# Discrimination Between inrush and fault Condition in transformer : A Probabilistic Neural Network Approch

Roshankumar U.khedkar
Dept. of Electrical Engineering
Shri Sant Gajanan Maharaj College of Engineering
Shegaon, Dist-Buldhana, 444203

Prof.R.Z.Fulare
Asst. Professor, Dept. of Electrical Engineering
Shri Sant Gajanan Maharaj College of Engineering
Shegaon, Dist-Buldhana, 444203

Ms.Preeti Ravindra Fulare
Dept. of Electrical Engineering
Shri Sant Gajanan Maharaj College of Engineering
Shegaon, Dist-Buldhana, 444203

Vijay S. Karale
Asst. Professor Dept. of Electrical Engineering
Shri Sant Gajanan Maharaj College of Engineering
Shegaon, Dist-Buldhana, 444203

**Abstract: In this paper, an algorithm has been developed around the theme of the conventional differential protection of the transformer. The proposed algorithm is based on probabilistic neural network (PNN) and use of the spectral energies of detail level wavelet coefficients of differential current signal for discriminating magnetising inrush and fault condition in the transformer. Performance of the proposed PNN is investigated with the conventional backpropagation feed forward (BPFF) multilayer perceptron neural network. To evaluate the developed algorithm, relaying signals for various operating condition (i.e., inrush and fault) of the transformer, are obtained from a custom-built single-phase transformer in the laboratory.**

**Keywords: differential protection; discrete wavelet transform; DWT; inrush current; internal fault; probabilistic neural network; PNN.**

## 1 Introduction

### What is a Power Transformer?

The Power transformer is one type of transformer that is used to transmit electrical energy in any component of the electronic or electrical circuit between the distribution primary circuits and the generator. These transformers are utilized in distribution networks to interface step down and step up voltages. The usual form of power transformer is fluid immersed, and the life cycle of these instruments is approximately 30 years. Power transformers can be divided into three types according to the ranges. They are large power transformers, medium power transformers, and small power transformers.

- The range of large power transformers can be from 100MVA and beyond
- The range of medium power transformers can be from -100MVA
- The range of low power transformers can be from 500-7500kVA

These transformers transmit the voltage. It keeps a low voltage, a high current circuit at one section of the transformer, and on the other side of the transformer, it keeps a high-voltage low current circuit. A Power transformer works based on the principle of Faraday's induction law. It explains the power network into areas where every gear attached to the system is designed per the rates set by the power transformer.

A power transformer is a static device employed for transforming power from one circuit to another without varying the frequency. This is a very simple definition of a transformer. Because there is no moving or rotating component, so a transformer is introduced as a static device. Power transformers perform based on an AC supply. A transformer operates on the rules of mutual induction. Power transformers are generally constructed to use the core part for maximum and will perform very much near to the apex of the B-H curve (Magnetic Hysteresis Loop). This takes down the mass of the core exceedingly. Typically, power transformers have the corresponding copper and iron loss. The electrical equipment and circuits in a substation must be protected in order to limit the damages due to abnormal currents and over voltages. All equipment installed in a power electrical system have standardized ratings for short-time withstand current and short duration power frequency voltage. The role of the protections is to ensure that these withstand limits can never be exceeded, therefore clearing the faults as fast as possible.

Power transformers are important elements of power system. So it is very important to avoid any maloperation of required protective system. For many years, differential protection has been used as the primary protection of power systems. It contains the differential relay, which operates for all internal fault types of power transformer and block due to inrush current. The major drawback of the differential protection relays stem from its potential for maloperation caused by the transient inrush current, which flow when the transformer is energised. The inrush current contains a large second harmonic component. Most of the methods for digital differential protection of transformers are based on detection of harmonic content of differential current. These methods are based on the fact that the ratio of the second harmonic to the fundamental component of differential current in inrush current condition is greater than the ratio in the fault condition. However, the second harmonic may also be generated during faults on the transformers. It might be due to saturation of CTs, or parallel capacitances. The second harmonic in these situations might be greater than the second harmonic in inrush currents. Thus, the commonly employed conventional differential protection based on second harmonic restraint will face difficulty in distinguishing inrush current and internal faults. Hence an improved technique of protection is required to discriminate between inrush current and internal faults (Moravej et al., 2000a, 2000b).

