Condition Monitoring of Oil-Impregnated Paper Bushings Using Extension Neural Network, Gaussian Mixture and Hidden Markov Models

W. S. Miya, L. J. Mpanza and T. Marwala
School of Electrical and Informantion Engineering
University of th Witwatersrand
Johannesburg, South Africa
Tshilidzi.marwala@wits.ac.za

Fulufhelo V. Nelwamondo
Modeling and Digital Science
Council for Scientific and Industrial Research
Johannesburg, South Africa
vnelwamondo@csir.co.za

Abstract— In this paper, a comparison between Extension Neural Network (ENN), Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) is conducted for bushing condition monitoring. The monitoring process is a two-stage implementation of a classification method. The first stage detects whether the bushing is faulty or normal while the second stage classifies the fault. Experimentation is conducted using dissolve gas-in-oil analysis (DGA) data collected from bushings based on IEEEc57.104; IEC60599 and IEEE production rates methods for oil-impregnated paper (OIP) bushings. It is observed from experimentation that there is no major classification discrepancy between ENN and GMM for the detection stage with classification rates at 87.93% and 87.94% respectively, outperforming HMM which achieved 85.6%. Moreover, HMM fault diagnosis surpasses those of ENN and GMM with a classification of 100%. However, for diagnosis stage HMM outperforms both ENN and GMM with 100% classification rate. ENN and GMM have considerably faster training and classification time whilst HMM's training is time-consuming for both detection and diagnosis stages.

Keywords—component, formatting, style, styling, insert *(key words)*

I. INTRODUCTION

Advancement in electric power transmission and distribution systems is critical for every nation's growth and development; however, key to these systems are expensive equipments such as power transformers [1]. According to Shoureshi, *et al* [2], 70% of transformer faults result from bushing related problems. In case of the utilization of oil-impregnated paper (OIP) bushing, incipient failure is indicated by the gaseous products that develop whenever a transformer is subjected to electrical and thermal stresses [3]. Reliability and accurate functioning of transformers influences both electric power availability of the supplied area as well as the economical operation of a utility [4]. In order to achieve the above, methodologies of incipient fault detection on transformer bushings are highly demanded by utilities. Numerous on-line condition monitoring systems such as partial discharge (PD) monitoring as well as dissolve gas-in-oil analysis (DGA) are currently used; the latter being the most favourable [4]. However, applying DGA methods such as key gas and ratio methods independently leads to non-automatic diagnostic rules, thus knowledge from new data samples cannot be acquired. In order to achieve automated monitoring systems, computational intelligence techniques are employed in conjunction with DGA data. These methods permit adaptive process for significant and new information [5].

In this paper, ENN, GMM and HMM are used to classify data obtained from OIP bushings via a DGA experiment. The overview of DGA, three computational intelligence techniques and a committee of classifiers is presented in Section II. In addition, an overall system design for the given application is detailed in Section III. Section IV discusses the results whilst Section V presents drawn conclusions.

II. BACKGROUND

DGA is considered the most crucial oil test for insulating liquids in electrical apparatus [5]. ENN has shown to overcome the drawbacks of the traditional neural networks [4]. In addition, GMM and HMM have been widely used as a classification tool for pattern recognition, particularly in speech and face recognition [6].

A. Dissolve Gas-in-oil Analysis

Over a long period of operation, internal insulation of an OIP bushing begin degrading when subjected to thermal and electrical stresses. The DGA technique involves taking a sample of oil randomly from the bushing and extracting the gases from the oil for measurement purposes [5]. The gases extracted from the oil sample are considered as fault indicators and the patterns as well as amounts at which they are generated characterize a possible fault in a bushing. Thus, the type and magnitude of gas determines the nature of the fault in the bushings. A fault diagnosing criteria is derived from the Key gas method specified in IEEE standard c57.104 and it is outlined in Table I [7].

1954

B. *B. Extension Neural Network (ENN)*

ENN is a relatively new neural network topology applied as a classification pattern system. It is a combination of extension theory and artificial neural networks. This technique has gained wide application in pattern recognition problems such as bearing fault classification as well as partial discharge recognition [8], [9].

