

# Artificial neural network approaches for fault classification: comparison and performance

Tapsi Nagpal · Yadwinder Singh Brar

Received: 18 February 2014 / Accepted: 14 July 2014 / Published online: 1 August 2014  
© The Natural Computing Applications Forum 2014

**Abstract** This manuscript focuses the implementation of artificial neural network-based algorithms to classify different types of faults in a power transformer, meant particularly for NonDestructive Test for transformer fault classification. The performance analysis of Probabilistic Neural Network (PNN) and Backpropagation Network classifiers has been carried out using the database of dissolved gases collected from Punjab State Electricity Board, Patiala, India. Features from the preprocessed data have been extracted using dimensionality reduction technique, i.e., principal component analysis. The selected features were used as inputs to the Backpropagation Network and PNN classifiers. A comparative study of the two intelligent classifiers has been carried out, which reveals that PNN classifier outperforms the Backpropagation Network classifier.

**Keywords** Power transformer · Probabilistic Neural Network · Backpropagation Neural Network · Fault Diagnosis

## 1 Introduction

Power transformer is one of the major apparatus in transmission and distribution system. The failure of power transformer causes huge economic losses to industry and

inconvenience to the general public. In power transformers, two types of insulations are used liquid insulation (mineral oil or transformer oil) and solid impregnated insulation (cellulose). Out of the two, the liquid insulation is very important. The transformer oil provides electrical insulation, dissipates heat, helps to preserve the core and winding and prevents the direct contact of atmospheric oxygen with cellulose-made paper insulation of windings [1–6].

The thermal and electrical stresses decompose the transformer oil, produce harmful gases such as hydrogen ( $H_2$ ), methane ( $CH_4$ ), acetylene ( $C_2H_2$ ), ethylene ( $C_2H_4$ ) and ethane ( $C_2H_6$ ), and subsequently, the oil gets damaged. This declines the performance of transformer oil. One can prevent this by knowing the exact amount of harmful gases dissolved in the transformer oil. The different conventional methods such as Roger's ratio method, Dornenburg's method, Duval's triangle method and key gas ratio methods are used to ascertain the exact amount of harmful gases dissolved in the transformer oil. These methods are either inconclusive on fault or give a false fault type [7–11]. To overcome these uncertainties in conventional methods, various intelligent methods such as artificial neural networks [12, 13], Wavelet Analysis [14], Least Vector Quotient [15], Probabilistic Neural Network (PNN) [16], fuzzy logic [17–19] Support Vector Machine classifiers [20–23] and Self-Organizing Map classifiers [24, 25] have been proposed.

This article deals with fault classification in power transformers using Backpropagation Neural Network (BPN) and PNN. The merits of BPN classifier include fair approximation of a large class of functions, relatively simple implementation, and mathematical formula used in BPN algorithm can be applied to any network. On the other hand, PNN classifier can generate accurate predicted target probability scores. They are insensitive to outliers and have faster convergence speed than BPN classifiers.

T. Nagpal (✉)  
Department of Electrical and Instrumentation Engineering,  
Thapar University, Patiala 147004, India  
e-mail: tapsiarora@gmail.com; tapsi.nagpal@thapar.edu

Y. S. Brar  
Department of Electrical Engineering, Guru Nanak Dev  
Engineering College, Ludhiana 141006, India

Performance of BPN and PNN classifiers has been compared for the transformer fault classification. PNN classifier gives better results compared to BPN classifier. The comparison has been carried out amongst other BPN architectures too to classify transformer faults.

## 2 Proposed fault classification scheme

Figure 1 shows the flow chart of the proposed transformer fault classification scheme. The raw data are collected and preprocessed. After that, the dimension reduction and feature selection are performed and in the last step of the classification, neural network-based classifiers have been applied to determine different faults.

