Artificial Neural Network and Rough Set for HV Bushings Condition Monitoring

L.J. Mpanza*, T. Marwala**
Department of Electrical and Electronics Engineering,
University of Johannesburg
Johannesburg, South Africa
ljmpanza@gmail.com, tmarwala@uj.ac.za

Abstract - Most transformer failures are attributed to bushings failures. Hence it is necessary to monitor the condition of bushings. In this paper three methods are developed to monitor the condition of oil filled bushing. Multi-layer perceptron (MLP), Radial basis function (RBF) and Rough Set (RS) models are developed and combined through majority voting to form a committee. The MLP performs better that the RBF and the RS is terms of classification accuracy. The RBF is the fastest to train. The committee performs better than the individual models. The diversity of models is measured to evaluate their similarity when used in the committee.

I. INTRODUCTION

Reliability engineering is concerned with the development of approaches to determine the maintenance strategies for monitored equipment so that the usability and the availability are optimized. To effectively implement a schedule maintenance plan, an efficient condition monitoring system is necessary.

The primary objective of a condition monitoring system is to detect changes in the characteristics of the equipment. This change should be quantified and analyzed to give information about the current condition of the equipment. This information is then used to advise the maintenance personnel. The advantage of a condition monitoring system is that faults can be identified early and corrected to avoid breakdowns which lead to unplanned and prolonged equipment downtime.

Transformers are crucial parts of the power transmission and distribution system. Breakdown of transformer leads to frustration to the consumer and financial loss to the supplier. Bushings are used as insulators in transformers. Studies have shown that most transformer failures are due to bushings failure [1]. Therefore, as an initiative to monitor and prevent transformer failure, the condition monitoring of bushings is advised [1].

This paper presents methods for monitoring oil-impregnated paper (OIP) bushings using dissolved gas analysis (DGA). The rest of the paper is outlined as follows: Section two presents the literature review. Section three describes the theory necessary for the development of the model. Section four presents the experimental setup. Section five details the results arrived at and the discussion. The conclusion then follows.

II. LITERATURE REVIEW

Artificial intelligence methods have been used extensively in condition monitoring. Monitoring bushings has had its share of intelligence method. Most notably the use of multi-layer perceptron (MLP), radial basis function (RBF), support vector machines (SVM), extension neural networks (ENN), Gaussian mixture model (GMM), hidden Markov model (HMM) by [2,3,4,5]. Other researchers [3,4] also used a committee of classifiers to improve the results. Rough sets (RS) have also been used to detect faults in bushings [6].

In this study three methods are developed based on the MLP, RBF and RS. A committee of networks is developed by combining these techniques. What is unique in this study is that a combination of neural network with rough set theory is modeled and the diversity of the classifiers is measured. The measure of diversity is done to justify the combination of neural networks with rough set instead of neural networks with each other.

III. SYSTEM FORMULATION

A. Dissolved gas analysis

Dissolved gas-in-oil analysis (DGA) is used to analyze the gases developed in oil-impregnated equipment. To monitor the condition of insulated paper bushings, an oil sample is taken on a regular basis. The gases are extracted from the oil using chromatography. The nature and the amount of gases developed can help experts to identify and isolate a fault in its development stage. The gases associated with DGA for bushings are: Hydrogen (H2), methane (CH4), ethylene (C2H4), ethane (C2H6), acetylene (C2H2), carbon monoxide (CO), carbon dioxide (CO2), nitrogen (N2) and Oxygen (O2).

The presence of simple hydrocarbons indicates a breakdown of insulating oil as a result of either electrical or thermal stress. Elevated levels of hydrogen, carbon oxides and methane indicates the decomposition of the insulating paper.

IEEEc57 [7] and ICE 60599 [8] details the methods used to identify faults in DGA. These are, among others, Key gas analysis and Ratio analysis. Transformer bushings are installed in different operation conditions around the world and are subject to a range of external influences which can affect the development of gases in the oil [9].

The abovementioned methods fall short in...
detecting faults due to their generalized nature that fails to cater for localized operation condition.

We develop a tool for analysis of bushings data based on the history of the operation of these bushings. This tool learns the bushings characteristic to develop a customized condition monitor. Two neural networks and rough set models are developed and tested for this purpose.

