An Integrated Neural Fuzzy Approach for Fault Diagnosis of Transformers

R. Naresh, Veena Sharma, and Manisha Vashisth

***Abstract*—This paper presents a new and efficient integrated neural fuzzy approach for transformer fault diagnosis using dissolved gas analysis. The proposed approach formulates the modeling problem of higher dimensions into lower dimensions by using the input feature selection based on competitive learning and neural fuzzy model. Then, the fuzzy rule base for the identification of fault is designed by applying the subtractive clustering method which is very good at handling the noisy input data. Verification of the proposed approach has been carried out by testing on standard and practical data. In comparison to the results obtained from the existing conventional and neural fuzzy techniques, the proposed method has been shown to possess superior performance in identifying the transformer fault type.**

***Index Terms*—Cluster centers, neural-fuzzy model, self-organizing network, subtractive clustering, transformer fault diagnosis.**

I. INTRODUCTION

THE POWER transformer is essential equipment of the electrical power system. Any fault in the power transformer may lead to the interruption of the power supply and accordingly, the financial losses will also be great. So it is of vital importance to detect the incipient fault of the transformer as early as possible. To monitor the serviceability of power transformers, many devices have evolved, such as Buchholz relays or differential relays. But the main shortcoming of these devices is that they only respond to the severe power failures which require removal of equipment from the service. Thus, techniques for early detection of the faults would be very valuable to avoid outages.

Among the existing methods for identifying the incipient faults, dissolved gas analysis (DGA) is the most popular and successful method [1]–[3]. When there is any kind of fault, such as overheating or discharge fault inside the transformer, it will produce a corresponding characteristic amount of gases in the transformer oil. This concept is the underlying principle of DGA. Through the analysis of the concentrations of dissolved gases, their gassing rates, and the ratios of certain gases, the DGA method can determine the fault type of the transformer. The commonly collected and analyzed gases are H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, and CO_2. An ANSI/IEEE standard and IEC publication 599 [4], [5] describes three DGA approaches: 1) key gas method; 2) Roger's ratio method; and 3) the Doernenberg ratio method. All three methods are computationally straightforward. However, these methods, in some cases, provide erroneous diagnoses as well as no conclusion for the fault type. The key gas method based on the determination of the key gas provides the basis for qualitative determination of fault types from the gases that are typical or predominant at various temperatures. Now, if the fault is very severe, then all of the gas concentrations will be high, yet insufficient to register a fault when using the values specified in IEEE standard [2]. Also, the gas ratios obtained for the particular transformer sample, may not fall within ANSI/IEEE-specified ranges, leading to the failure of the ratio methods for transformer diagnosis [6].

In recent years, many researchers have studied the application of artificial intelligence, such as neural networks and fuzzy set theory to increase diagnosis accuracy [6]–[15]. The fuzzy systems, though good at handling uncertainties, could not learn from previous diagnosis results and, hence, are not able to adjust the diagnostic rules automatically [10]–[13]. To account for uncertainties, the artificial neural networks (ANNs) have been proposed to diagnose the transformer's faults because of their superior learning capabilities [6]–[9]. In general, fuzzy systems and neural networks deal efficiently with two different areas of information processing. Fuzzy systems are good at various aspects of uncertain knowledge representation, while neural networks are efficient structures that are capable of learning from examples. Both techniques complement each other. The generalized regression neural network was used in [14] but since this network is a one-pass network, efficiency is somewhat low for fault detection. An application of fuzzy clustering and a radial basis function neural network was reported in [15]; however, when one type of fault is in the neighborhood of other types of faults, then the chances of false diagnosis increase. In this paper, a combination of neural network and fuzzy system is proposed for enhancing the performance of the diagnostic system.

The objective of this work is to develop an efficient neural fuzzy model for providing transformer diagnosis. The model is developed from the available data for the five fault types. The self-organizing network, based on the concept of competitive learning, operates by dividing input data into a suitable number of clusters. The neural fuzzy model then employs cluster centers information to rank the importance of these gases that are used to select the significant input features over insignificant ones. The resulting input fault gases data are used by the subtractive clustering method to extract the fuzzy rule base for identifying each class of fault data. The main advantage of the proposed

Manuscript received June 5, 2007; revised January 30, 2008. Current version published September 24, 2008. Paper no. TPWRD-00339-2007.

The authors are with the Electrical Engineering Department, National Institute of Technology Hamirpur, Hamirpur 177005, Himachal Pradesh India (e-mail: rnareshnith@gmail.com; veena@nitham.ac.in; manisha.vashisth@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2008.2002652

0885-8977/$25.00 © 2008 IEEE

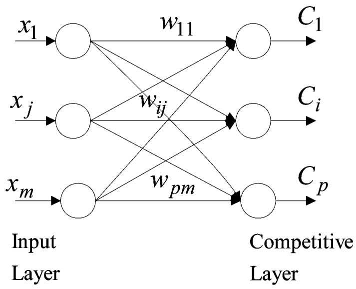

Fig. 1. Self-organizing network.

approach over others is that the fuzzy rule generation process of subtractive clustering employed in this work is dynamic in that the variation in radius parameter changes the rule base and the diagnosis accuracy. Therefore, an optimal value of this parameter can be chosen to obtain the better diagnosis. Also, when little information is available about an optimal number of cluster centers, the proposed approach is advantageous over other optimization and clustering algorithms as an adverse effect of outliers on diagnostic accuracy is decreased significantly.