## 3 Methodology

### 3.1 Dissolved gas analysis (DGA)

Dissolved gas analysis (DGA) provides advance warning of developing faults. Some of the methods used in industry for DGA are IEEE std. C 57.104:1991, IEC std. 60599:1999, Duval's triangle, CIGRE, Nomograph methods. IEEE std. C 57.104:1991 and IEC std. 60599:1999 methods are key gas ratio methods [26, 27]. Ratio method

does not cover the entire range of data. So, the fault classification sometimes gives no results. CIGRE method is combination of key gas ratio method and gas concentration method [28, 29]. Duval's triangle method and Nomograph method are graphical methods. These methods have a limitation. When multiple DGA faults occur in the system, none of these methods are able to detect it. Table 1 shows the allowable range of harmful gases (in ppm) in transformer oil for OLTC and commutating OLTC [30]. Table 2 shows different faults of power transformer defined by a combined IEC/IEEE and CIGRE criteria [26].

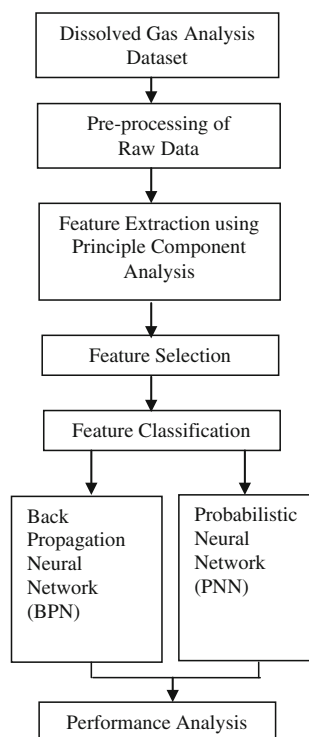
### 3.2 Data collection

Transformers from ten substations of Punjab State Electricity Board, Patiala (India), are used to collect gas samples. The data are collected as per the ASTM standards. The transformer rating ranges from 52–63 MVA. The range of voltage is 132/33/11 kV. After the data collection, they are preprocessed by removing linear trends, outliers, etc. Table 3 shows the preprocessed data of samples obtained from Punjab State Electricity Board. The raw gas data, collected from different transformers, are statistically analyzed. The variance plot of each of the gas sample has been represented by ANOVA plot in Fig. 2. For normalization purpose, the mean value is subtracted from each point, i.e., the data can be said zero-mean data.

### 3.3 Feature selection

The process of mapping original features of the data into fewer, more effective features is called feature extraction. Linear Discriminant Analysis (LDA) and Principle Component Analysis (PCA) are some of the well-known feature extraction methods [31, 32]. LDA is a supervised feature extraction technique, whereas PCA is unsupervised feature extraction technique. Principal component analysis (PCA) also known as Karhunen–Loeve transform is one of the most popular statistical technique, which reduces the dimensions of a dataset but preserves the correlation structure in the data. It is basically used for feature extraction. Steps of PCA are as follows:

- (i) Get the input data
- (ii) Calculate the mean of data and subtract the mean
- (iii) Calculate the covariance  $\text{cov}(x, y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{n-1}$  where  $X_i$ , i.e., input, is the DGA Dataset,  $i = 1-600$ ,  $\bar{X}$  is the mean value of the dataset and  $y$  (i.e., output) is the fault type of the transformer
- (iv) Calculate the eigenvector and eigenvalue of covariance matrix



**Fig. 1** Block diagram of proposed transformer fault classification scheme

**Table 1** Allowable range of harmful gases (in ppm) in transformer oil for OLTC and commutating OLTC as per IEC 60599

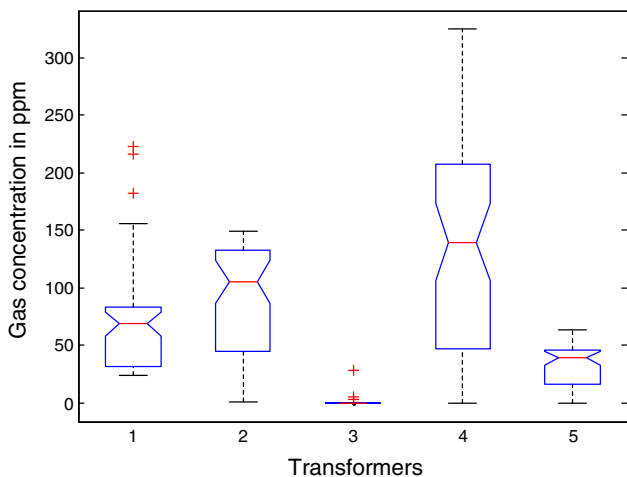
Transformer type	Hydrogen (H <sub>2</sub> )	Methane (CH <sub>4</sub> )	Ethane (C <sub>2</sub> H <sub>2</sub> )	Ethylene (C <sub>2</sub> H <sub>4</sub> )	Acetylene (C <sub>2</sub> H <sub>2</sub> )	Carbon monoxide (CO)	Carbon dioxide (CO <sub>2</sub> )
No OLTC	60–150	40–110	50–90	60–280	3–50	540–900	5,100–13,000
Communicating OLTC	75–150	35–130	50–70	110–250	80–270	400–850	5,300–12,000

