A Learning System for Detecting Transformer Internal Faults

Abdullah Md. Saleh, Md. Zakir Hossain, Md. Jubayer Alam Rabin, A. N. M Enamul Kabir, Md. Fazle Elahi Khan, Md. Shahjahan
Khulna University of Engineering and Technology, Khulna-9203, Bangladesh

Abstract—**Miniature transformer is one of the most important components of electronic devices. A serious failure of such kind of transformer may cause loss of time and money. This paper presents a learning system to recognize internal fault of such kind of transformer. The different kinds of faults are made to occur intentionally and data are collected at various conditions. The faults include turn to turn, winding to ground, and dielectric faults. The data are then processed and entered in the learning algorithms to recognize the type of fault. We devise a learning system to recognize the various types of faults. Several versions of learning algorithms such as standard back propagation, Levenberg-Marquardt, Bayesian regulation, Resilient back propagation, Gradient descent, One-step secant, Elman recurrent network are used. The result of Levenberg-Marquardt algorithm was found to be faster than that of other algorithms. Therefore it is suitable for real time fault detection.**

Keywords-Miniature transformer, internal fault, neural network (NN), Back propagation algorithm, fault detection

I. INTRODUCTION

The transformer is defined as a static piece of device with two or more windings which by electromagnetic induction, transforms a system of alternating voltage and current into another level of voltage and current usually of different values and at the same frequency for the purpose of transmitting electrical power.

Miniature transformer is used almost every electronic devices on everyday life, like as IPS, UPS, adapter, television, radio etc. Damages to the miniature transformer can be caused by different stresses, which are due to overheating, open circuits and short circuits. The major concern of this type of transformer is short circuit fault because open circuits do not cause too much hazard. Usually there is less protection scheme for this kind of transformer. Therefore, it is necessary to take care on the transformer during operation. This paper deals with the different faults recognition using neural network approach.

Different techniques have been used in the area of transformer fault detection and diagnosis. However, the implementations of the existing monitoring methods [1]–[4] tend to cost too much to be applied to distribution transformers. The appropriate protection scheme must be selected to ensure the safety of power apparatus and reliability of the system. Generally, transformers can be protected by over current relays, pressure relays and differential relays depending on purposes [5]. For differential protection, the differential current, which is generated by a comparison between the primary current and the secondary current detected via current transformers, is required. The differential protection is aimed at detecting internal faults in transformer windings. In a normal operation or in a fault condition due to the external short circuits, the differential current is relatively small, and the differential relay should not function [5,6]. However, there are some factors that can cause a needless operation of the differential protection such as effects from magnetizing inrush current. To avoid the malfunction of the differential relay, the discrimination between internal faults, magnetizing inrush current and external short circuit current is required [5-7]. Several transformer models and decision algorithms have been proposed and discussed for such a task [8-10]. Recently, with the advance of signal processing technologies and artificial intelligent tools, the development of more sophisticated protection systems as well as fault diagnosis for the transformer has been progressed with the applications of artificial neural networks (ANNs) [11-14].

Back-propagation is a common method of training artificial neural networks so as to minimize the objective function. Arthur E. Bryson and Yu-Chi Ho described it as a multi-stage dynamic system optimization method in 1969.[15] It wasn't until 1974 and later, when applied in the context of neural networks and through the work of Paul Werbos. [16] David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams, [17-18] that it gained recognition, and it led to a “renaissance” in the field of artificial neural network research.

This paper is organized as follows. Section II describes about data acquisition and experimental setup. The learning system of our work is explored in section III. Overview of different learning algorithms are described in section IV. Experimental result and discussion of work are analyzed in section V. Finally section VI concludes the paper.

978-1-4799-0400-6/13/$31.00 ©2013 IEEE

II. DATA ACQUISITION AND EXPERIMENTALE SETUP

A 50 Hz, 3000 mA, 220/24 V two winding single phase step down transformer was employed for determining various parameters under different faulty conditions. Three types of transformer fault - internal turn-to-turn fault, transformer winding to ground fault and transformer dielectric fault were employed in this experiment. For fault detection the output voltage and current were measured under different conditions. The measured data are used for training and testing purposes.

For inter-turn fault, the high voltage side of 220/24V single phase step down transfer was connected to the supply line and the output was taken from low voltage side. There are several tapings are taken from both low and high voltage sides. Faults are intentionally occurred making short between any pair of tapings. The transformer winding turns were taken to such a position that they were very close to be short circuited but not totally connected. Therefore a lot of data was recorded at this pre-fault condition. For clear understanding an example is given here.