This paper is organized as follows: In Section I, there is a brief introduction of the topic; followed by feature selection algorithm using the self-organizing network in Section II. In Section III, methods of diagnosis, conventional, and existing neural-fuzzy approaches and proposed techniques using subtractive clustering are presented. In Section IV, the entire model formulation is summarized in one algorithm. The results of the testing the proposed diagnosis system on DGA data of power transformers and a comparison of this proposed technique with conventional neural fuzzy methods are dealt with in Section V, followed by the conclusion in Section VI.

II. Feature Selection

From the training data selected from [16], it is observed that for the major five fault types, the gases dissolved in the oil are hydrogen, methane, acetylene, ethylene, ethane, carbon monoxide and carbon dioxide. The first step for developing the neural fuzzy model is feature selection. In feature selection the gases which are most important for the diagnosis of the major faults are obtained. Feature selection for the major faults has been carried out by using the neural fuzzy model [17] is briefly described in the following paragraph.

A. Generation of Clusters

An algorithm based on the self-organizing Kohonen network has been developed in this work [17] for the input fault pattern selection. The structure of the network is shown in Fig. 1. It has one input layer and one competitive layer. In the input layer, there is m number of neurons according to the size of input patterns. In a competitive layer, the number of neurons is generated dynamically by selecting the appropriate value of threshold. All of the input fault data patterns are divided into p number of clusters using the unsupervised learning algorithm in the competitive layer. The weights of the network are adjusted according to the distance calculated between the input pattern and the previously present weights.

Notations $n =$ number of input patterns in m dimensional space; $X_k = (x_{k1}, x_{k2}, \ldots, x_{km})$ is the kth training pattern; C_i is the ith cluster center; Nsi is the number of fault points in the ith cluster; $W_i = (w_{i1}, w_{i2}, \ldots, w_{im})$ is the weight vector of the ith neuron in the competitive layer; $\alpha_0 \in [0, 1]$ is the initial learning rate; p is the total neurons created dynamically in the competitive layer.

First, the input pattern is presented to the input layer of the self-organizing network and then its Euclidean distances from the weight vectors of competitive layer neurons are calculated. The set of neurons having distances less than the predefined threshold is identified. If there is only one neuron in the set, then it is the winner; otherwise, the neuron with a lower index will be the winner [17]. Further, a new neuron is created in the competitive layer if all of the calculated Euclidean distances are larger than the predefined threshold. If a higher value of threshold is chosen, then the number of clusters generated will be less and the results obtained may be vague because of underfitting. And if a smaller value of threshold is chosen, then it would lead to overfitting by generating a large number of clusters. So it is very important to choose the appropriate value of threshold. Here, by testing various values of threshold on the model, the value which provides the best results is selected as the threshold value. If neuron I is the winner for input pattern X_k, then it belongs to cluster I and the weight vector of that neuron is modified as

$$W_I = W_I + \alpha \| X_k - W_I \|$$

where α is the learning rate which is determined by $\alpha = \alpha_0/(N_{SI} + 1)$, where N_{SI} is the number of input patterns belonging to cluster I.

The DGA data taken from [16] were divided into two types (i.e., training and testing data). Out of 117 random samples, 87 samples were selected for training and the remaining 30 were left for testing. The training data were first normalized in the range [0, 1]. After applying data to this self-organizing network and selecting the threshold value as 0.1, $p = 69$ clusters were formed dynamically.

B. Initial Fuzzy Model Derived From Competitive Learning

Here, each cluster center C_i is considered as a fuzzy rule that describes the system local behavior. Intitutively, cluster center C_i represents the rule:

"If input is around C_i, then input belongs to cluster i"

We can also represent this rule in terms of a fuzzy inference system, employing traditional fuzzy IF-THEN rules

$$R_i: \textbf{ If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2}, \ldots, \text{ and } x_m \text{ is } A_{\mathrm{im}} \textbf{ Then } \text{cluster is } i.$$

where R_i denotes the ith rule; and A_{ij} is the Gaussian membership function in the ith rule. The membership function A_{ij} is given by

$$A_{ij} = \exp\left\{ -\left(\frac{x_{kj} - c_{ij}}{\sigma_j} \right)^2 \right\} \quad (1)$$

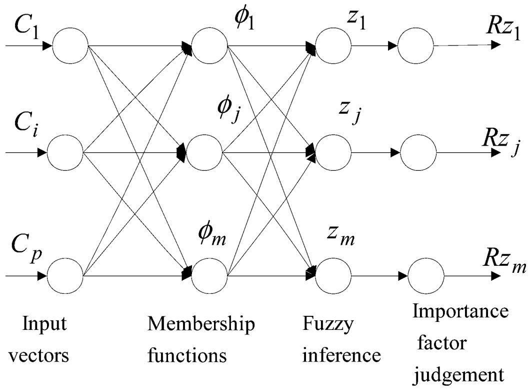

Fig. 2. Neural fuzzy model for feature selection.

and, consequently, the rule is given as

$$y_i^* = \max(A_{ij}) \tag{2}$$

where $i = 1, 2, \ldots, p; j = 1, 2, \ldots, m; k = 1, 2, \ldots, n; c_{ij}$ is the jth element of the ith cluster center and σ_j is the span in the input data. The degree of fulfillment of each rule is computed by using multiplication as the AND operator, and because the fuzzy system performs classification, we simply select the consequence of the rule (i.e., y^*) with the highest degree of fulfillment to be the output of the fuzzy system. Thus, these $p = 69$ clusters form the initial fuzzy model with $p = 69$ rules that are used for selecting the important features.