**Table 2** Combined criterion of IEC/IEEE and CIGRE standards for integrating fault types

Fault type	Ratio of C <sub>2</sub> H <sub>2</sub> /C <sub>2</sub> H <sub>4</sub> (R <sub>1</sub> )	Ratio of CH <sub>4</sub> /H <sub>2</sub> (R <sub>2</sub> )	Ratio of C <sub>2</sub> H <sub>4</sub> /C <sub>2</sub> H <sub>6</sub> (R <sub>3</sub> )	Ratio of CO/CO <sub>2</sub> (R <sub>4</sub> )
<300 °C thermal fault	<0.1	>1	<1	<0.1
300–700 °C thermal fault	<0.1	>1	1–3	<0.1
300–700 °C thermal fault with cellulose decomposition	<0.1	>1	1–3	0.1–1
>700 °C thermal fault	<0.1	>1	>3	<0.1
>700 °C thermal fault involving cellulose decomposition	<0.1	>1	>3	0.1–3.3
Low-energy partial discharge (corona)	<0.1	<0.1	<1	0.1–0.3
High-energy partial discharge (corona)	0.1–3	<0.1	<1	0.1–0.3
Low-energy discharge (arcing)	>0.1	0.1–1	>1	0.3–3.3
High-energy discharge (arcing)	0.1–3	0.1–1	>3	0.3–3.3

**Table 3** Preprocessed samples of data of dissolved gases in power transformers of Punjab state electricity board

Sample No.	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>
1	336	419	105	1,074	21
2	160	130	33	96	0
3	57	77	58	21	0
4	565	93	34	47	0
5	650	53	34	20	0



**Fig. 2** The variance plot of raw data collected from different transformers

(v) Choosing components and forming a feature vector

(vi) Derive new dataset from following formula:

$$\text{Final data} = \text{Row Feature Vector} \times \text{Row Data Adjust}$$

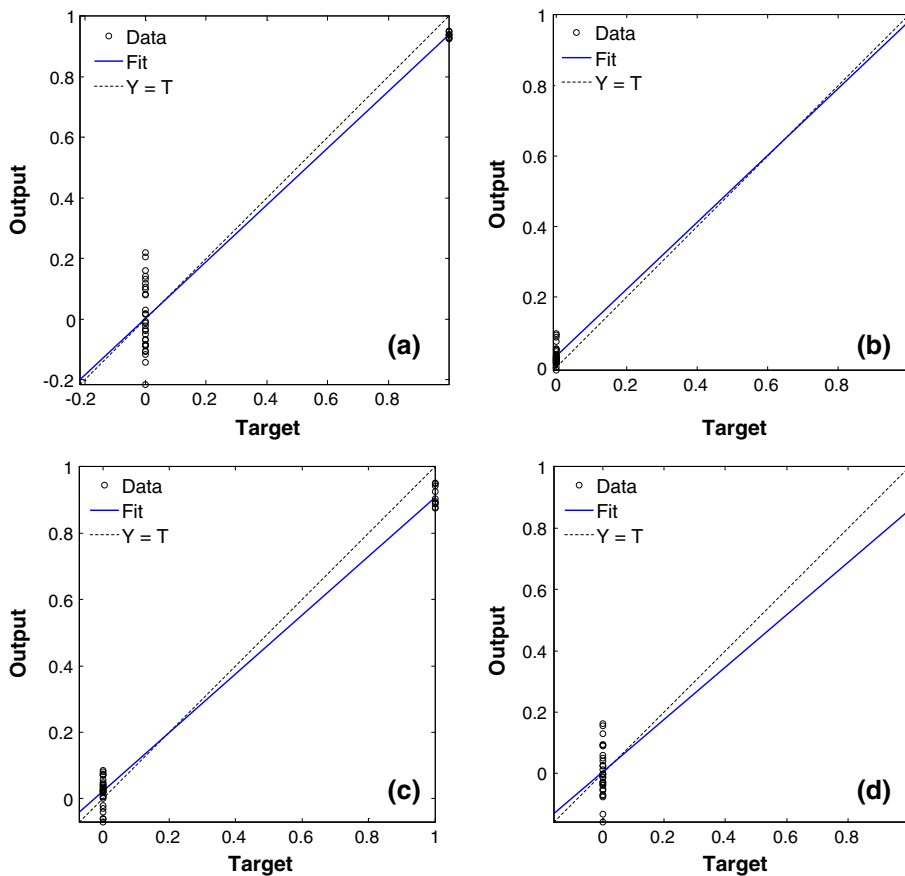
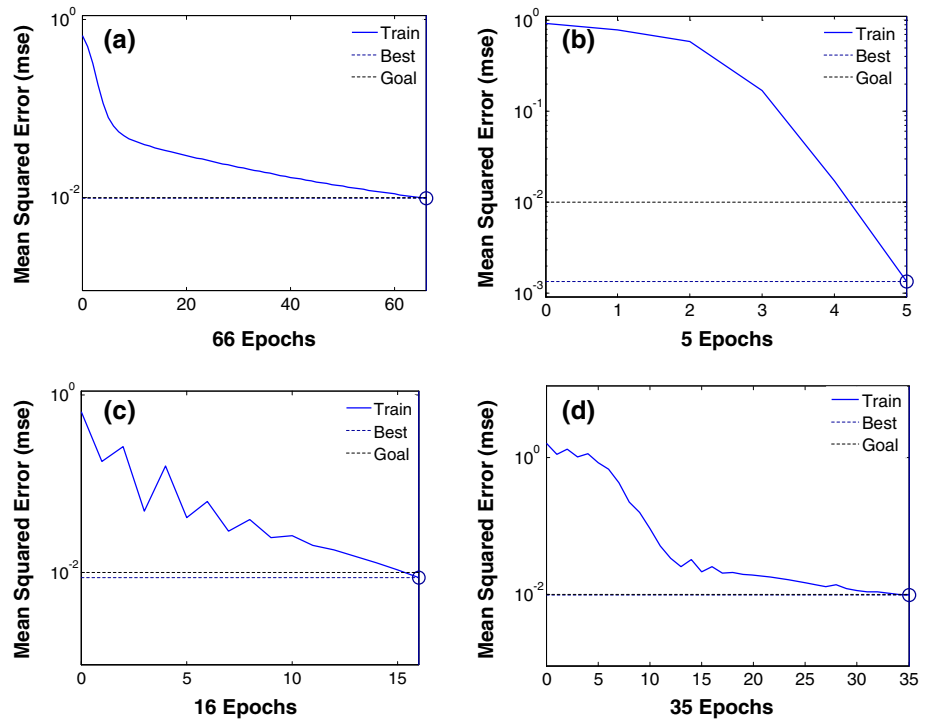
where Row Feature Vector is the matrix with eigenvectors in the columns transpose and Row data Adjust is the mean-adjusted transpose. An assumption made for feature extraction and dimensionality reduction by PCA is that most information of the observation vectors is contained in the subspace spanned by the first  $m$  principal axes, where  $m < p$  for a  $p$ -dimensional data space. Therefore, each original data vector can be represented by its principal component vector with dimensionality  $m$ .

### 4 Results and discussion

The Backpropagation Neural Network and Probabilistic Neural Network have been used as fault classifiers, to classify different transformer fault types according to the international standard IEC 60599. The following section describes the two fault classifiers briefly.