To overcome this difficulty and to prevent the malfunction of differential relay, many methods have been presented to analyse and recognise inrush current and internal fault currents. As both inrush current and internal faults are non-stationary signals, wavelet-based signal processing technique can be effectively used (Omar and Youssef, 2003). However, the wavelet-based methods have better ability of time-frequency analysis but they usually require long data windows and are also sensitive to noise. The method presented in Monsef and Lotfifard (2007) uses wavelet transform (WT) to discriminate internal faults from inrush current. Since the values of wavelet coefficients at detail 5 levels (d5) are used for pattern recognition process, the algorithm is very sensitive to noise.

In Faiz and Lotfifard's (2006) paper, a new algorithm is presented which discriminate between the inter-turn fault and magnetising inrush current. The algorithm used wavelet coefficients as a discriminating function. Two peak values corresponding to the mod of d5 (|d5|) level following the fault instant is used to discriminate the cases studied. As the criterion compare the two peak values, hence no threshold settings are necessary in this algorithm, but it is observed that in noisy environment it is difficult to identify correct switching instant and there the strategy fails.

Moreover, feed forward neural network (FFNN) (Mao and Aggarwal, 2001) has found wide application for detection of inrush current from internal faults but they have two major drawbacks: First, the learning process is usually time consuming. Second, there is no exact rule for setting the number of neurons to avoid over-fitting or underfitting. To avoid these problems, a radial basis function network (RBFN) has been developed in Moravej et al. (2001) paper. RBFs are well suited for these problems due to their simple topological structure and their ability to reveal how learning proceeds in an explicit manner. In Shin et al. (2003) paper, differential current harmonics are used as inputs to fuzzy logic.

The problem associated with these methods is the need to design neural networks and fuzzy laws, which require a huge number of training patterns produced by simulation of various cases. In Jazebi et al. (2009), an energy index is defined by calculation of nine-level frequency contours using S-transform to distinguish inrush current from internal fault currents. But the disadvantage of this method is determining the threshold value which can be different in transformers with different capacity and may change in noisy environment. Support vector machine (SVM) (Jazebi et al., 2009), hidden Markov model (HMM) (Jazebi et al., 2008) and Gaussian mixture models (GMM) (Jazebi et al., 2009) are used as new classifiers for detection of internal fault and inrush currents. In Jazebi et al. (2009), the extracted features are chosen from differential a current, which due to large data window is not effective as compared to those methods which use fewer features based on preprocessing step, like WT. But the performance and detection capability of SVM is better than HMM and GMM.

Although the proposed method in Tripathi et al. (2010) is without application of any deterministic index, the logarithm values of models probabilities for different inrush and internal fault currents are sensitive to noise condition and discrimination accuracy is reduced. Intrinsic sensitivity of wavelet analysis to noise, finding suitable mother wavelet from different types of mother wavelet and preprocessing data using k-mean clustering algorithm increase the complexity of the algorithm encountering huge computational task.

In Specht (1990), the optimal probabilistic neural network (PNN) is proposed as the core classifier to discriminate between the magnetising inrush and the internal fault of a power transformer. The particle swarm optimisation is used to obtain an optimal smoothing factor of PNN which is a crucial parameter for PNN.

This paper presents an algorithm which use energies of d3, d4 and d5 level wavelet coefficients of signal as an input to the PNN. The features extracted from DWT are given to PNN for training and subsequently it is tested for an effective classification. The performance of this algorithm is demonstrated on custom-built single-phase transformer, used in the laboratory to collect the data from controlled experiments.

## 2 Wavelet transform

Wavelet analysis is about analysing the signal with short duration finite energy functions which transform the considered signal into another useful form. This transformation is called WT. Let us consider a signal $f(t)$, which can be expressed as:

$$f(t) = \sum_{l} a_l \ \phi_l(t) \tag{1}$$

where, $l$ is an integer index for the finite or infinite sum. Symbol $a_l$ are the real valued expansion coefficients, while $\phi_l(t)$ are the expansion set. If the expansion (1) is unique, the set is called a basis for the class of functions that can be so expressed. The bases are orthogonal if

$$\left[\phi_l(t), \phi_k(t)\right] = \int \phi_l(t) \phi_k(t)\, dt = 0 K \neq l \tag{2}$$