C. Significant Input Feature Selection

From the modeling point of view, the variables which make the model simpler, more useful, reliable, and more practical to apply are incorporated into the model. The initial fuzzy model discussed in Section II-B is employed for determining the importance of each fault gas. It is known that the change of the system output is contributed to by all input variables but the larger the output change caused by a specified input variable, the more important this input variable is. The importance of all m input variables which, in the present problem, are seven fault gases can be tested simultaneously under a given predefined index by using a simplified fuzzy inference neural-network model which can generate in parallel all fuzzy outputs with respect to every individual input variable. The structure [17] of neural fuzzy inference is shown in Fig. 2.

The model is a three-layer feedforward network. Unlike common neural fuzzy models, the input and output of each neuron are vectors. The prototypes of the p clusters $\{C_1, C_2, \ldots, C_p\}$ form a matrix C which is generated by competitive learning and stores the input patterns for a neural fuzzy model for feature selection. The first layer is called the fuzzification layer. The activation function of each neuron consists of a set of membership functions, i.e., $\phi_k = (\phi_{1k}, \phi_{2k}, \ldots, \phi_{Pk})$, where ϕ_{ik} is the membership function of the kth fuzzy subset of the ith input variable, which is defined as

$$\phi_{ik} = \exp\left\{ -\left(\frac{c_{ik} - c_{ij}}{\sigma_j'} \right)^2 \right\}$$
$$i = 1, 2, \ldots, p; j = 1, 2, \ldots, m;$$
$$k = 1, 2, \ldots, m \tag{3}$$

where c_{ij} and c_{ik} are elements of matrix C, and σ_j' is the spread or span in the cluster centers.

The basic idea is that all antecedent clauses are assigned the value 1 except for one dominant testing input variable, then fuzzy inference using multiplication as the AND operator and defuzzification using the center of gravity algorithm can be merged into one procedure which is implemented by the second layer of the network.

In this layer, the output of the kth neuron $z_k = (z_{1k}, z_{2k}, \ldots, z_{Pk})$ denotes the fuzzy inference output corresponding to the contribution of the kth input variable. The output vector is computed by

$$z_{ik} = \frac{\sum_{i=1}^{p} \phi_{ik}(c_i) y_i^*}{\sum_{i=1}^{p} \phi_{ik}(c_i)}$$
$$i = 1, 2, \ldots, p; k = 1, 2, \ldots, m \tag{4}$$

where y_i^* is obtained from (2).

On the basis of m fuzzy output vectors, the importance of input variables can be recognized by calculating the change range of corresponding z_k, which can be obtained in the output layer of the network by $Rz_k = \max(z_k) - \min(z_k)$.

The input selection is carried out according to the following steps.

Step 1) Define the importance factor of the kth input by

$$F_k = Rz_k / R_m \tag{5}$$

where $R_m = \max\{Rz_1, Rz_2, \ldots, Rz_m\}$.

Step 2) Rank the importance of all input variables according to their corresponding F_k values.

Step 3) Remove all input variables with respect to $F_k < \lambda$, where $\lambda \in (0, 1)$ is the predefined threshold.

Obviously, $F_k = 1$ corresponds to the most important input variable, the large varying range of the fuzzy output R_k indicating the big influence of the corresponding input variable. A small value of F_k corresponds to a relatively unimportant input. When F_k is less than the threshold (i.e., $F_k < \lambda$), the corresponding input variable is believed to be unimportant and can be removed. Assume that there are r inputs with the values of $F_k > \lambda$; thus, a collection of r inputs are selected from m input variables.

In this paper, the importance factors for different input variables obtained on the basis of aforesaid clusters are shown in Table I. On the basis of trial-and-error simulation runs, the most suitable value for the threshold was considered to be 0.5 and using this value, the gases carbon dioxide and carbon monoxide were discarded. The gases carbon dioxide and carbon monoxide are helpful in cellulose degradation [18], [19] and their effect on DGA is less prominent. Also from Table I, it is evident that the importance factor of both gases is low, so these gases have been discarded in feature selection. The value of the threshold, if taken to be large and close to 1.0, results in the elimination of important features which leads to poor diagnosis. And if the threshold that is taken is very small, then all features remaining in the network, would further lead to an increase in dimensionality of the network, and the partitioning of the fault region

TABLE I
IMPORTANCE FACTOR (IF) FOR INPUT VARIABLES (IV)

IV	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2
IF	0.6729	0.6841	0.5755	0.6914	1.0000	0.0435	0.3347

would become difficult, because the region of overlap will be more among different fault types.

Thus, when the problem of feature selection is complete, we prune the training data to include only five significant input gases, namely: 1) hydrogen; 2) methane; 3) acetylene; 4) ethylene; and 5) ethane.

III. METHODS OF DIAGNOSIS

A. Module 1: Conventional Techniques

1) Rogers Ratio Method: This method is based on three gas ratios. The algorithm is based on IEEE C57.104-1991 standard [4]. The simulation code was developed in Matlab.

2) Fuzzy C Means Method: Fuzzy C-means (FCM) is the data clustering technique which suggests that each input pattern belong to the cluster which is defined by its membership function. This technique was defined first by J. Bezedek [12]. In this technique, the n number of input patterns having m dimensions is clustered in c number of clusters. These c different clusters are defined as faults for the DGA as described in [13].

3) Generalized Regression Neural Network: The generalized regression neural network (GRNN) has been found to be advantageous in solving a great variety of difficult mapping and prediction problems. In this network, the input patterns are distributed in all of the pattern units. These pattern units have the same number as that of the input patterns. Now, if a new input pattern is given to the network, it will calculate its distance from the respective cluster centers and this distance after multiplication with the bias will be applied to the radial basis function and, finally, the response of the network can be obtained [14].