Artificial neural network maps the input samples and output in a nonlinear fashion. Backpropagation is one of the age-old learning algorithms and is used to train a multilayer feedforward neural network. Apart from the input and output layer, there is one hidden layer. The network has been trained for different numbers of neurons

**Fig. 3** Error and epoch graph of **a** Gradient Descent Method. Best Training Performance is 0.0099 at epoch 66. **b** Levenberg–Marquardt method. Best Training Performance is 0.099 at epoch 5. **c** Conjugate Gradient Descent Method. Best Training Performance is 0.0087 at epoch 16. **d** Resilient Method. Best Training Performance is 0.0097 at epoch 35



**Fig. 4** Regression Plot of **a** Gradient Descent Method.  $R = 0.97$ . Output =  $0.94 \times \text{Target} + 0.0022$ . **b** Levenberg–Marquardt method.  $R = 0.99$ . Output =  $0.95 \times \text{Target} + 0.032$ . **c** Conjugate Gradient

Descent Method.  $R = 0.96$ . Output =  $0.89 \times \text{Target} + 0.020$ . **d** Resilient Method.  $R = 0.95$ . Output =  $0.86 \times \text{Target} + 0.004$

in the hidden layer. It has been found by trial and error that eighteen hidden neurons for this problem gave the best result. The networks are trained until the mean square error of the training samples fell below 0.005. In this paper, four different backpropagation learning algorithms, namely Gradient Descent, Levenberg–Marquardt, Conjugate Gradient and resilient backpropagation algorithms, have been compared for transformer fault classification. Levenberg–Marquardt algorithm was designed to attend second-order training speed without calculating Hessian Matrix. It has been proved that Levenberg–Marquardt training algorithm provides superior performance than conventional Gradient Descent algorithm. In conjugate Gradient Descent algorithm, the step size is adjusted in every iteration. In resilient backpropagation algorithm, only the sign of the derivative is used to update the weight. The magnitude of the derivative has no effect on the weight update [33].

Five key gas ratios are considered as input to the neural network, and six output codes are treated as the output of the neural network. ANN of (5 × 18 × 6) is designed, and backpropagation algorithm is used to train the neural network. The dataset is divided into two categories such as training set (50 %) and testing set (50 %). The network parameters for backpropagation algorithm include the following.

Gradient =  $9.29 \times 10^{-6}$ ,  $\mu = 1 \times 10^{-6}$ , learning rate = 0.02, momentum factor = 0.8, number of neuron in hidden layer = 18 and tolerance = 0.005.

Figure 3 shows the error and epoch graph for different backpropagation learning algorithms. This figure gives a plot between the error value, i.e., the difference between the desired output and actual output, and the number of iterations it takes to reach to minimum error. From this figure, it is clear that Levenberg–Marquadt algorithm takes less number of iterations to converge to the required tolerance level. Figure 4 shows the regression (R) plot of different learning algorithms. It is plotted to measure the correlation between outputs and targets. An R value of one means a close relationship, zero a random relationship. The Regression Plot drawn in this figure shows that the value of R is closest to one indicating that output of the training network is quite close to the targets, when the dataset was trained with Levenberg–Marquardt.

There are some drawbacks of backpropagation training algorithm. It is too slow for practical applications, especially when too many hidden layers are employed. An appropriate selection of training parameters in the backpropagation algorithm is difficult and purely based on trial and error. There are many learning algorithms or modifications of the backpropagation algorithm in the literature but none of these methods are able to completely solve the problems associated with the backpropagation algorithm [34]. To overcome these

**Table 4** Confusion Matrix of Backpropagation Neural Network showing fault classification results of different algorithms

Algorithms	Fault Type →	PD	D1	D2	T1	T2	T3
	↓						
GD	PD	89	0	11	0	0	0
LM		100	0	11	0	0	0
CGD		100	0	0	0	0	0
RA		100	0	0	1	1	0
GD	D1	0	99	0	0	0	0
LM		0	97	0	0	0	0
CGD		0	84	0	4	0	0
RA		0	95	0	0	0	0
GD	D2	0	0	75	0	2	0
LM		0	0	98	2	4	0
CGD		9	0	85	0	1	1
RA		0	0	98	2	4	0
GD	T1	0	0	2	95	0	0
LM		0	0	0	95	0	0
CGD		2	0	0	98	0	0
RA		0	0	0	92	0	0
GD	T2	0	0	2	1	87	0
LM		0	0	0	0	90	1
CGD		0	0	1	0	95	0
RA		0	0	0	0	88	1
GD	T3	0	2	0	0	0	88
LM		0	0	0	0	1	88
CGD		1	0	0	0	0	99
RA		0	0	0	0	1	85