B. Artificial Neural Network

Artificial Neural networks (ANNs) are universal approximators. They are able to map almost any input-output relationship. This is attributed to the fact that ANNs are based on the biology of a human brain. They try to mimic the operation of a brain by the use of neurons. Each neuron is connected to the other using synaptic weights. A neuron can have multiple inputs and one output. The input (x) is transformed into output y by summing the weighted values and adding a bias and then using the activation function as in (1) [10].

\[ y = f_{out} \sum_{i=1}^{N} w_{i}x_{i} + b \]  

(1)

Using the training algorithm, the weights are adjusted for a given input to be transformed into a target output and reduce the error. This process is called error back-propagation supervised learning. A number of algorithms are available to train ANN such as the scaled conjugate gradient, conjugate gradient, quasi-newton and other optimization algorithms. Scaled gradient is used in this experiment as it offers faster convergence and higher accuracy. Two ANNs are investigated and these are the multi-layer perceptron (MLP) and radial basis function (RBF).

1) Multi-Layer Perceptron (MLP)

MLP is a network of connected neurons forming two or more layers. A two layer perceptron is used in this work as it is able to model any input-output relationship [10]. Data is presented to the MLP as input (x) and is transformed into output (y) using (2) [10].

\[ y = f_{out} \sum_{j=1}^{M} w_{j}f_{in} \sum_{i=1}^{N} w_{ij}x_{i} + b_{j} + b \]  

(2)

Where \( x_{i} \), \( w_{ij} \), \( b_{i} \), \( f_{in} \), \( w_{j} \), \( b \), \( f_{out} \) and \( y \) are input vectors, hidden weights, hidden bias, hidden activation function, output weights, output bias, output activation function and output respectively. A logistic function is used in the output layer and a hyperbolic tangent activation function is used in the hidden layer. The combination of the two activation functions offers better results [10]. The weights are adjusted using the scaled gradient conjugate method. The bias is used to force an output even when there is zero input.

2) Radial Basis Function (RBF)

Similar to the MLP, the RBF is a network of neurons forming two layers with radially activated functions in the hidden nodes. RBF has unity weights (\( w_{j} = 1 \)) in the hidden layer. The input \( x_{i} \) is transformed into output \( y \) using (3) [10].

\[ y = \sum_{i=1}^{N} w_{i} \phi_{i}(x) + b \]  

(3)

where, \( \phi_{i} \) is a hidden activation function. A linear function is chosen for the output layer and a thin plate spline function is chosen for the hidden layer. The thin plate spline is expressed as in (4).

\[ \phi(x) = r^2 \log(r) \]  

(4)

where \( r \) is the radial distance given by (5).

\[ r = \|x_i - x_m\| = \sqrt{x_i^2 - x_m^2} \]  

(5)

where \( x_i \) is the input and \( x_m \) is the mean of input values. The RBF is trained using backpropagation which used the sum of squared error to optimize the performance.

C. Rough set theory

Rough Set (RS) is a rule based pattern recognition tool. It is used when dealing with vagueness, uncertainty and imprecision in the data set. It differs from classical set in that, instead of using purely defined sets with hundred percent membership it uses upper and lower approximation to approximate sets without purely defined boundaries. This property allows the RS to classify uncertain objects into sets which would have been impossible by just using classical crisp set theory.

RS model is defined in terms of an Information System (IS). IS is defined mathematically as in (6) [11].

\[ \Lambda = (U, A) \]  

(6)

where \( U \) is a non-empty set of objects is called the universe and \( A \) is a non-empty set of attributes of an object \( x \in U \). For any attribute \( a \in A \), there exist a set \( V_{a} \) of values such that \( a : V \to V_{a} \). In supervised learning, the learning model is presented with inputs and a target output. An IS with a target decision attribute \( d \) is transformed in a decision system (DS) and is defined as \( \Lambda = (U, A \cup \{d\}) \). In binary classification system, \( d \) takes value 1 or 0.

Central to RS is the concept of indiscernibility. Using a subset of attributes to define objects \( x \) and \( y \), we say \( x \) and \( y \) are indiscernible if \( x \) cannot be distinguished from \( y \). Mathematically, if \( B \subseteq A \) for any DS, then

\[ (x, y) \in IND(B) \text{ iff } a(x) = a(y) \]

\[ a \in B, x, y \in U \]  

(7)

\( IND(B) \) is a representative of all objects that are indiscernibly in terms of \( B \) and is called an elementary set. Given a set \( X \) of interest based on the decision attribute \( d \), there is a possibility that objects cannot be fully definable using attribute \( B \). In classic set theory it is impossible to classify objects with similar attributes to different sets. To overcome this limit, RS defines two approximation sets. The lower approximation set \( BX \) is used for objects that completely belong to set \( X \) of interest, while the upper approximation set \( \bar{BX} \) is the set of objects that do not belong in set \( X \) with certainty. Equation (8) and (9) are used to find the lower and upper approximation [12].

\[ BX = \{x \in U : B(x) \subseteq X\} \]  