Fault	Principal Gas	Gas Proportions
Corona	Hydrogen (H_2)	H_2: 85% CH_4: 13%
Arcing	Acetylene (C_2H_2)	H_2: 60% C_2H_2: 30%
Overheated oil	Ethylene (C_2H_4)	C_2H_4: 63% C_2H_6: 19%
Overheated cellulose	Carbon Monoxide (CO)	CO: 92%

ENN uses a modified extension distance (ED) to evaluate similarities between the given data and cluster centers, in addition, it adapts well when provided with new and significant information [10]. ENN is a two layer network with input layer that receives input training patterns and the output layer that represents classes in which the incoming data is allocated accordingly. The schematic structure of ENN is depicted in Fig. 1.
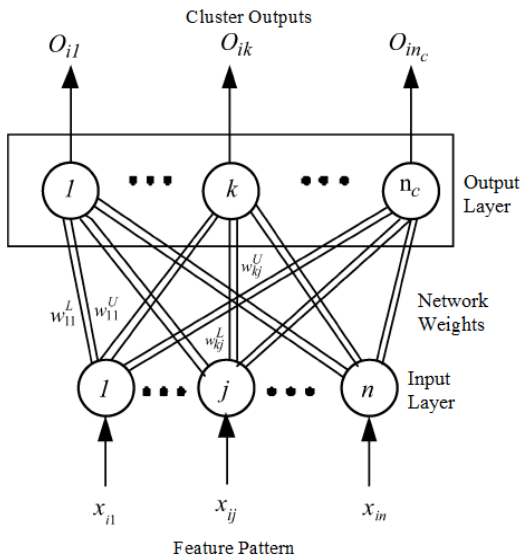

Fig. 1 Structure of extension neural network [10]

The set-up of ENN is such that a number of input nodes, n, correspond to the number of features in each training pattern and likewise, the number of classes or pattern clusters is equal to the number of output nodes, n_c. A set of two weights (i.e. upper and lower) connect each input node to an output node

[8]. The operation mode of ENN entails input layer nodes receiving input feature patterns and further employs a set of weighted parameters to generate an image of the input pattern [9].

C. *Gaussian Mixture Model (GMM)*

GMM is a probabilistic model that is formed by a weighted linear combination of Gaussian distributions. GMM is characterized by three parameters, expressed as:

$$\lambda = \{\omega, \mu, \Sigma\} \tag{1}$$

where ω, μ, Σ are vectors of the weights, means and co-variances corresponding to the features, respectively [11]. A mixture of Gaussians is given by a superposition of K Gaussian densities described in (2)

$$p(x \mid \lambda) = \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k) \tag{2}$$

where π_k, is the mixing coefficient, x refers to the feature data and N represents a Gaussian distribution. The validity of π_k parameter is achieved if the conditions in (3) are satisfied [11].

$$0 \le \pi_k \le 1 \quad \text{and} \quad \sum_{k=1}^{K} \pi_k = 1 \tag{3}$$

Initializing values of the model given in (1) is conducted through the use of maximum likelihood. The maximum likelihood of a GMM is computed from a distribution function and is given by:

$$\ln p(x \mid \lambda) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k) \tag{4}$$

where N is the number of data examples [11]. The model parameters are estimated using an Expectation-Maximization (EM) algorithm. The EM algorithm finds the parameters that increase the likelihood of cluster from the given data.

D. *Hidden Markov Model (HMM)*

HMM is defined as a set of states connected by a set of transitions and emitting an output during each transition [12]. The states of the model are not directly observable; hence, the model is called 'hidden' [12]. An ergodic or fully connected HMM was implemented for this work such that every state can be reached from every other state of the model [11]. An HMM implemented in this work is characterized by the number of states in the model, N, the number of distinct observation symbols per states, M, and the model parameters presented in (5), (6) and (7) [13].

$$A = \{a_{ij}\} = P[q_{t+1} = S_j \mid q_t = S_i], \quad 1 \le i, j \le N \tag{5}$$

$$B = \{b_{jk}\} = P[v_k @ t \mid q_t = S_j], \quad \begin{array}{l} 1 \le j \le N \\ 1 \le k \le M \end{array} \tag{6}$$

$$\pi_i = P[q_1 = S_j], \quad 1 \le i \le N \tag{7}$$

$$\lambda = (A, B, \pi) \qquad\qquad (8)$$

where (5), (6) and (7) refers to state transition, observation symbol and initial state probability distribution, respectively. (8) is a compact notation of HMM [13].