GD Gradient Descent Algorithm, LM Levenberg–Marquardt algorithm, CGD Gradient Descent Scaled Conjugate RA resilient, PD partial discharge, D1 = Low-energy discharge, D2 = High-energy discharge, T1 = Thermal faults < 300 °C, T2 = Thermal faults < 700 °C, T3 = Thermal faults > 700 °C

drawbacks of backpropagation algorithms, a new neural network called PNN classifier is used in this article. In 1990, Specht introduced PNN architecture as a three-layer feedforward neural network architecture. The layers in PNN are input layer, pattern layer and summation layer. PNN is the neural network implementation of Parzen window kernel discrimination analysis. PNN is implemented using probabilistic model. Unlike backpropagation algorithm, it is bound to converge. No learning process is required for PNN, and there is no need to set weight [35, 36].

Parzen’s windowing estimation is given by

$$\varphi_{ki}(x) = \frac{1}{(2\pi)^{d/2} \sigma^d} \sum_{i=1}^m \exp \left[ -\frac{(x - x_{ki})^T (x - x_{ki})}{2\sigma^2} \right] \quad (1)$$

Output of pattern layer is calculated as

**Table 5** Confusion matrix of Probabilistic Neural Network showing fault classification results

Fault type	PD	D1	D2	T1	T2	T3
Partial discharge (PD)	89	0	0	0	0	0
Low-energy discharge ( $D_1$ )	0	99	0	0	0	0
High-energy discharge ( $D_2$ )	0	0	98	0	0	0
Thermal faults < 300 °C ( $T_1$ )	0	0	0	95	0	0
Thermal faults < 700 °C ( $T_2$ )	0	2	0	0	96	0
Thermal faults > 700 °C ( $T_3$ )	1	0	0	0	0	97

**Table 6** Comparison of regression and accuracy amongst different intelligent methods

Methods	Regression (R)	Training recognition accuracy (%)
Gradient Descent	0.98128	88.3
Levenberg–Marquardt method (LM method)	0.98232	93.6
Conjugate Gradient	0.97905	93.5
Resilient BPN	0.97544	93
PNN	0.9875	95.6

$$\varphi_{ki}(x) = \frac{1}{(2\pi)^{d/2} \sigma^d} \exp \left[ -\frac{(x - x_{ki})^T (x - x_{ki})}{2\sigma^2} \right] \quad (2)$$

where  $x_{ki}$  is the neuron vector,  $\sigma$  is the smoothing parameter,  $d$  is the dimension of pattern vector  $x$ ,  $\varphi_{ki}$  is the output of pattern layer.  $T$  is the transpose of the distance between the neuron vector  $x_{ki}$  and pattern vector  $x$ .

Output of summation layer for  $k$ th neuron is

$$p_k(x) = \frac{1}{(2\pi)^{d/2} \sigma^d N_i} \exp \left[ -\frac{(x - x_{ki})^T (x - x_{ki})}{2\sigma^2} \right] \quad (3)$$

Here,  $N_i$  is total number of samples in  $k$ th neuron.

The output of decision layer is

$$c(x) = \arg \max \{p_k(x)\} \quad k = 1, 2, 3, \dots, m \quad (4)$$

$m$  denotes the number of classes in training sample;  $c(x)$  is estimated class of the pattern  $x$ .

**Table 7** Comparison of the classification results with the actual faults

Sample no.	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	Actual fault	BPN				PNN
							GD	LM	CGD	RA	
1	336	419	105	1,074	21	High-temp. overheat	×	√	×	√	√
2	160	130	33	96	0	Low-temp. overheat	×	√	√	×	√
3	57	77	58	21	0	Low-temp. overheat	√	×	√	√	√
4	565	93	34	47	0	Partial discharge	×	√	×	√	√
5	650	53	34	20	0	Partial discharge	√	√	×	×	×

*BPN* Backpropagation Neural Network, *PNN* Probabilistic Neural Network, *GD* Gradient Descent Algorithm, *LM* Levenberg–Marquardt algorithm, *CGD* Gradient Descent Scaled Conjugate, *RA* resilient

The accuracy of different BPN learning algorithms has been checked using confusion matrix. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see, if the system is confusing two classes (i.e., commonly mislabeling one as another).