Then coefficients can be calculated by the inner product as

$$\langle f(t), \phi_k(t) \rangle = \int f(t)\phi_k(t)dt \quad (3)$$

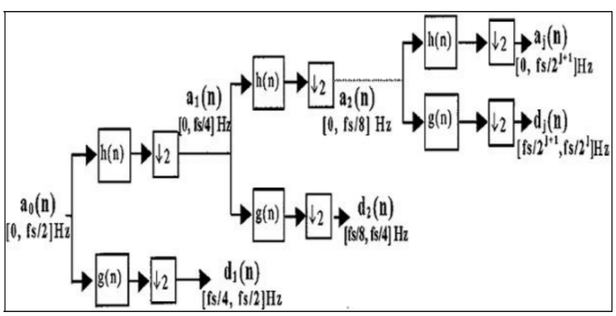
If the basis set is not orthogonal, then a dual basis set $\phi_k(t)$ exists such that using (3) with the dual basis gives the desired coefficients. For wavelet expansion, equation (1) becomes

$$f(t) = \sum_k \sum_j a_{j,k} \phi_{j,k}(t) \quad (4)$$

In (4), $j$ and $k$ are both integer indices and $\phi_{j,k}(t)$ are the wavelet expansion function that usually form an orthogonal basis. The set of expansion coefficients $a_{j,k}$ are called discrete wavelet transform (DWT).

There are varieties of wavelet expansion functions (or also called as a mother wavelet) available for useful analysis of signals. Choice of particular wavelet depends upon the type of applications. If the wavelet matches the shape of signal well at specific scale and location, then large transform value is obtained, vice versa happens if they do not correlate. This ability to modify the frequency resolution can make it possible to detect signal features which may be useful in characterising the source of transient or state of post disturbance system. In particular, capability of wavelets to spotlight on short time intervals for high frequency components improves the analysis of signals with localised impulses and oscillations particularly in the presence of fundamental and low order harmonics of transient signals. Hence, wavelet is a powerful time frequency method to analyse a signal within different frequency ranges by means of dilating and translating of a single function called mother wavelet.

**Figure 1** Multiresolution signal decomposition




The DWT is implemented using a multiresolution signal decomposition algorithm to decompose a given signal into scales with different time and frequency resolution. In this sense, a recorder-digitised function $a_0(n)$, which is a sampled signal of $f(t)$, is decomposed into its smoothed version $a_1(n)$ (containing low-frequency components), and detailed version $d_1(n)$ (containing higher-frequency components), using filters $h(n)$ and $g(n)$, respectively. This is a first-scale decomposition. The next higher scale decomposition is now based on signal $a_1(n)$ and so on, as demonstrated in Figure 1.

The analysis filter bank divides the spectrum into octave bands. The cutoff frequency for a given level $j$ is found by

$$fc = \frac{fs}{2^{j+1}} \quad (5)$$

where $fs$ is the sampling frequency. The sampling frequency in this paper is taken to be 10 kHz and Table 1 shows the frequency levels of the wavelet function coefficients.

**Table 1** Frequency levels of wavelet functions coefficients

| *Decomposition level* | *Frequency components (HZ)* |
|---|---|
| d1 | 5,000–2,500 |
| d2 | 2,500-1,250 |
| d3 | 1,250–625 |
| d4 | 625–312.5 |
| d5 | 312.5–156.25 |
| a5 | 0–156.25 |

## 3 Neural networks

### *3.1 Artificial neural network*

ANNs are highly interconnected processing units inspired in the human brain and its actual learning process. Interconnections between units have weights that multiply the values which go through them. Also, units normally have a fixed input called bias. Each of these units forms a weighted sum of its inputs, to which the bias is added. This sum is then passed through a transfer function.

Prediction with NNs involves two steps: training and learning. Training of FFNNs is normally performed in a supervised manner. The success of training is greatly affected by proper selection of inputs. In the learning process, a neural network constructs an input-output mapping, adjusting the weights and biases at each iteration based on the minimisation or optimisation of some error measure between the output produced and the desired output. This process is repeated until an acceptable criterion for convergence is reached. The most common learning algorithm is the backpropagation (BP) algorithm, in which the input is passed layer through layer until the final output is calculated, and it is compared to the real output to find the error. The error is then propagated back to the input adjusting the weights and biases in each layer. The standard BP learning algorithm is a steepest descent algorithm that minimises the sum of square errors. In order to accelerate the learning process, two parameters of the BP algorithm can be adjusted: the learning rate and the momentum. The learning rate is the proportion of error gradient by which the weights should be adjusted. Larger values can give a faster convergence to the minimum. The momentum determines the proportion of the change of past weights that should be used in the calculation of the new weights.