There are a total of four tapings and two power lines (positive and negative terminals) in low voltage side. There are also three tapings and two power lines (positive and negative terminals) in the high voltage side. 100% length is considered between two power lines. For the inter-turn faults in low voltage side, the position of point 'a' on the transformer winding, as shown in Fig. 1, was varied at the length of 25%, 37.5%, 50% and 75% measured from the positive terminal and the tapings. Similarly, the position of point 'b' on the transformer winding, as shown in Fig.1, was varied at the length of 12.5%, 25% and 50% measured from the point 'a' of the tapings.

For the inter-turn faults in high voltage side, the position of point 'a' on the transformer winding, as shown in Fig.1, was varied at the length of 63.64%, 72.73%, 81.82% and 90.91% measured from the positive terminal and the windings. Also the position of point 'b' on the transformer winding, as shown in Fig.1, was varied at the length of 9.09%, 18.18% and 27.27% measured from the point 'a' and the tapings.

For the winding to ground faults in low voltage side, the fault locations were designated on transformer windings at the length of 25%, 37.5%, 50% and 75% measured from the supply line end and the tapings. For the winding to ground faults in high voltage side, the fault locations were designated on transformer windings at the length of 63.64%, 72.73%, 81.82% and 90.91% measured from the supply line end and the tapings. The transformer winding turns were taken to such a position that they were very close to be grounded but not totally grounded.

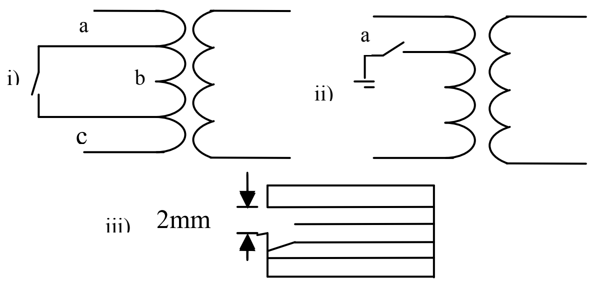

Fig. 1: Faultclasses i) internal turn-to-turn fault, ii) winding to ground fault, iii) Dielectric fault

The dielectric medium of the 220/24 two winding single phase step down transformer was air. So in order to put a change in the air gap, single and multiple fractures were put in the core and the corresponding variations in output voltage and current were determined which were very precious. For every case, the load resistance was 10 KΩ. It is to recognize these faults with Levenberg-Marquardt algorithm.

III. THE LEARING SYSTEM

Physically different faultsare created in a transformer and a machine learning algorithm is used to detect these faults. Fig. 2 depicts the entire flow of the fault detection scheme and alarming to transformer. Firstly different kinds of fault are created and collected. Secondly different learning algorithms are used to recognize a particular fault. Finally an alarming signal is generated for the transformer premises. Our objective is to test different learning algorithms for different kinds of faults.

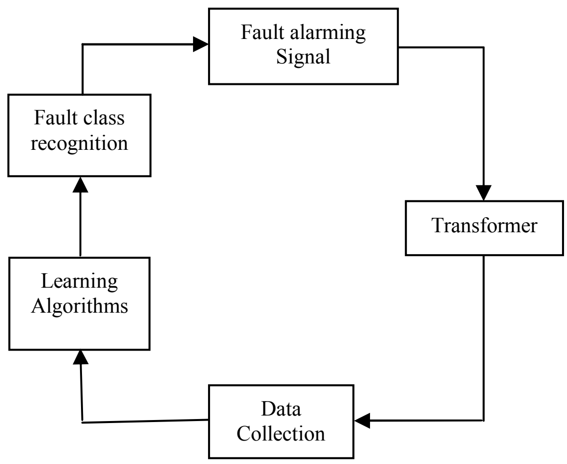

Fig. 2: Fault detection scheme

The procedure used to perform the learning process is called a learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective.

The neural network in Fig.3 is said to be fully connected in the sense that every node in each layer of the network is connected to every other node in the adjacent forward network.