Table 4 shows the confusion matrix for Backpropagation Neural Network and its various learning algorithms. Accuracy percentage, to classify different transformer fault types of the fault classification algorithms, i.e., Gradient Descent Algorithm, Levenberg–Marquardt algorithm, Gradient Descent Scaled Conjugate and Resilient, is 88.3, 93.6, 93.5 and 93 %, respectively. The accuracy percentage is higher, when LM method is used to classify different fault types of power transformer. Table 5 shows the confusion matrix for PNN and accuracy of fault classification using PNN comes out to be 95.6 %, which is an improved accuracy than backpropagation learning algorithms. To evaluate the performance of the classifiers to classify six fault classes, PNN is compared with different learning algorithms of backpropagation algorithm. The neural network is trained for 600 samples of training data (100 samples of each class). The network is further tested with 600 samples (100 samples of each class). Table 6 gives a comparative analysis of accuracy and regression among different learning algorithms of backpropagation method and PNN. From this table, it is seen that PNN is a better classifier than others. The Table 7 gives the classification results and compares the actual fault with the simulated fault, and from the comparison, it is concluded that PNN is a better classifier.

### 5 Conclusions

A comparative study of backpropagation algorithm and Probabilistic Neural Network algorithms has been carried out, to classify transformer faults, using these two algorithms. Highest accuracies are 95.6 % for Probabilistic

Neural Network classifier and 93.6 % for Backpropagation Network classifier (Levenberg–Marquardt method). The findings show that Probabilistic Neural Network classifier outperforms the Backpropagation Network classifier. The proposed technique (Probabilistic Neural Network classifier) is capable of identifying the faults even if the dissolved gas ratio data lie outside the specified ranges defined by conventional ratio methods. From simulation point of view, early convergence, no learning process and no need to set weight are the added advantages, which make Probabilistic Neural Network a very useful fault classification tool for power transformers.

**Acknowledgments** The authors are thankful to Dr. B. N. Chudasama, Assistant Professor, School of Physics and Material Science, Thapar University, for his valuable suggestions.