In this paper, the fully-connected multilayer FFNNs is used and trained for discrimination of inrush and fault with a supervised BP learning algorithm. The FFNN consists of *an input layer representing the input data to the network, hidden layers and an output layer representing the response of the network. Each layer consists of a certain number of neurons; each neuron is connected to other neurons of the previous layer through adaptable synaptic weights w and biases b. If the inputs of neuron j are the variables $x_1, x_2, \ldots, x_i, \ldots, x_N$*, the output $u_j$ of neuron *j* is obtained as

$$u_j = \phi \sum_{i=1}^{N} w x + b$$

*where $w_{ij}$ is the weight of the connection between neuron j* and $i^{th}$ input; $b_j$ is the bias of neuron *j* and $\phi$ is the transfer (activation) function of neuron *j*.

An FFNN of three layers (one hidden layer) is considered with *N*, *M* and *Q* neurons for the input, hidden and output layers, respectively. The input patterns of the ANN represented by a vector of variables $x = x_1, x_2, \ldots, \ldots, x_N)$ submitted to the NN by the input layer are transferred to the hidden layer. Using the weight of the connection between the input and the hidden layer and the bias of the hidden layer, the output vector $u = (u_1, u_2, \ldots, u_j, \ldots, u_M)$ of the hidden layer is determined.

*The output $u_j$ of neuron j is obtained*

Training is the process of adjusting connection weights *w* and biases *b*. In

$$u_j = \phi\, hid \sum_{i=1}^{N} w_{ij}^{hid} x_i + b_j^{hid} \tag{7}$$

where $w_{ij}^{hid}$ is the weight of connection between neuron *j* in the hidden layer and the $i^{th}$ neuron of the input layer, $b_j^{hid}$ represents the bias of neuron *j* and $\phi$ *hid* is the activation function of the hidden layer.

The values of the vector *u* of the hidden layer are transferred to the output layer. Using the weight of the connection between the hidden and output layers and the bias of the output layer, the output vector $y = (y_1, y_2, \ldots, y_k, \ldots, y_O)$ of the output layer is determined. The output $y_k$ of neuron *k* (of the output layer) is obtained as

$$y_k = \phi out \sum_{j=1}^{M} w_{jk}^{out} u_j + b_k^{out} \tag{8}$$

where $w_{jk}^{out}$ is the weight of the connection between neuron *k* in the output layer and the $j^{th}$ neuron of the hidden layer, $b_k^{out}$ is the bias of neuron *k* and $\phi$*out* is the activation function of the output layer.

The output $y_k$ is compared with the desired output (target value) $y_k^d$. The error *E* in the output layer between $y_k$ and $y_k^d$ $\left(y_k^d - y_k\right)$ is minimised using the mean square error at the output layer (which is composed of *Q* output neurons), defined by

the first step, the network outputs and the difference between the actual (obtained) output and the desired (target) output (i.e., the error) is calculated for the initialised weights and biases (arbitrary values). In the second stage, the initialised weights in all links and biases in all neurons are adjusted to minimise the error by propagating the error backwards (the BP algorithm). The

network outputs and the error are calculated again with the adapted weights and biases, and this training process is repeated at each epoch until a satisfied output $y_k$ is obtained corresponding with minimum error. This is by adjusting the weights and biases of the BP algorithm to minimise the total mean square error and is computed as

$$w = w^{new} - w^{old} = -\eta \frac{\partial E}{\partial w} \tag{10a}$$

$$b = b^{new} - b^{old} = -\eta \frac{\partial E}{\partial b} \tag{10b}$$

biases. For the output layer, we have,

$$w_{jk}^{new} = \alpha\, w_{jk}^{old} + \eta\, \delta_k y_k \tag{11a}$$

$$b_k^{new} = \alpha\, b_k^{old} + \eta \delta_k \tag{11b}$$

where α is the momentum factor (a constant between 0 and 1) and $\delta_k = y_k^d - y_k$, for the hidden layer, we get,

$$w_{ij}^{new} = \alpha\, w_{ij}^{old} + \eta \delta_j y_j \tag{12a}$$

$$b_j^{new} = \alpha\, b_j^{old} + \eta \delta_j \tag{12b}$$

where $\delta_j = \sum^{Q}_{k} \delta_k w_{jk}$ and $\delta_k = y_k^d - y_k$.