E. Committee of Classifiers

If different classifiers are used in pattern recognition problems, there is a need to integrate them in order to improve efficiency and accuracy of the classification process [14]. This releases classification dependency on a single classifier and instead, results are based on a consensus decision of base classifiers [14]. Base classifiers operate concurrently during classification and their outputs are integrated to form a final output. The reader is referred to [14] for a theory-based discussion of majority voting combination strategy that is used in this paper. Its structure is provided in Fig. 2.
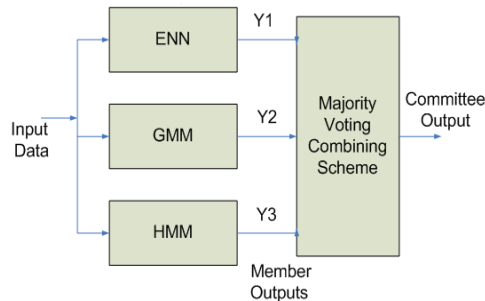

Fig. 2 Structure of committee of classifiers

III. SYSTEM DESIGN

System design implies the implementation methodology designed for monitoring OIP bushings. The proposed method comprises three phases, i.e. data analysis, detection and diagnosis (i.e. classification sub-system) phases. The first stage of a classification system identifies whether the bushing is faulty or not and the second stage determines the nature of the detected fault. The proposed system is represented in a block diagram shown in Fig 3.

A. Data Analysis

This is a primary phase of the implemented method where collection of data, inspection and preprocessing of data is examined.

1) Data Collection and Specification: Data were collected from the DGA tests conducted according to IEEEc57.104, IEC60599 and production rates methods [7], [15]. Combustible and non-combustible gases (i.e. CH_4, C_2H_6, C_2H_4, C_2H_2, H_2, CO, CO_2, N_2, O_2 and TDCG) were recorded for each experiment and some of these gases characterize the condition of the bushing at a particular instance. Extracted gases are provided in parts per million (ppm) and the status of

the bushing condition corresponding to a certain pattern is indicated with a zero and one for a faulty and non-faulty bushing, respectively.
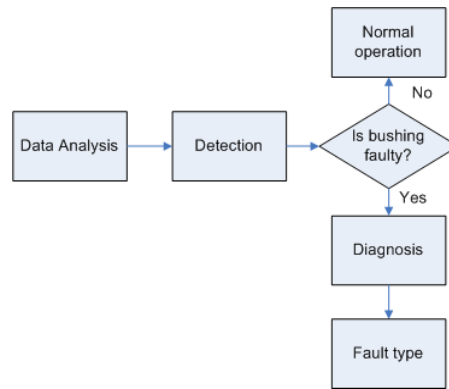

Fig. 3 Block diagram of a proposed solution

Feature extractors comprising time and time-frequency domain analysis were examined and it was discovered that this process is not a requirement for the problem at hand since data were provided with measured gases forming necessary features [11]. However, in a case where raw data are provided, feature extraction process is a necessity since it ensures that data is reduced to only minimal features that characterize the system, hence, accurate and faster classification process is achieved.

2) Data Inspection: During this step data quality is improved by examining outliers as well as missing data. Only one outlier was identified as a possible recording error because it was located in isolation from other data sets and it was then eliminated [16]. There are several methods that can be used to deal with missing data, e.g. list-wise deletion, multiple imputation and neural network approach [11]. However, in this work input patterns with missing data were discarded since the ratio of deleted data was very small compared to the training instances that remained [17, 18]. It was assumed that data entries with recorded zeros are due to the absence of results during the course of an experiment or to an insignificantly low outcome. Therefore, all training instances with recorded zeros were also eliminated since they added no value to the training process, instead, they could have prolonged the training time unnecessarily. In addition, non-faulty gases were also discarded since they do not characterize a faulty condition. Only faulty gases were used for this project as features.

