SVM for incipient fault classification of power transformers. In *Foundations of Intelligence Systems*. Springer Berlin Heidelberg: Berlin, 2006; 84–90.
61. Liao RJ, Zheng HB, Grzybowski S, Yang LJ. A multiclass SVM-based classifier for transformer fault diagnosis using a particle swarm optimizer with time-varying acceleration coefficients. *International Transactions on Electrical Energy Systems* 2013; **23**:181–190.
62. Xiong H, Sun C. Artificial immune network classification algorithm for fault diagnosis of power transformer. *IEEE Transactions on Power Delivery* 2007; **22**:930–935.
63. Xu W, Wang D, Zhou Z, Chen H. Fault diagnosis of power transformers: Application of fuzzy set theory, expert systems and artificial neural networks. *IEE Proceedings-Science Measurement and Technology* 1997; **144**:39–44.

DOI: 10.1002/etep

64. Wang Z, Liu Y, Griffin P. A combined ANN and expert system tool for transformer fault diagnosis. *IEEE Transactions on Power Delivery* 1998; **13**:1224–1229.
65. Miranda V, Castro A. Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks. *IEEE Transactions on Power Delivery* 2005; **20**:2509–2516.
66. Morais D, Rolim J. A hybrid tool for detection of incipient faults in transformers based on the dissolved gas analysis of insulating oil. *IEEE Transactions on Power Delivery* 2006; **21**:673–680.
67. Yang Z, Tang W, Shintemirov A, Wu Q. 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* 2009; **39**:597–610.
68. Liao R, Zheng H, Grzybowski S, Yang L, Zhang Y, Liao Y. An integrated decision-making model for condition assessment of power transformers using fuzzy approach and evidential reasoning. *IEEE Transactions on Power Delivery* 2011; **26**:1111–1118.
69. Bhalla D, Bansal RK, Gupta HO. Integrating AI based DGA fault diagnosis using dempster-shafer theory. *International Journal of Electrical Power & Energy Systems* 2013; **48**:31–38.
70. Malik H, Yadav AK, Mishra S, Mehto T. Application of neuro-fuzzy scheme to investigate the winding insulation paper deterioration in oil-immersed power transformer. *International Journal of Electrical Power & Energy Systems* 2013; **53**:256–271.
71. Huang YC, Yang HT, Huang CL. Developing a new transformer fault diagnosis system through evolutionary fuzzy logic. *IEEE Transactions on Power Delivery* 1997; **12**:761–767.
72. Yang HT, Liao CC. Adaptive fuzzy diagnosis system for dissolved gas analysis of power transformers. *IEEE Transactions on Power Delivery* 1999; **14**:1342–1350.
73. Wang M, Tseng Y, Chen H, Chao K. A novel clustering algorithm based on the extension theory and genetic algorithm. *Expert Systems with Applications* 2009; **36**:8269–8276.
74. Fei S, Zhang X. Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications* 2009; **36**:11352–11357.
75. Specht DF. A general regression neural network. *IEEE Transactions on Neural Networks* 1991; **2**:568–576.
76. Ashkezari AD, Ma H, Saha TK, Ekanayake C. Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; **20**:965–973.

DOI: 10.1002/etep