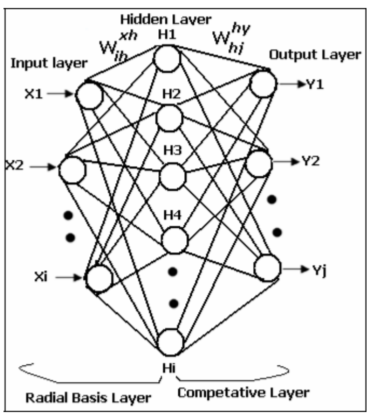
### 3.2 Probabilistic neural network

Probabilistic neural networks can be used for classification problems. When an input is presented, the first layer computes distances from the input vector to the training input vectors and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a1 for that class and a0 for the other classes. The PNN model is one among the supervised learning networks and has the following features distinct from those of other networks in

- it is implemented using the probabilistic model, such as Bayesian classifiers
- a PNN is guaranteed to converge to a Bayesian classifier provided that it is given enough training data
- no learning processes are required
- no need to set the initial weights of the network
- no relationship between learning processes and recalling processes.
- the difference between the inference vector and the target vector are not used to modify the weights of the network.

The learning speed of the PNN model is very fast making it suitable in real time for fault diagnosis and signal classification problems. Figure 2 shows the architecture of a PNN model that is composed of the radial basis layer and the competitive layer.

**Figure 2** Architecture of PNN




In the signal-classification application, the training examples are classified according to their distribution values of probabilistic density function (pdf), which is the basic principle of the PNN. A simple probabilistic density function is as follows:

$$fx(X) = \frac{1}{Nk} \sum_{j=1}^{Nh} \exp\left( - \frac{X - X_{kj}}{2\sigma^2} \right)$$

Modifying and applying (6) to the output vector H of the hidden layer in the PNN is as follows:

$$H_h = \exp\left( - \frac{\sum \left| \left( Xi - W_{ij}^{xh} \right)^2 \right|}{2\sigma^2} \right)$$

$$net_j = \frac{1}{N_j} \sum_h W_{hj}^{hy} H_h \text{ and } N_j = \sum W_{hj}^{hy} \tag{13}$$

where

*i* – number of input of layers, *h* – number of hidden layers

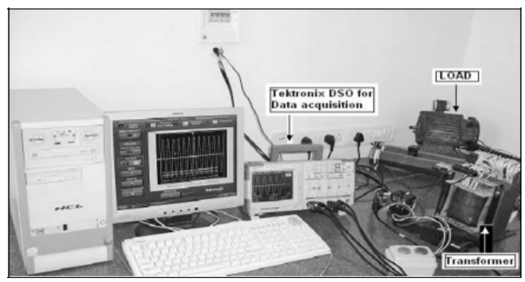
*j* – number of output layers; *k* – number of training examples

*N* – number of classifications (clusters); $\sigma$ – smoothing parameters (standard deviation)

*X* – input vector; and where $\left| X - X_{kj} \right|$ is the Euclidean distance between the vectors *X* and $X_{kj}$, i.e., $\left| X - X_{kj} \right| = \sum_j \left( X - X_{kj} \right)^2$, $W_{ih}^{xh}$ is the connection weight between the input layer *X* and the hidden layer *H* and $W_{ih}^{hy}$ is the connection weight between the hidden layer *H* and the output layer *Y*.

**Figure 3** Experimental setup

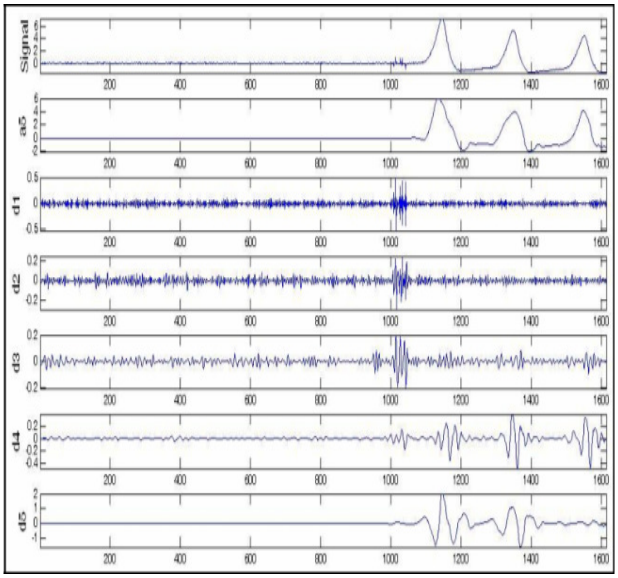
$net_j = \max_k (net)$ then $Y_j = 1$, else $Y_j = 0$