3) Data Preprocessing: This process is required to ensure accurate classification operation and it involves scaling, distribution and partitioning of data. Minimum and maximum values of the data sets are acquired and manipulated according to (9), in order to accomplish a normalized value.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad \textbf{(9)}$$

where X_{norm} refers to the normalized value, X_{min} and X_{max} represent the minimum and maximum values respectively within the data sets and X refers to the actual data [17, 18]. This operation aims at normalizing data to fall between zero and one, ensuring that larger values in a data set are not treated as a priority than others, thus leading to a biased classifier. Additionally, a normal distribution of data was conducted to ensure that every pattern has the same likelihood of occurring. In order to avoid under-fitting and over-fitting, data was partitioned into three equal sets; training, validating and testing data. After partitioning, each set had 11 704 instances. Out of the training instances, 1 000 balanced patterns (500 faulty and 500 non-faulty) were used when developing base classifiers in order to prevent the network from being biased and also from memorizing the training data.

B. Classification System

1) Detection Stage: Supervised learning was employed in developing an ENN toolbox for the given application. The reader is referred to the existing literature on ENN by Wang [3] for theory details of the developed toolbox. The learning process of ENN requires the adjustment of weights and the learning rate is a critical consideration during this process because it determines the extent at which the weights change for each training step [11]. Once this is achieved then the final optimal weights are stored. Each training pattern adjusts the network's weights and cluster centers depending on learning rate [9]. The algorithm takes longer to converge if η is too small and it diverges if the learning rate is too large. As a result, numerous ENNs were developed with typical learning rates, η, varying from 0.01 to 0.1 with the aim of obtaining the network with less training error. The best network was recorded and trained with a large value of epochs, 1 000, so that it can be observed when the training error tends to be constant. The tested ENNs' results with various learning rates are demonstrated graphically in Fig. 4. From the graph, it is seen that the best learning rate and training epochs are 0.09 and 13 respectively.

The resulting ENN structure for this stage has two output nodes corresponding to two classes or clusters (i.e. faulty or non-faulty bushing) and seven input nodes which are equal to the number of faulty gases. Again the reader is referred to [3] for a discussion on the algorithm details of ENN operation phase. In this phase, new data is presented to the network in order to test if accurate classification is achieved.
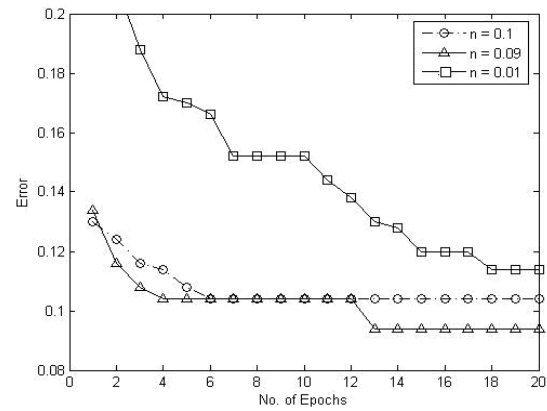

Fig. 4 The learning curve for ENN detection stage

The development of second and third classifiers entailed a construction of two GMMs and HMMs corresponding to the number of output conditions for this stage. The number of inputs and outputs used for a GMM and HMM models are similar to those employed in developing an ENN. GMM and HMM are statistical classifiers, hence, their development and operation is comparable. In this stage, two Gaussians and HMMs are trained using two sets of input, the faulty and the non-faulty sets (e.g. hmm0 and hmm1). GMM is optimized using Expectation Maximization (EM) algorithm during a training process. HMM's are optimized using EM embedded with Baulm-Welch algorithm. The two algorithms maximize the likelihood of detecting the probable output cluster for a given testing data. Two learning curves for HMM's developed during training in the first stage are presented in Fig. 5 where it is indicated that the training stops when the log-probability has stabilized. The results obtained from training and testing all three models are presented in Section IV.
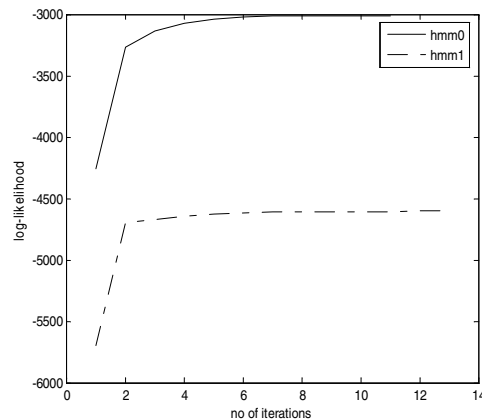