The activation functions, denoted by φ(x),defines the output of a neuron in terms of the net input x. Sometimes it referred as transfer function since it limits and rectifies the output signal according to its characteristics. The sigmoid function is the most common form of activation function used in construction of artificial neural network, defined by

$$\varphi(x) = \frac{1}{1+\exp(-x)} \quad \ldots\ldots\ldots (1)$$

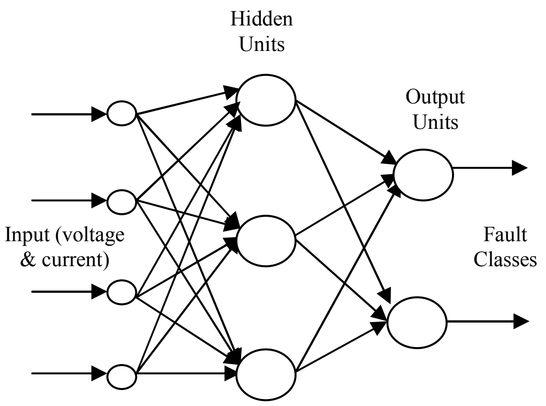

Fig. 3: A fully connected feed forward neural network model. Elman recurrent networkcan be formed bringing output from hidden units to input units.

IV. OVERVIEW OF LEARNING ALGORITHMS

a) Levenberg-marquardt algorithm

This method provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function. These minimization problems arise especially in least squarescurve fitting and nonlinear programming. The LMA interpolates between the Gauss–Newton algorithm (GNA) and the method of gradient descent.LMA is designed to approach second-order training speed without having to compute the Hessian matrix. The LMA uses approximation to the Hessian matrix in the following Newton-like update:

$$X_{k+1} = X_k - [J^T J + \mu I]^{-1} J^T e \quad \ldots\ldots\ldots\ldots\ldots (2)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard back-propagation technique that is much less complex than computing the Hessian matrix. When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible.

The LMA is a very popular curve-fitting algorithm used in many software applications for solving generic curve-fitting problems. It is often the fastest back propagation algorithm and highly recommended as a first-choice supervised algorithm, although it does require more memory than other algorithms. However, the LMA finds only a local minimum, not a global minimum.

b) Standard Back Propagation (SBP) algorithm

Back-propagation is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks.

The input from unit i into unit j is denoted x_{ji}, and the weight from unit i to unit j is denoted w_{ji}. For each (x, t), in training examples, propagate the input forward through the network [19].

SBP(training_examples, ή)

Each traing example is a pair of the form (x, t), where x is the vector of input values, and t is the target output value, ή is the learning rate.

- Initialize each w_i to some small random value
- Until the termination condition is met, Do
- Initialize each Δw_i to zero
- For each (x, t), in training_examples, Do
- Input the instance x to the unit and compute the output o
- For each linear unit weight w_i, Do $\Delta w_i = \Delta w_i + ή(t-o)x_i$
- For each linear unit weight w_i, Do $w_i = w_i + \Delta w_i$

Fig. 4: Weight updates process of a generalized SBP

c) Bayesian regulation back (BRP) propagation

Bayesian regulation back (BRP) propagation is a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well.

d) Resilient back propagation (RBP)

The purpose of the resilient backpropagation (RBP) training algorithm is to eliminate the harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative can determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update. The size of the weight change is determined by a separate update value.Multilayer networks typically use sigmoid transfer functions in the hidden layers. These functions are often called "squashing" functions, because they compress an infinite input range into a finite output range. Sigmoid functions are characterized by the fact that their slopes must approach zero as the input gets large. This causes a problem when you use steepest descent to train a multilayer network with sigmoid functions, because the gradient can have a very small magnitude and, therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values.

e) The one step secant (OSS)

The one step secant (OSS) method is an attempt to bridge the gap between the conjugate gradient algorithms and the quasi-Newton (secant) algorithms. This algorithm does not store the complete Hessian matrix. It assumes that at each iteration the previous Hessian was the identity matrix. This has the additional advantage that the new search direction can be calculated without computing a matrix inverse.

f) Elman recurrent network (ERN)

A recurrent neural network (RNN) is a class of neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Unlike feed forward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs. This makes them applicable to tasks such as un-segmented connected handwriting recognition, where they have achieved the best known results. Elman network is a Recurrent Neural Network (ERN) which is constructed by taking feed forward network architecture and adding feedback connections to previous layers. Such networks are trained by the standard back propagation algorithm except that patterns must always be presented in time sequential order.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Here, three-layer back propagation neural network with one input layer, one hidden layer and one output layer is employed. Hyperbolic tangent sigmoid function is used as an activation function in the hidden layer while linear function is used as an activation function in output layers. A training process was performed using neural network toolboxes in MATLAB. A structure of the back propagation neural network consists of 2 neuron inputs and 6 neuron outputs. The inputs are the normalized value of output voltages and currents of the transformer. The output targets of the neural networks are designated as either 0 or 1 with various types of faults as shown in Table 1.