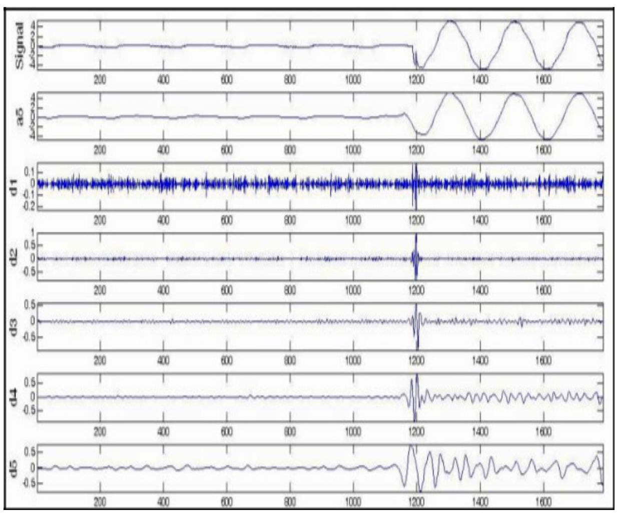

## 4 Experimentation and data collection

During transformer operation it encounters any one of the following conditions: Normal condition, magnetising inrush/sympathetic inrush condition, and internal fault condition. To evaluate the developed algorithm, relaying signals for various operating condition of the transformer, are obtained from a custom-built 220 V/220 V, 2 KVA, and 50 Hz single-phase transformer with externally accessible taps on both primary and secondary to introduce interturn faults. The primary winding and secondary winding has 272 turns each. The load on the secondary comprises of static and rotating elements. Data acquisition card by Tektronix Instruments is used to capture the voltages and current signals. These signals are recorded at a sample rate of 10,000 samples/sec. Different cases of inter turn short circuit are staged, considering the effect of number of turns shorted on primary and secondary and load condition. Experimental setup is as shown in Figure 3. Primary (*Ip*) and secondary (*Is*) currents are captured using the experimental setup. The Tektronix DSO, TPS 2014 B, with 100 MHz bandwidth and adjustable sampling rate of 1 GHz is used to capture the currents. The Tektronix current probes of rating 100 mV/A, input range of 0 to 70 Amps AC RMS, 100 A peak and frequency range DC to 100 KHz are used.

**Figure 4** Wavelet decomposition of differential current for (a) inrush and (b) fault (see online version for colours)




(a)




(b)

## 5 Feature selection by WT

The captured current signals for inrush and faulted condition staged on mains feed custom built transformer are decomposed up to fifth level using Daubechies-4 (db4) as a mother wavelet. Several wavelets have been tried but db4 is found most suitable due to its compactness and localisation properties in time frequency frame. The threshold value for initiating the calculations of DWT depends upon the capacity of transformer. In this paper, the threshold value is set to 5% of the rated primary current of the transformer. Figure 4(a) and Figure 4(b) shows decomposition of differential current signal for inrush and fault respectively.

From Figures 4(a) and 4(b) there is no discrimination between inrush and fault current wave form by visual inspection and from changes in the detailed and approximation coefficients for both cases it is very difficult to draw any inference. Therefore some suitable artificial intelligence technique should be used to classify these events in transformer. Hence energies of differential current of level d3, d4, and d5 for one cycle are calculated and are used as an input to the neural networks.

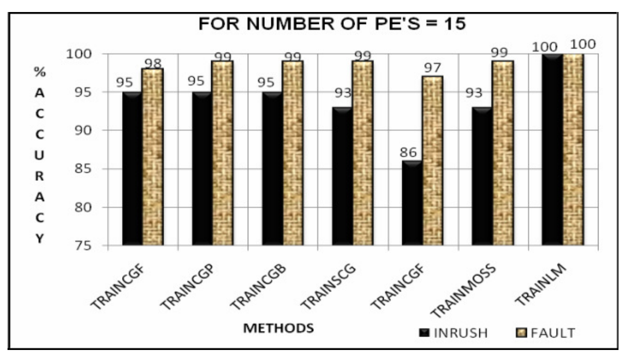
## 6 ANN training and testing

### *6.1 BP feed forward ANN*

In this paper, BP feed multilayer perceptron (MLP) neural network is used as a classifier network. Energies of decomposed level d3, d4, and d5 for one cycle are used as an input to the network. This input dataset obtained through experiments on custom built transformer is found to be non-linear non-separable mixed data. ANN possesses ability to classify such mixed datasets and can be used effectively in obtaining the correct classifications of the events in transformer. For generalisation the randomised data is fed to the network and is trained for different hidden layers. The numbers of processing elements (PEs) in the hidden layer are varied to obtain minimum mean square error.