4) Fuzzy Clustering and Radial Basis Function Neural Network: The input patterns in this case are normalized by a fuzzy membership function called sigmoidal function. And then, using the fuzzy clustering for selecting the efficient training data and finally the faults were determined using the radial basis function neural network (RBFNN) [15].

B. Module 2: Proposed Technique

1) Subtractive Clustering Method: Subtractive clustering has been described in detail in [20]–[22]. Subtractive clustering is a very fast and efficient clustering method designed for a moderate number of input patterns, because its computation grows linearly with the data dimension and as the square of the number of data points. The subtractive clustering method is available in the fuzzy logic toolbox for MATLAB [23].

In the subtractive clustering method, the training data are divided according to their respective class labels and then the subtractive clustering algorithm is applied on each group of data individually to extract the rules for identifying each class of data. Let a group of nc data points $\{X_1, X_2, \ldots, X_{nc}\}$ be specified for a particular class in the feature space. The first step in subtractive clustering is to normalize the data in the feature space in the range [0, 1]. Each data point in the class is considered a potential cluster center and the measure of the potential of data point X_i to serve as a cluster center is defined as

$$\Omega_i = \sum_{j=1}^{nc} \exp(-\delta \|X_i - X_j\|^2); \quad i = 1, 2, \ldots, nc; \tag{6}$$

where

$$\delta = \frac{4}{r_a^2}. \tag{7}$$

$\|\cdot\|$ denotes the Euclidean distance, and the positive constant r_a is effectively a normalized radius defining a neighborhood. The data points outside this radius have little influence on the potential of the data points within this radius. Thus, the measure of the potential of a data point is a function of its distances to all other data points. A data point with many neighboring data points will have a high potential value. After the potential of every data point has been computed, we select the data point with the highest potential as the first cluster center.

Let X_1^* be the location of the first cluster center with Ω_1^* as its potential value. Now, the potential of each data point X_i is revised by the formula

$$\Omega_i \Leftarrow \Omega_i - \Omega_1^* \exp\left(-\gamma \|X_i - X_1^*\|^2\right)$$

where

$$\gamma = \frac{4}{r_b^2} \tag{8}$$

and r_b is a positive constant. Thus, there is a subtraction of an amount of potential from each data point as a function of its distance from the first cluster center. Thus, the potential of the data points near the first cluster center will be greatly reduced and, therefore, their possibility of getting selected as the next cluster center will be eliminated. The constant r_b is an effective radius defining the neighborhood which will have measurable reductions in potential. To avoid obtaining closely spaced cluster centers, choose $r_b = 1.25 r_a$.

When the potential of all data points has been reduced according to (8), then select the data point with the highest remaining potential as the second cluster center. Also, the potential of each data point is reduced according to its distance to the second cluster center.

The whole procedure is generalized as follows: After the kth cluster center has been obtained, the potential of each data point is revised by the formula

$$\Omega_i \Leftarrow \Omega_i - \Omega_k^* \exp\left(-\gamma \|X_i - X_k^*\|^2\right) \tag{9}$$

where X_k^* is the location of the kth cluster center and Ω_k^* is its potential value.

The process of acquiring a new cluster center and reducing potential repeats until the remaining potential of all data points are below some fraction of the potential of the first cluster center Ω_1^*; typically using $\Omega_k^* < 0.15\Omega_1^*$ as the stopping criterion.

2) Generating Fuzzy Rule Base From Clusters: Fuzzy rules are obtained from cluster centers generated by subtractive clustering for identifying a particular class. Suppose 1 number of cluster centers are obtained for class $c1$. Consider cluster center C_k^* in the group of data for class $c1$, this cluster center is translated into the rule

$$\text{Rule } k: \text{ If } \{X. \text{ is near } C_k^*\} \text{ then class is } c1.$$

The degree of fulfillment of $\{X. \text{ is near } C_k^*\}$ is defined as

$$\mu_k = \exp\left(-\delta \left\|X - C_k^*\right\|^2\right) \tag{10}$$

where $\delta = 4/r_a^2$ and $k = 1, 2, \ldots, l$. r_a is the radius selected for subtractive clustering. We can also write this rule in the more familiar form

$$\text{Rule } k: \text{ If } x_1 \text{ is } \Lambda_{k1} \;\&\; x_2 \text{ is } \Lambda_{k2} \;\&\; \ldots \text{ then class is } c1.$$

where x_j is the jth input feature and Λ_{kj} is the membership function in the kth rule associated with the jth input feature (j varies in the input feature dimensions). The membership function Λ_{kj} is given by

$$\Lambda_{kj}(x_j) = \exp\left\{-\frac{1}{2}\left(\frac{x_j - c_{kj}^*}{\sigma_{kj}^*}\right)^2\right\} \tag{11}$$

where c_{kj}^* is the jth element of C_k^*, and σ_{kj}^* is the spread. The degree of fulfillment of each rule is computed by using multiplication as the AND operator.

Thus, by using subtractive clustering, a set of rules for identifying each individual class of data is obtained. The individual sets of rules are then combined to form the rule base of the classifier.