Fig. 5 The learning curve for HMM detection stage

2) Diagnosis Stage: In this stage, only patterns of the faulty bushings (i.e. instances corresponding to zero values in the target output) are considered for classification. However,

considering the fact that detection classification is not 100% accurate, it can be concluded that some of these input patterns may not result to a known faulty bushing. Hence, it is necessary to include an unknown fault condition amongst the faults specified in Table 1. Furthermore, unknown faults refer to those possible faults not specified by IEEE c57.104 standard. Ultimately, OIP bushings undergo five conditions which correspond to five output clusters used in developing all classifiers for a diagnosis stage. As a result, ENN's structure composed of five output nodes and five Gaussians as well as an HMM were developed corresponding to the number of clusters in this stage. In the same way as in the detection stage, EM algorithm is used for optimizing GMM's and HMM's maximum likelihood for the classification of the nature of the fault.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Experiment was performed in MATLAB simulation environment with an Intel^{TM} Core^{TM}2 Duo computer operating at a processor speed of 3 GHz. ENN simulation procedures entail the use of optimal learning rate, a minimum number of epochs (also used as a training-stopping criteria) as well as 1 000 and 500 training examples for detection and diagnosis stages, respectively. Simulation process for the other two methods involved further data processing where data was rearranged according to five key-gas patterns. These observed patterns were then used to develop two and five models, depending on the system stage. Results are recorded in terms of model's accuracy which is one of the network parameters. Accuracy is computed using elements of a confusion matrix which is defined by (10).

$$Accuracy = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} \qquad (10)$$

where TP, TN, FP and FN are true positive, true negative, false positive and false negative respectively. TP is when a fault is classified correctly whereas FP is when a non-fault is classified as a fault. On the other hand, TN is when a non-fault is classified correctly, while FN is when a non-fault is incorrectly classified as a fault. Table II compares all three models in terms of accuracy as well as training and classification duration for the first phase of fault diagnosis.

Table II also indicates that GMM and ENN give approximately the same performance, surpassing HMM by 2.3 % whilst a combining scheme indicated no improvement in classification. Moreover, ENN trains faster than all other models for detection stage and it has shorter learning and classification times for diagnosis stage. Results for the diagnosis stage are presented in Table III.

TABLE II
COMPARING PERFORMANCE OF ENN, GMM, HMM AND A COMMITTEE FOR THE DETECTION STAGE

Criteria	ENN	GMM	HMM	Committee
Accuracy (%)	87.93	87.94	85.6	87.94
Training time (sec)	0.78	3.59	7.09	n/a
Classification time (sec)	0.17	0.13	0.67	n/a

TABLE III
COMPARING PERFORMANCE OF ENN, GMM, HMM AND A COMMITTEE FOR THE DIAGNOSIS STAGE

Criteria	ENN	GMM	HMM	Committee
Accuracy (%)	98.08	86	100	98
Training time (sec)	0.82	3.67	8.13	n/a
Classification time (sec)	0.54	0.99	1.63	n/a

Classifier comparisons between stages indicate an improvement in classification performance. However, this improvement seems to be of less significance. Detection stage identifies margins set by faulty gases. On the other hand, diagnosis stage evaluates the concentration of the instance of gases and identifies the most probable fault to be associated with the given instance. These comparisons are presented in Fig. 6. The black histogram indicates the detection stage whereas the other one is for classification.
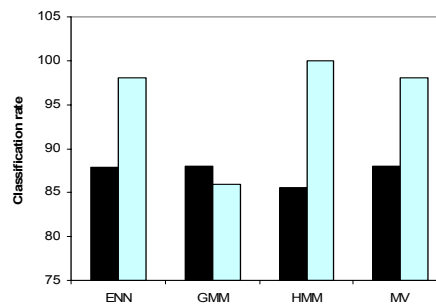

Fig. 6 Performance comparison between detection and diagnosis stages

V. CONCLUSION

A bushing condition monitoring scheme based on ENN, GMM and HMM has been presented. The overall system is designed such that it examines the condition of the bushing and if the condition is faulty, it classifies the identified fault. The operation of the system was executed using DGA data based on IEEEc57.104; IEC60599 and IEEE production rates methods for OIP bushings. A minimum number of epochs was used as a stopping criteria for the training process of ENN

which also depends upon an optimal rate. Results indicate that all three models are suitable for bushing fault detection and diagnosis application. GMM outperforms both ENN and HMM in detecting the faults and HMM diagnose faults the best. It is also evident from the results that ENN trains and classifies the fastest for diagnosis stage. Moreover, the results depict that ENN has shorter learning rates while GMM classifies fastest. It is also evident from experimentation results that a committee of classifiers contributed no significant improvements in the overall classification process.