Various training methods are used to train the network, for all training methods, the network is build with configuration 0.8 learning rate, 0.6 momentum, 60% samples used for training purpose, and 40% for testing purpose. With this network the variation of average mse and correct percent accuracy for both inrush and fault with respect to number of PEs in the hidden layer is obtained. It is found that for training method 'trainlm' of Levenberg-Marquardt for one hidden layer and fifteen PEs in the hidden layer the mean square error is minimum and correct classification accuracy is hundred percent as shown in Figure 5.

**Figure 5** Variation of correct percent accuracy with 15 number of PEs in the hidden layer (see online version for colours)

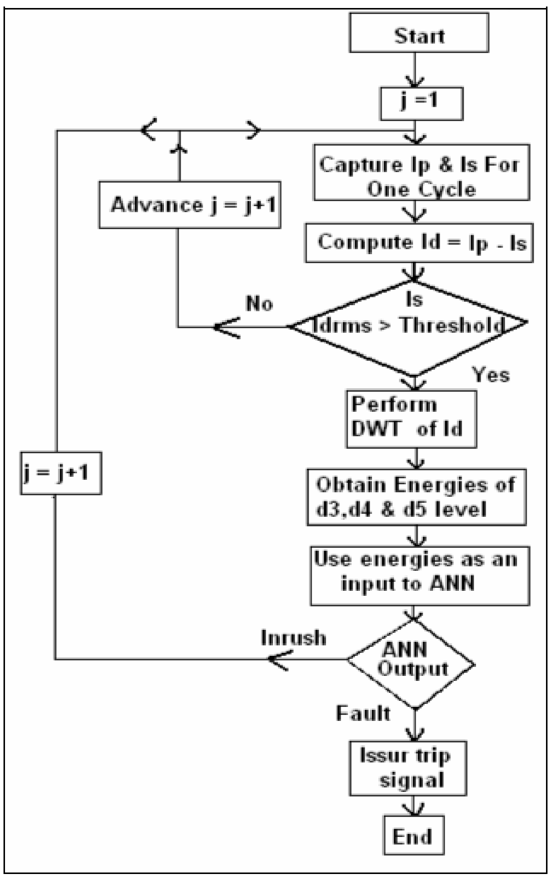



The training of BP MLP involves heuristic searches, which involves small modifications of the network parameter that result in a gradual improvement of system performance. Heuristic approaches are associated with long training times with no guarantee of converging to an acceptable solution within a reasonable timeframe, hence not suggested for on line condition monitoring of equipment.

### *6.2 Probabilistic neural network as a decision classifier*

In this paper, two layer fully-connected probabilistic neural network, consists of one input layer, and one output layer is used. The input layer consists of three neurons: the inputs to these neurons are the spectral energy contained in the detail d3, d4 and d5 level. The output layer consists of two neuron representing the magnetising inrush and fault. With respect to the hidden layer it is customary that the number of neurons in the hidden layer can be obtained by trial and error. Sixty experiments for each case (inrush, interturn fault in primary winding, interturn fault in secondary winding) have been performed at zero switching instant (switching at zero produces Sevier transients). Total of 180 readings data is available as an input to PNN, out of this 60% data is used for training the network and 40% data is used for testing purpose. With a spread of 0.02, randomisation of data and forward tagging network is trained and tested to classify inrush from fault conditions.

**Figure 6** Flow chart for on-line classification scheme




This method gives 100% accuracy to discriminate inrush from the fault in a transformer. PNN has shown the capability to discriminate the inrush current from internal fault current. The online discrimination process of inrush and fault current is illustrated in Figure 6. The step for carrying out the on-line detection scheme is presented as under:

1. Capture one cycle of primary (*Ip*) and secondary (*Is*) current by data acquisition system
2. Compute differential current, $Id = Ip - Is$
3. If the rms value of differential current value is less than threshold value, go to step 1
4. Obtain detail coefficient by performing DWT of differential current up to fifth level if it exceeds the threshold
5. Obtain the energies of decomposed levels from d3 to d5
6. Give energy of decomposed levels d3 to d5 to PNN as input data to discriminate the fault and inrush that is healthy condition
7. If PNN output is discriminated as fault, then issue trip signal otherwise proceed further (i.e., monitor the differential current).

## 7 Discussion

Two ANNs have been reviewed: the PNN and the backpropagation feed forward neural network (BPFFNN). The most important advantage of PNN networks is that training is easy and instantaneous. They can be used in real time cause as soon as one pattern representing each category has been observed; the network can begin to generalise to new patterns.