3) Membership Function Optimization: For achieving the satisfactory modeling accuracy, the optimization of model parameters under a given performance index is required. To find the optimal values of individual c_{kj}^* and σ_{kj}^* parameters in the membership functions (11), the classification error measure is minimized using the neural-network back propagation algorithm [18], [20]. The performance index as a classification error measure for a data sample that belongs to some class $c1$ is defined as

$$E = \frac{1}{2}(1 - \mu_{c1,\max} + \mu_{c1\to\max})^2 \tag{12}$$

where $\mu_{c1,\max}$ is the highest degree of fulfillment among all rules that infer class $c1$, and $\mu_{c1\to\max}$ is the highest degree of fulfillment among all rules that do not infer class $c1$. Note that this error measure is zero only if a rule that would correctly classify the sample has a degree of fulfillment of 1 and all rules that would misclassify the sample have a degree of fulfillment of 0.

The membership function parameters (11) are updated according to the following formulae:

$$\left.\begin{array}{l} c_{kj}^* \Leftarrow c_{kj}^* - \lambda \frac{\partial E}{\partial c_{kj}^*} \text{ and} \\ \sigma_{kj}^* \Leftarrow \sigma_{kj}^* - \lambda \frac{\partial E}{\partial \sigma_{kj}^*} \end{array}\right\} \tag{13}$$

where λ is a positive learning rate in the back propagation algorithm.

In this study, the pruned training data mentioned at the end of the Section II-C are divided into five fault types. Then, subtractive clustering was applied on each fault type. Taking the radius value r_a as 0.5, 41 cluster centers are obtained. These cluster centers translate into an equal number of fuzzy rules and the fuzzy rule base has been generated. In the fuzzy rule base, four rules for partial discharge, eight rules for discharge of low energy, 13 rules for discharges of high energy, nine rules for low thermal faults, and seven rules for high thermal faults have been obtained.

In Fig. 3, the membership functions for partial discharge are shown. These membership functions were optimized using a two-layer neural network back propagation algorithm with the learning rate λ value chosen as 10^{-6}. The network has five neurons in the input layer and 41 neurons in the second layer with cluster centers acting as the weight vectors. The four rules thus formed for partial discharge are obtained from subtractive clustering. The degree of fulfillment of each rule is obtained by multiplication and the AND operator. Whenever the degree of fulfillment of any of these four rules is highest among the complete fuzzy rule base, the transformer is then diagnosed with a partial discharge fault. The x-axis in Fig. 3 represents the normalized universe of discourse (i.e., input range of 0 to 1.0) and the y-axis represents the membership grade, also in the range 0 to 1.

In the same manner, the membership function for other faults type is obtained and optimized. The membership functions for all five faults represent the complete rule base.

IV. Main Algorithm

The feature selection and derivation of the fuzzy rule base have been described in Sections II and III-B. Here, the stepwise details of the proposed integrated neural fuzzy approach for transformer fault diagnosis are presented as follows.

Step 1) The input fault data are divided into training and testing data.

Step 2) Evaluate cluster centers for the training data by competitive learning.

Step 3) Calculate antecedents and consequents using the initial fuzzy model.

Step 4) Using the neural fuzzy model, find significant input features.

Step 5) Apply subtractive clustering to selected training feature data to obtain cluster centers.

Step 6) Develop the fuzzy rule base using cluster centers obtained in Step 5).

Step 7) Optimize the membership functions of the fuzzy rule base using the back propagation algorithm.

Step 8) Apply the model that has been developed in Steps 1) to 7) on testing fault data and print the results.

V. Test Results and Discussion

After network formulation is over, the proposed model has been tested on two test cases for checking the validity of feature selection and its superiority over the conventional and existing neural fuzzy methods, such as Roger's ratio method, FCM, GRNN, and fuzzy clustering and RBFNN. The two test cases include one standard test data from [16] and other practical fault data from working transformers of the Himachal Pradesh