Table 1: Output target patterns for neural networks for various fault types.

Fault classes	Neural Network outputs					
Inter-turn short circuit(LVIT)	1	0	0	0	0	0
Winding to ground(LVWG)	0	1	0	0	0	0
Inter-turn short circuit(HVIT)	0	0	1	0	0	0
Winding to ground(HVWG)	0	0	0	1	0	0
Dielectric Fault (D)	0	0	0	0	1	0
Healthy condition(H)	0	0	0	0	0	1

If the inter-turn short circuit fault is made on the low-voltage side, it is named as LVIT. Similarly, ifit is made on the high voltage side of the transformer,it is designated as HVIT.Other faults are self explanatory as also explained previously in Sec II.

The initial number of neurons for the hidden layer can be calculated as shown in (3).

$$z = \frac{2}{3}(r+q) \qquad \ldots\ldots\ldots(3)$$

Where, z is the initial number of neurons in the first hidden layer, r is number of neurons input and q is number of neurons output. When the initial number of neurons in the hidden layer had been determined, the final number of the neurons in the same layer had to be calculated in order to stop the training process. The final number can be obtained from:

$$z_{st} = z + z1 \qquad \ldots\ldots\ldots(4)$$

Where, z_{st} is the final number for the neurons in the hidden layer,

$$z1 = \begin{cases} 5 \text{ if } 2 \le z \le 6 \\ 4 \text{ if } 7 \le z \le 10 \\ 3 \text{ if } 11 \le z \le 14 \\ 2 \text{ if } 14 \le z \end{cases}$$

During the training process, the weight and biases were adjusted, and there were 100 iterations. The training procedure was stopped when reaching the final number of iterations or the training error was less than 0.01. The various results from the training process can be shown in Table 2 with the initial number of neurons for the hidden layer obtained from equation (3).

Table 2: Error and execution time(Performed on an Intel(R) Core(TM) i3 CPU M 330 @ 2.13GHz 917 MHz, 1.85 GB of RAM).

Number of neurons in the hidden layer	Error on test set	Execution time (training, seconds)
5	0.0347	52.23
6	0.0348	57.01
7	0.0330	62.91
8	0.0350	68.71
9	0.0338	74.65
10	0.0348	89.33

We added noise to test values and checked fault detection capability of the neural network algorithms. In this case white Gaussian noise was added to the input data. The noise level was varied by varying the signal to noise ratio (SNR). Fig.5 shows the accuracy of fault detection of the Standard Back Propagation (SBP) algorithm. The accuracy decreases with increase of noise in terms of SNR. When the SNR is high the accuracy is 100%. Then with decreasing SNR the accuracy falls. The data sets we got experimentally were employed to different back propagation models provided in MATLAB toolbox.

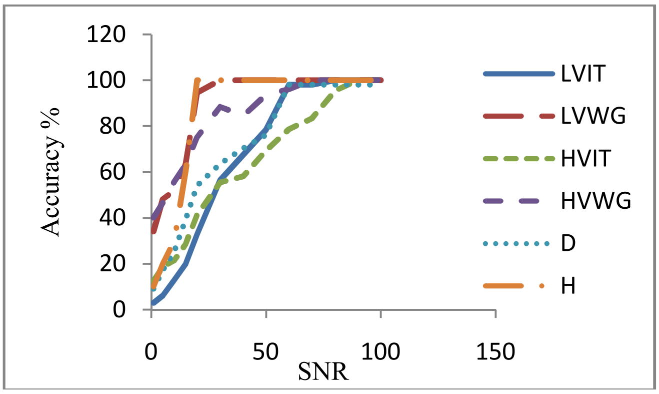

Fig. 5: variation of accuracy with SNR for different conditions

Table 3 describes the recognition ability of different algorithms under various fault conditions.It is seen that inter-turn faults are critical for the network to recognize, because data are considerably different for same class. On the other hand WG fault at LV side is very easy to recognize for all algorithms (100%), while it is difficult for faults in the HV side (<100%). This is because examples in the HV side are considerably coinciding with other classes. In case of dielectric faults the patterns are comparatively easier than inter-turn faults.