## References

- Ozgonenel O, Kilic E, Khan MA, Rahman MA (2008) A new method for fault detection and identification of incipient faults in power transformers. *Electric Power Compon Syst* 36:1226–1244
- Knapp GM, Jovadpour R, His-Pin Wang (2000) An ARTMAP neural network: based machine condition monitoring system. *J Qual Maint Eng* 6:86–105
- Yadaiah N, Ravi N (2011) Internal fault detection techniques for power transformers. *Appl Soft Comput* 11:5259–5269
- Parvin Darabad V, Vakilian M, Phung BT, Blackburn TR (2013) An efficient diagnosis method for data mining on single PD pulses of transformer insulation defect models. *IEEE Trans Dielectr Electr Insul* 20:2061–2072
- Liu J, Zheng K, Zhang H, Peng D (2013) A comparative research on power transformer fault diagnosis based on several artificial neural networks. *J Comput Inf Syst* 18:7501–7508
- Singh J, Sood YR, Jarial RK (2008) Condition monitoring of power transformers—bibliography survey. *IEEE Electr Insul Mag* 24:11–25
- Verma P, Singh J, Sood YR (2012) The influence of service aging on transformer insulating oil parameters. *IEEE Trans Dielectr Electr Insul* 19:421–426
- Bhalla D, Bansal RK, Gupta HO (2012) Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis. *Int J Electr Power Energy Syst* 43:1196–1203
- Rogers RR (1978) IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis. *IEEE Trans Electr Insul* 13:348–354
- Bhalla D, Bansal rk, Gupta HO (2013) Integrating AI based DGA fault diagnosis using Dempster-Shafer Theory. *Int J Electr Power Energy Syst* 48:31–38
- Liu CE, Ling JM, Huang CL (1996) An expert system for transformer fault diagnosis using dissolved gas analysis. *IEEE Trans Power Deliv* 8:231–238
- Malik H, Yadav AK, Mishra S, Mehto T (2013) Application of neuro-fuzzy scheme to investigate the winding insulation paper deterioration in oil-immersed power transformer. *Int J Electr Power Energy Syst* 53:256–271
- Huo-Ching Suna, Huang Yann-Chang, Huang Chao-Ming (2012) Fault diagnosis of power transformers using computational intelligence: a review. *Energy Procedia* 14:1226–1231
- Jayaswal P, Verma SN, Wadhvani AK (2010) Application of ANN, fuzzy logic and wavelet transform in machine fault diagnosis using vibration signal analysis. *J Qual Maint Eng* 16:190–213
- Zheng HB, Liao RJ, Grzybowski S, Yang LJ (2011) Fault diagnosis of power transformers using multi-class least square support vector machines classifiers with particle swarm optimization. *Electric Power Appl IET* 5:691–696
- Perera N, Rajapakse AD (2011) Recognition of fault transients using a probabilistic neural network classifier. *IEEE Trans Power Deliv* 26:410–419
- Islam Mofizul S, Wu T, Ledwich G (2000) A novel fuzzy logic approach to transformer fault diagnosis. *IEEE Trans Dielectr Electr Insul* 7:177–186
- Youssef OAS (2004) Combined fuzzy: logic wavelet—based fault classification technique for power system relaying. *IEEE Trans Power Deliv* 19:582–589
- Huang YC, Yang HT, Huang CL (1997) Developing a new transformer fault diagnosis system through evolutionary fuzzy logic. *IEEE Trans Power Deliv* 12:761–767
- Theofilatos K, Pylarinos D, Likothanassis S, Melidis D, Siderakis K, Thalassinakis E, Mavroudi S (2014) A hybrid support vector fuzzy inference system for the classification of leakage current waveforms portraying discharges. *Electric Power Compon Syst* 42:180–189
- Ganyun LV, Haozhong C, Haibao Z, Lixin D (2005) Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electr Power Syst Res* 74:1–7
- Wei CH, Tang WH, Wu QH (2014) A hybrid least-square support vector machine approach to incipient fault detection for oil-immersed power transformer. *Electric Power Compon Syst* 42:453–463
- Fei SW, Sun Y (2008) Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm. *Electr Power Syst Res* 78:507–514
- Alhoniemi E, Hollmén J, Simula O, Vesanto J (1999) Process monitoring and modelling using the self-organizing map. *Integr Comput Aided Eng* 6:3–14
- da Silva ACM, Garcez Castro AR, Miranda V (2012) Transformer failure diagnosis by means of fuzzy rules extracted from Kohonen Self-Organizing Map. *Int J Electr Power Energy Syst* 43(1034–1042):26
- Sun HC, Huang YC, Huang CM (2012) A review of dissolved gas analysis in power transformers. *Energy Procedia* 14:1220–1225
- Pahlavanapur B (1995) Power Transformer insulation ageing. *CIGRE SC 15 Symp., Sydney, Australia*
- Beniwal NS, Dwivedi DK, Gupta HO (2010) Creep life assessment of distribution transformers. *Eng Fail Anal* 17:1077–1085
- Srinivasan M, Krishnan A (2013) Effects of environmental factors in transformer's insulation life. *WSEAS Trans Power Syst* 8:35–44
- Duval M (2001) Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electr Insul Mag* 17:31–41
- Talavera L (2000) Dependency: based feature selection for clustering symbolic data. *Intell Data Anal* 4:19–28
- Dash M, Liu H (2013) Feature selection for classification. *Intell Data Anal* 1:131–156
- Sun YJ, Zhang S, Miao CX, Li JM (2007) Improved BP neural network for transformer fault diagnosis. *J China Univ Min Technol* 17:138–142
- Li S, Sun Y, Miao C, Feng Y (2006) Transformer fault diagnosis based on neural network of BPARM Algorithm. *Sixth World Congress on Intelligent Control and Automation vol 2*, pp 5734–5738

35. Paidarnia H, Hajiaghasi S, Abbaszadeh K (2013) New method for transformer fault diagnosis using Probabilistic Neural Network based on Principle Component Analysis. CIREC Regional, Iran, Tehran 13–14
36. Jiang SF, Fu C, Ziang C (2011) A hybrid data fusion system using modal data and probabilistic neural network for damage detection. *Adv Eng Softw* 42:368–374