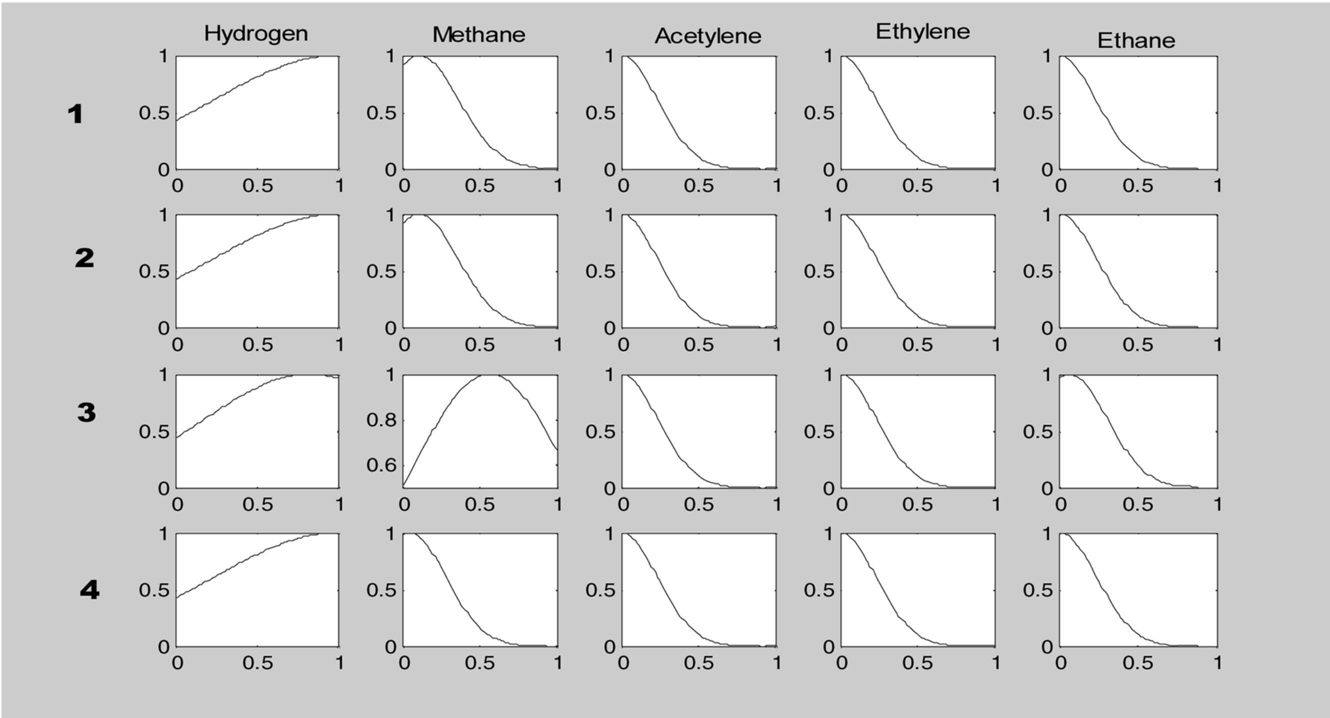

Fig. 3. Rules for partial discharge.

State Electricity Board (H.P.S.E.B.), a part of the Northern Power Grid of India. All of the coding has been performed in MATLAB 7.0, according to algorithmic steps given in Section IV.

Test Example 1: Based on the 117 data samples, out of which 87 samples were selected randomly for training to take into account all possible variations in faults present in the oil during diagnosis of transformers and the remaining 30 samples were selected for testing. In the training data, fault gases are H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, and CO_2. These are the input features. In feature selection, the training data are first normalized between [0, 1]. And these normalized data are then applied to the self-organizing network. Using the threshold as 0.1, 69 clusters were formed dynamically. These cluster centers are then converted to an equal number of fuzzy rules. Finally, the importance of each feature is obtained by using the information obtained from the cluster centers using a three-layer feedforward network. The importance of carbon monoxide and carbon dioxide was found to be less than the threshold value (i.e., 0.5), which was selected on the basis of a trial-and-error run. After feature selection, only first five fault gases are selected and considered to be important for the diagnosis of the five faults present in the training data. This developed model has also been tested on the same testing patterns without a feature selection concept which means that all seven input gas features are taken into account and the results have been produced in Table II. It has been observed that the rate of misclassification is very high with all of the input variables present compared to selected variables. With all input variables present, 15 samples out of 30 were misclassified in the testing data. Also, it has been seen in the analysis that with selected variables, out of 30 samples, only 1 was misclassified in the testing data, yielding a 96.67% success rate. Thus, it is concluded that the neural fuzzy model, when trained with the selected variables, gives a better diagnosis compared to the neural fuzzy model with all input variables present.

TABLE II
PROPOSED MODEL DIAGNOSIS WITH AND WITHOUT FEATURE SELECTION

Fault types	Number of testing patterns	Fault detected with proposed model (without feature selection)	Fault detected with proposed model (with feature selection)
PD	3	3	3
LED	9	3	8
HED	10	5	10
LTF	4	1	4
HTF	4	3	4

The following notations for the fault types are considered:

$$\begin{aligned} \text{PD} &= \text{Partial discharge} \\ \text{LED} &= \text{Low energy discharge} \\ \text{HED} &= \text{High energy discharge} \\ \text{LTF} &= \text{Thermal faults } < 700\,^\circ\text{C} \\ \text{HTF} &= \text{Thermal faults } > 700\,^\circ\text{C} \\ \text{ND} &= \text{No diagnosis.} \end{aligned}$$

On these input patterns, other models, such as Roger's ratio method [2], FCM [12], [13]; GRNN [14]; and fuzzy clustering and RBFNN [15] were applied. The comparison of the results obtained for the testing patterns using existing neural fuzzy methods and the proposed approach are shown in Fig. 4.

As is evident from Fig. 4, 50% model accuracy is achieved with FCM, because the various fault points corresponding to input training data may be lying in the overlapping regions and,

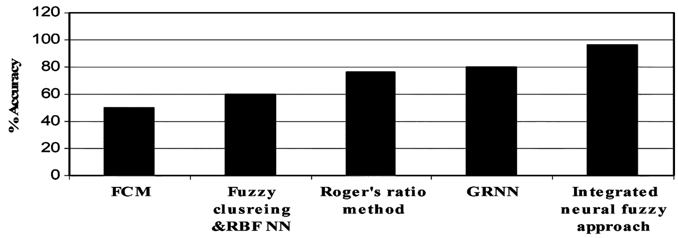

Fig. 4. Comparison of the proposed approach with the existing models.

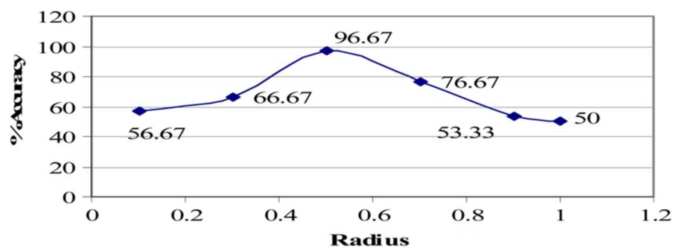

Fig. 5. Comparison of the proposed model for different values of radii.