Table 3: Accuracy (%) of different back propagation models at different conditions

Fault classes	SBP	LMA	BRP	RBP	OSS	ERN
LVIT	84.38	83.33	84.38	93.75	81.25	84.38
LVWG	100	100	100	100	100	100
HVIT	73.91	82.61	86.96	65.22	65.22	91.30
HVWG	85.71	83.33	83.33	88.89	78.57	93.75
D	92.31	100	100	100	98.08	98.08
H	100	100	100	100	100	100

In addition to the visualization, mean absolute percentage error (MAPE) is computed for performance analysis as

$$\mathrm{MAPE} = \frac{100\%}{\mathrm{n}} \sum_{t=0}^{n} \left| \frac{\mathrm{A(t)} - \mathrm{F(t)}}{\mathrm{A(t)}} \right|$$

Where A(t) is the actual value, F(t) is the measured value and the number of fitted points are *n*.
The MAPE describe above and execution times for different algorithms are presented in Table 4. It is seen that LMA is faster than other methods. Therefore it is suitable for real time fault detection.

Table 4: MAPE and time elapsed comparison between different models

Model	MAPE (%)	Execution time (sec)
LMA	1.25	0.713498
BRP	2.3012	17.201119
RBP	3.3221	1.419542
SBP	51.5789	0.637860
OSS	12.8958	2.209640
ERN	2.8516	12.15294

Dielectric fault of transformer are analyzed using six different learning algorithms shown in Fig. 6. Where LMA shows best result, at 60 iterations it shows error 0.01, but for other methods it needs above 100 iterations.

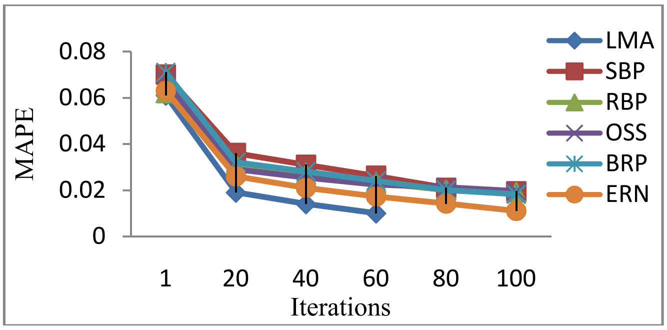

Fig.6: MAPE for different algorithms in case of dielectric fault

Fig.7. shows inter-turn short circuit fault of transformer. Here also LMA reach minimum error .009 at 80 iterations, but others reach almost 100 iterations.

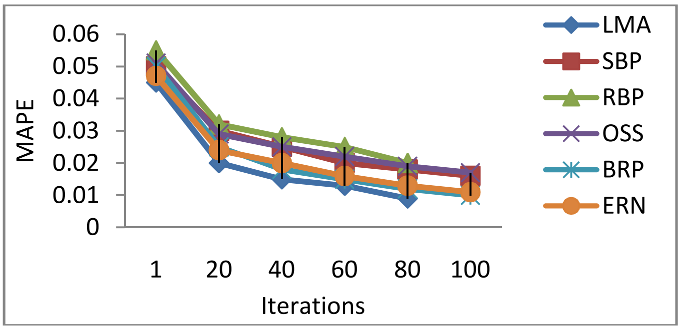

Fig. 7: MAPE for different algorithms in case of inter-turn fault

Even for winding to ground fault LMA shows best result, this is shown in Fig.8. LMA shows minimum error 0.008 at 60 iterations, but other methods show larger error at higher iterations.For above result, we can suggest that LMA is a good choice to detect transformer fault with minimum errors, execution time and iterations.

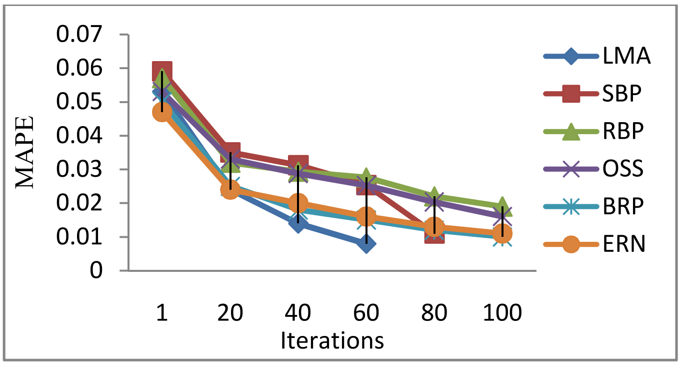

Fig. 8: MAPE for different algorithms in case of winding to ground fault

VI. CONCLUSION

In this study, a learning system is discussed with different kinds of learning algorithms. Our focus is to test different fault conditions of a single phase transformer with various learning algorithms. Various case studies have been studied including the addition of noise to the test values to evaluate the accuracy of the learning algorithm more precisely. We can clearly identify that LMA is the fastest method and it provides good accuracy and MAPE. Bayesian regulation back propagation and Elman back propagation network also provides better result but they require more time than any other models. It can be suggested that LMA is appropriate choice for real time fault detection. The technique is equally useful for large power transformer, although it is developed for a miniature transformer. The online inspection of faulty machines may be developed with the current implementation. This will be a topic of further study.