PNN has other advantages

1. the shape of the decision surfaces can be made as complex as necessary, or as simple as desired, by choosing the appropriate value of the smoothing parameter $\sigma$
2. the decision surfaces can approach optimal minimum-risk decision surfaces (Bayes criterion)
3. erroneous samples are tolerated
4. sparse samples are adequate for network performance
5. $\sigma$ can be made smaller as $n$ gets larger without retraining
6. for time-varying statistics, old patterns can be overwritten with new patterns.

PNN does have the disadvantage of requiring the storage and use of the entire training data base for testing of unknown patterns.

Real-time implementation of differential relaying using the proposed algorithm applying PNN as the core classifier would essentially require PNN processor, as software implementation of PNN would be slow for protective relaying. Long term memory weights can then be used at the processor level to take decisions regarding the classification of inrush and fault current.

## 8 Conclusions

This paper presents a novel approach to discriminate between transformer internal fault and magnetising inrush condition based on PNN in digital differential protection scheme. The proposed PNN algorithm is more accurate than traditional harmonic restraint-based technique, especially in the case of modern power transformers which use high-permeability, low-corrosion core materials. The conventional harmonic restrain technique may fail because high second harmonic components are generated during internal faults and low second harmonic components are generated during magnetising inrush with such core materials. PNN overcomes the shortcomings of entrapment in local optimum; slow convergence rate corresponds to BP algorithm. Because of the fast training rate, the training samples can be added into PNN at any time. So, PNN is fit to diagnose the fault of power transformer and has auto-adaptability. The algorithm has an acceptable accuracy in the recognition of unused patterns for learning. This fact highlights the on line practical importance of algorithm.

## References


Faiz, J. and Lotfifard, S. (2006) 'A novel wavelet based algorithm for discrimination of internal faults from magnetizing inrush current in power transforms', *IEEE Transaction on Power Delivery*, October, Vol. 21, No. 4.

Jazebi, S., Vahidi, B. and Hosseinian, S.H. (2008) 'A new stochastic method based on hidden Markov models to transformer differential protection', *IEEE Proceeding OPTIM'08*, Vol. 1, No. 1, pp.179–184.

Jazebi, S., Vahidi, B., Hosseinian, S.H. and Faiz, J. (2009) 'Magnetizing inrush current identification using wavelet based Gaussian mixture models', *Simulation Modeling Practice and Theory*, Vol. 17, No. 6, pp.991–1010.

Mao, P.L. and Aggarwal, R.K. (2001) 'A novel approach to the classification of the transient phenomena in power transformers using combined wavelet transform and neural network', *IEEE Transaction on Power Delivery*, Vol. 16, No. 4, pp.654–660.

Monsef, H. and Lotfifard, S. (2007) 'Internal fault current identification based on wavelet transform in power transformers', *Electric Power System Research*, Vol. 77, No. 1, pp.1637–1645.

Moravej, Z., Vishvakarma, D.N. and Singh, S.P. (2000a) 'Digital filtering algorithms for the differential relaying of power transformer: an over view', *Electric Machines and Power Systems*, 1 June, Vol. 28, No. 6, pp.485–500.

Moravej, Z., Vishwakarma, D.N. and Singh, S.P. (2000b) 'ANN based protection scheme for power transformer', *Electric Machines and Power Systems*, Vol. 28, No. 9, pp.875–884.

Moravej, Z., Vishwakarma, D.N. and Singh, S.P. (2001) 'Radial basis function (RBF) neural network for protection of a power transformer', *Electric Power Components and Systems*, Vol. 29, No. 4, pp.307–320.

Omar, A. and Youssef, S. (2003) 'A wavelet-based technique for discrimination between faults and magnetizing inrush currents in transformers', *IEEE Transaction on Power Delivery*, January, Vol. 18, No. 1, pp.170–176.

Shin, M.C., Park, C.W. and Kim, J.H. (2003) 'Fuzzy logic-based for large power transformer protection', *IEEE Transaction on Power Delivery*, July, Vol. 18, No. 3, pp.718–724.

Specht, D.F. (1990) 'Probabilistic neural network', *IEEE Transaction on Neural networks*, Vol. 3, No. 1, pp.109–118.

Tripathi, M., Maheshwari, R.P. and Verma, H.K. (2010) 'Power transformer differential protection based on optimal probablistic neural network', *IEEE Transaction on Power Delivery*, January, Vol. 25, No. 1, pp.102–112.