TABLE III
GAS CONCENTRATIONS OF OIL SAMPLES

S. No.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2
1	14	92	279	156	31	129	1211
2	2	14	<1	11	47	84	2095
3	7	2	<1	<1	1	48	1241
4	4	27	<1	27	40	205	3746
5	10	3	2	2	1	43	467
6	52	132	406	206	62	377	1657

TABLE IV
DIAGNOSIS RESULTS OF TRANSFORMERS

S. No.	FCM	Fuzzy Clustering and RBF NN	Roger's Ratio Method	GRNN	Proposed Integrated Neural Fuzzy Model
1	PD	LED	ND	HED	HED
2	HTF	LTF	LTF	HTF	LTF
3	PD	HED	ND	HED	LED
4	HTF	LTF	LTF	HTF	HTF
5	LTF	HED	LED	HED	HED
6	HED	HED	LED	LED	HED

hence, the results obtained are highly inaccurate. In fuzzy clustering and RBFNN, the accuracy achieved is 60%, because the training data are very noisy and if one type of fault is very close to the other type of fault, then there is a prominent likelihood of erroneous diagnosis.

In the Roger's ratio method, the accuracy achieved is 76.67%. The Roger's ratio method, though accurate, sometimes tends to have no diagnosis. The main reason is that in this method, the obtained ratios may not match the standard ratios. In other words, we can say that the Roger's ratio method is not able to cover the entire input space. The generalized regression neural network has an accuracy of 80%. The accuracy is higher than other methods as each input pattern here is considered to be a separate cluster center and the overlapping problem is efficiently handled. But it is a one-pass network and that is why efficiency is somewhat low.

It has been shown in Table II that with the proposed integrated neural fuzzy network, 96.67% accuracy has been achieved. Such a high success rate of the proposed neural fuzzy approach has been made possible by a suitable choice of radius parameter r_a in the subtractive clustering algorithm. In this clustering approach, the number of fuzzy rules generated is equal to the number of cluster centers formed. It may be noted that the fuzzy rule generation process via subtractive clustering is dynamic in nature. This is because the number of fuzzy rules generated in this process is sensitive to the value of parameter r_a and, hence, its optimal selection is important for better results. Therefore, subtractive clustering algorithm has been simulated for various values of r_a in the range 0.0 to 1.0. The results so obtained in terms of efficiency of the algorithm on the testing data are plotted in Fig. 5.

From Fig. 5, it clear that the judicious selection of parameter r_a is required for achieving better transformer diagnostic results with the proposed approach, as smaller r_a results in underfitting and larger r_a leads to overfitting. From the aforementioned discussion, it is clear that when little information is available regarding the optimal number of clusters, then the proposed technique is advantageous over optimization and other classes of clustering algorithms. Also, this technique is noise robust, so the effect of outliers on the choice of cluster centers is not significant. The neural-network back propagation algorithm has been used for improving the results by optimizing the membership functions for various fault types.

Test Example 2: The proposed neural fuzzy model has now been tested on the H.P.S.E.B. working transformers oil samples. The concentrations for the faulty transformers as detected by the laboratory equipment Transport X, which is a portable DGA [24], are given in Table III. The Transport X has an embedded software package which provides transformer fault diagnosis by three methods, namely, the key gas method, Roger's ratio method, and Duval's triangle method. The incorporated software contains instructions to guide the user through the operation of the system and algorithms to assist the diagnosis of the equipment. The algorithms for key gas and Roger's ratio method are based on the IEEE C57.104-1991 [4] standard, and the algorithm for Duval's triangle method is based on TechCon 2004-Michel Duval [25]. The results of testing the integrated neural fuzzy model on the H.P.S.E.B. working power transformers oil samples are given in Table IV.

From the portable DGA, the concentration levels of the gases were obtained and using those values of DGA gas concentrations, the diagnoses are as follows.

In Table IV, the results of fault diagnosis obtained by the portable DGA by Roger's ratio method and the proposed neural fuzzy model are provided. Also, these results are compared with FCM, fuzzy clustering and RBFNN and GRNN. From the results, it is evident that the proposed neural fuzzy model, when applied to the practical data, gives a better diagnosis compared to the conventional Roger's ratio and other existing neural fuzzy techniques.

VI. CONCLUSION

A reliable and efficient integrated neural fuzzy fault diagnosis approach has been developed and implemented in this paper.

This neural fuzzy model was formulated by applying the feature selection concept on the training data and thereafter on the selected input features, the subtractive clustering technique was applied and the desired rules were obtained which were further optimized using the neural-network back propagation algorithm. After completion of training, the rule base was applied on the two testing data. The results obtained from the proposed neural fuzzy model were then compared with the conventional Roger's ratio method and the existing neural fuzzy approaches. The comparison with the different conventional methods leads to the observation that the proposed approach was successfully tested and provided better results. This may be due to the reason that the entire fault input space has been covered for fault diagnosis and analysis. Another important aspect of input feature selection has also been incorporated in this developed model and it has been observed that the results obtained with feature selection were better than without feature selection.

Acknowledgment

The authors would like to thank the management of TIFAC-CORE project on "Power Transformer Diagnostics," at NIT Hamirpur (H.P.) India, for providing the necessary support, facilities, and encouragement to carry out this work. We are also grateful to anonymous reviewers for useful comments.

References

[1] J. J. Kelly, "Transformer fault diagnosis by dissolved-gas analysis," *IEEE Trans. Artif. Intell.*, vol. 16, no. 6, pp. 777–782, Nov. 1980.

[2] R. R. Rogers, "IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis," *IEEE Trans. Elect. Insul.*, vol. EI-13, no. 5, pp. 348–354, Oct. 1978.