REFERENCES

[1] C. E. Lin, J. M. Ling, and C. L. Huang, "Expert system for transformer fault diagnosis using dissolved gas analysis," IEEE Trans. Power Del.,vol. 8, pp. 231–238, Jan. 1993.

[2] W. Yue, H. Zhuxiao, and Z. Jiaxiang, "Expediting cellulose insulation aging evaluation and life prediction thorough degree of polymerization measurements," in Proc. 3rd Int. Conf. Properties Appl. Dielectric Materials, 1988, pp. 328–331.

[3] A. Noro, K. Nakamura, T. Watanabe, and T. Morita, "Acoustic-basedreal-time partial d ischarge location in model transformer," in Proc. IC-SPAT, pp. 1077–1082.

[4] J. Bak-Jensen, B. Bak-Jensen, and S. D.Mikkelsen, "Detection of faults and aging phenomena in transformers by transfer functions," IEEE Trans. Power Del., vol. 10, pp. 308–314, Jan. 1995.

[5] S. H. Horowitz and A. G. Phadke, Power System Relaying, John Wiley & Sons, Inc, 1992.

[6] A. G. Phadke and J. S. Thorp, "A new computer-based flux restrained current-differential relay for power transformer protection," IEEE Trans. on Power Apparatus and System, vol. PAS-102, no. 11, pp. 3624-3629, November 1983.

[7] T. S. Sidhu and M. S. Sachdev, "On-line identification of magnetizing inrush current and internal faults in three-phase transformers," IEEE Trans. on Power Delivery, vol. 7, no. 4, pp. 1885-1891, October 1992.

[8] P. Bastard, P. Bertrand, and M. Meunier, "A transformer model for winding fault studies," IEEE Trans. on Power Delivery, vol. 9, no. 2, pp. 690-699, April 1994.

[9] S. M. Islam and G. Ledwich, "Locating transformer faults through sensitivity analysis of high frequency modeling using transfer function approach," Proc. of IEEE International Symposium on Electrical Insulation, pp. 38-41, June 1996.

[10] H. Wang and K. L. Butler, "Modeling transfor-mers with internal incipient faults," IEEE Trans. on Power Delivery, vol. 17, no. 2, pp. 500-509, April 2002.

[11] Y. Zhang, X. Ding, Y. Liu, and P. J. Griffin, "An artificial neural network approach to transformer fault diagnosis," IEEE Trans. on Power Delivery, vol. 11, no. 4, pp. 1836-1841, October 1996.

[12] M. G. Morante and D. W. Nocoletti, "A wavelet -based differential transformer protection," IEEE Trans. on Power Delivery, vol. 14, no. 4, pp. 1352-1358, October 1999.

[13] O. A. S. Youssef, "A wavelet-base technique for discrimination between faults and magnetizing inrush currents in transformers," IEEE Trans. on Power Delivery, vol. 18, no. 1, pp. 170-176, January 2003.

[14] P. Purkait and S. Chakravorti, "Wavelet transform-based impulse fault pattern recognition in distribution transformers," IEEE Trans. on Power Delivery, vol. 18, no. 4, pp. 1588-1589, October 2003.

[15] Arthur Earl Bryson, Yu-Chi Ho (1969). Applied optimal control: optimization, estimation, and control. Blaisdell Publishing Company or Xerox College Publishing. pp. 481.

[16] Paul J. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, 1974

[17] Alpaydın, Ethem (2010). Introduction to machine learning (2nd ed. ed.). Cambridge, Mass.: MIT Press. p. 250. ISBN 978-0-262-01243-0. "...and hence the name backpropagation was coined (Rumelhart, Hinton, and Williams 1986a)."

[18] Rumelhart, David E.; Hinton, Geoffrey E., Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors". Nature 323 (6088): 533–536. doi:10.1038/323533a0.

[19] Tom M. Mitchell, Machine Learning, McGraw-Hill Science/Engineering/Math; (March 1, 1997), page no. 93