REFERENCES

[1] Pathak J, Jiang Y, Honavar V, McCalley J. *Condition Data Aggregation with Application to Failure Rate Calculation of Power Transfomers.* Proceeding s of the 39th Hawaii International Conference on System Sciences-2006, Page(s) 241

[2] Shoureshi R, Pernama V, Wood R, Swartzendruber R, Simoes M. *Intelligent Transformer Monitoring System Utilizing Neuro-Fuzzy Technique Approach* http://www.pserc.org/cgipserc/getbig/publicati/reports/2004report/shureshi_smart_sensor_final_report.pdf, Last accessed 27 September 2007.

[3] Wang M H. *Extension Neural Network for Power Transformer Incipient Fault Diagnosis*, IEE Proc.Gen. Transm. Distrib., Vol. 150, No. 6, 2003, Page(s). 679-685.

[4] Wang Z. *Artificial Intelligence Applications in the Diagnosis of Power Transformers Incipient Faults.* Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fullmillment of the requirement for the degree of Doctor of Philosophy in Electrical Engineering,http://scholar.lib.vt.edu/theses/available/etd-08102000-21510032/unstricted /Chapter04.pdf, Last accessed 27 September 2007.

[5] Duval M. *Dissolved Gas Analysis: it can Save Your Transformer.* IEEE Electr. Insul. Mag. Vol. 5, No.6, Page(s) 22-27

[6] Schlapbach A, Bunke H. *Off-line Writer Verification: A Comparison of a Hidden Markov Model (HMM) and a Gaussian Mixture Model (GMM) Based System*, Proceedings of the Tenth International Workshop on Frontiers in Handwriting Recognition, October 2006, La Baule. Page(s) 386-396

[7] IEEE Standards, Mineral Oil Impregnated Electrical Equipment in Service-IEEE Guide to Interpretation of Dissolved and Free Gases Analysis, IEEE c57.104, Page(s) 1-78.

[8] Mohamed S, Tetty T, Marwala T. *An extension Neural Network and Genetic Algorithm for Bearing Fault Classification.* Proceedings of the International Joint Conference on Neural Networks, 2006, Page(s) 7673-7679.

[9] Wang M. *Partial Discharge Pattern Recognition of Current Transformers Using an ENN.*, IEEE Transactions on Power Delivery, Vol. 20 No.3, 2005, Paga(s). 1984-1990.

[10] Wang M H, Hung C P. *Extension Neural Network and its Application.* Neural Networks, Vol. 16, No. 5-6, 2003, Page(s) 779-784.

[11] Bishop C B. *Pattern Recognition and Machine Learning.* Singapore, Springer Science +Business Media, LCC, 2006.

[12] Satish L, Gururaj B I. *Use of Hidden Markov Models for Partial Discharge Pattern Classification.* IEEE Transactions on Electrical Insulation, Vol 28, No. 2, 1993

[13] Rabiner L R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, Vol. 77, No. 2, 1989.

[14] Kittler J, Hatef M, Duin R, Matas J. *On Combining Classifiers.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, 1998.

[15] IEEE Standards IEC60599, IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, Page(s) 1-36.

[16] Motulsky H J, Brown R E. Detecting Outliers when Fitting Data with Non-linear Regression –A New Method Based on Rhobust Non-linear Regression and the False Discover Rate. http://www.biomedcentral.com/1471-2105-7-123.pdf, Last accessed 27 September 2007.

[17] Patel P B, Marwala T. *Forecasting closing Price Indices using Neural Networks.* IEEE International Conference on Systems, Man and Cybernetics, 2006. ICSMC apos; 06. Volume 3, Issue, 8-11 Oct. 2006, Page(s). 2351 - 2356.

[18] S.M. Dhlamini, F.V. Nelwamondo and T. Marwala. "Condition monitoring of HV bushings in the presence of missing data using evolutionary computing", *WSEAS Transactions on Power Systems*, 2006, Page(s). 280-287.