[3] M. Duval, "Dissolved gas analysis: It can save your transformer," *IEEE Elect. Insul. Mag.*, vol. 5, no. 6, pp. 22–27, Nov./Dec. 1989.

[4] "IEEE guide for the interpretation of gases generated in oil-immersed transformers," *IEEE Power Eng. Soc.*, 1992, ANSI/IEEE Std. C57.104-1991.

[5] IEC Publ. 599, "Interpretation of the analysis of gases in transformers and other oil-filled electrical equipment in service,"1st ed. 1978.

[6] S. S. Kalyani, "ANN approach for condition monitoring of transformer using DGA," in *Proc. IEEE Region 10 Conf.*, 2004, vol. 3, pp. 444–447.

[7] J. L. Guardado, J. L. Naredo, P. Moreno, and C. R. Fuerte, "A comparative study of neural network efficiency in power transformer diagnosis using dissolved gas analysis," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 643–647, Oct. 2001.

[8] V. Miranda and A. R. G. Castro, "Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks," *IEEE Trans. Power Del.*, vol. 20, no. 4, pp. 2509–2516, Oct. 2005.

[9] Y.-C. Huang, "Evolving neural nets for fault diagnosis of power transformers," *IEEE Trans. Power Del.*, vol. 18, no. 3, pp. 843–848, Jul. 2003.

[10] K. Tomsovic, M. Tapper, and T. Ingvarsson, "A fuzzy information approach to integrating different transformer diagnostic methods," *IEEE Trans. Power Del.*, vol. 8, no. 3, pp. 1638–1646, Jul. 1993.

[11] Q. Su, C. Mi, L. L. Lai, and P. Austin, "A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer," *IEEE Trans. Power Del.*, vol. 15, no. 2, pp. 593–598, Apr. 2000.

[12] J. C. Bezedek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[13] A.-P. Chen and C.-C. Lin, "Fuzzy approaches for fault diagnosis of transformers," *Fuzzy Sets Syst. 118*, pp. 139–151, 2001.

[14] D. R. Morais and J. G. Rolim, "A hybrid tool for detection of incipient faults in transformers based on the dissolved gas analysis of insulting oil," *IEEE Trans. Power Del.*, vol. 21, no. 2, pp. 673–680, Apr. 2006.

[15] J. P. Lee, D. J. Lee, P. S. Ji, J. Y. Lim, and S. S. Kim, "Diagnosis of power transformer using fuzzy clustering and radial basis function neural network," in *Proc. Int. Joint Conf. Neural Networks*, Vancouver, BC, Canada, 2006, pp. 1398–1404.

[16] M. Duval, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," *IEEE Elect. Insul. Mag.*, vol. 17, no. 2, pp. 31–41, Mar/Apr. 2001.

[17] D. A. Linkens and M.-Y. Chen, "Input selection and partition validation for fuzzy modeling using neural network," *Fuzzy Sets Syst. 107*, pp. 299–308, 1999.

[18] M. Dong, Z. Yan, and G. J. Zhang, "Comprehensive diagnostic and aging assessment method of solid insulation in transformer," in *Proc. Conf. Electrical Insulation Dielectric Phenomenon, Annu. Rep.*, 2003, pp. 137–140.

[19] V. G. Arakelian, "Effective diagnostics for oil filled equipment," *IEEE Elect. Insul. Mag.*, vol. 18, no. 6, pp. 26–38, Nov./Dec. 2002.

[20] S. Chiu, "Method and software for extracting fuzzy classification rules by subtractive clustering," in *Proc. Biennial Conf. North American Fuzzy Information Processing Soc.*, 1996, pp. 461–465.

[21] S. Chiu, "An efficient method for extracting fuzzy classification rules from high dimensional data," *J. Advanced Comput. Intell.*, vol. 1, no. 1, pp. 1–7, 1997.

[22] S. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 267–278, 1994.

[23] MATLAB 7.0, Fuzzy Logic Toolbox 2.1.3 The Mathworks. Natick, MA.

[24] *"Transport X, Portable Dissolved Gas Analyzer,"* Kelman Ltd., Lisbon, Ireland, version 1.4, 40-0069-01.

[25] M. Duval, "Dissolved gas analysis: A powerful maintenance tool for transformers," presented at the TechCon, San Antonio, TX, Jan. 25–29, 2004.

R. Naresh received the M.E. degree in power systems from Punjab Engineering College, Chandigarh, India, in 1990 and the Ph.D. degree from the Indian Institute of Technology, Roorkee, India, in 1999.

Since 2007, he has been a Professor in the Electrical Engineering Department with the National Institute of Technology Hamirpur, where he has been since 1989. His research interests are artificial-intelligence applications to power system optimization problems, transformer diagnostics, evolutionary computation, neural networks, and fuzzy systems.

Veena Sharma received the B.Tech. degree in electrical engineering from the National Institute of Technology Hamirpur, Hamirpur, India, in 1990 and the M.Tech. degree in instrumentation and control engineering from Punjab Agricultural University, Ludhiana, India, in 1993.

In 1994, she joined the Electrical Engineering Department as a Lecturer at the National Institute of Technology Hamirpur, where she has been Assistant Professor since 2007. Her current research interests include the applications of artificial neural optimization networks to power system optimization and control.

Manisha Vashisth received the B.Tech. degree in electrical engineering from the National Institute of Technology (NIT) Hamirpur, Hamirpur, India, and the M.Tech. degree in advanced power system and control from NIT Hamirpur.

Currently, she is a Lecturer in the Electrical Engineering Department at NIT Hamirpur. Her area of interest is power transformer diagnostics.