(8)

\[ \bar{BX} = \{x \in U : B(x) \cap X \neq \emptyset\} \]  

(9)
There are three regions of interest, these are, the inner region for objects in $\overline{B}X$, the outer region for objects not in $\overline{B}X$ and the boundary region for the objects that are uncertain. These are defined mathematically as $IB_x = BX$, $BN_x = BXB$ and $OB_x = U - BXB$ respectively. If the boundary region is empty ($BN_x = \emptyset$) then the system is crisp, else it is rough. The objects in $BN_x$ are classified into $X$ or $\overline{X}$ using a membership function. Rules are generated from the data and these rules are used to classify objects using the plausibility of belonging. The membership function is defined as (10) [12].

$$\mu^X_i : U \rightarrow [0,1] \text{ and } \mu^X_i = \frac{|B(x) \cap X|}{B(x)} \quad (10)$$

The rules extracted are represented in the form: if (condition) - then (decision). Theses rules are used to classify objects and are also used to define the events leading to decision table.

D. Classifier Ensemble

An ensemble of classifier is developed by combining different classifiers to perform one task. The reason for combining classifiers is to improve the classification accuracy by harnessing the abilities of diverse classifiers. The classifiers are combined using majority voting scheme.

E. Classifier Diversity

To develop a useful committee of classifiers it is necessary that different classifiers are used. The measure of how different classifier is called diversity. The diversity measure can either be structural based or outcome based. The later is used in this study. Several diversity measures have been proposed [13]. This study is only concerned with the pair-wise measure.

Given two classifiers, the following relationship is defined in table I [13].

<table>
<thead>
<tr>
<th></th>
<th>$D_{K_{correct}}$</th>
<th>$D_{K_{wrong}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{K_{correct}}$</td>
<td>$N^{11}$</td>
<td>$N^{10}$</td>
</tr>
<tr>
<td>$D_{K_{wrong}}$</td>
<td>$N^{01}$</td>
<td>$N^{00}$</td>
</tr>
</tbody>
</table>

$N^{11}$, $N^{01}$, $N^{10}$ and $N^{00}$ are the frequency of both classifier being correct, the $i$ classifier being correct, the $j$ classifier being correct and both classifiers wrong respectively.

Four diversity measures are defined for pair-wise: Q statistic, Correlation measure, Disagreement measure and Double fault measure.

1) Q statistic

Q statistic measures the degree of similarity and dissimilarity. Q statistic is measured by (11) [13].

$$Q_{i,j} = \frac{N^{11}N^{00} - N^{10}N^{01}}{N^{11}N^{00} + N^{10}N^{01}}, \quad (11)$$

where $i$ and $j$ are pairs of classifiers. The average of all pairs is computed to get the overall diversity of the ensemble.

2) Correlation measure

This measures the correlation of the outputs of the classifiers and is calculated using (12) [13].

$$P_{i,j} = \frac{N^{01}N^{10} - N^{00}N^{11}}{\sqrt{(N^{11}+N^{00})(N^{00}+N^{10})(N^{11}+N^{10})}} \quad (12)$$

3) Disagreement measure

This is the measure of the frequency of disagreement between the two classifiers. It is calculated using (13) [13].

$$D_{i,j} = \frac{N^{01} + N^{10}}{N^{11} + N^{00} + N^{10} + N^{01}} \quad (13)$$

4) Double fault measure

This is the measure of the frequency of both classifiers getting it wrong. It is measured using (14) [13].

$$D_{F_{i,j}} = \frac{N^{10}}{N^{11} + N^{00} + N^{10} + N^{01}} \quad (14)$$

IV. EXPERIMENT SETUP

The aim of the experiment is to develop three classifiers to use DGA data to detect faults in bushings. A MLP, RBF and RS classifiers are developed. The classifiers are then combined in an ensemble to form a committee of classifiers. Figure 1 shows the setup of the experiment. The diversity of the classifiers is measured to evaluate the effectiveness of the committee.

A. DGA Data

DGA data of 60966 bushings is used for developing the models. Ten inputs attributes are identified, these are: Hydrogen (H$_2$), methane (CH$_4$), ethylene (C$_2$H$_4$), ethane (C$_2$H$_6$), acetylene (C$_2$H$_2$), carbon monoxide (CO), carbon dioxide (CO$_2$), nitrogen (N$_2$), Oxygen (O$_2$) and total combustible gas (TCG). One output attribute is identified. Fault is a binary attribute, with 1 indicating healthy bushing and 0 indicating faulty bushing.

B. Data preprocessing

Preprocessing is done to improve the quality of the data by removing the outliers, dealing with missing values and verifying the data integrity, to name a few methods for improving data quality.
1) **Data Normalization**

To make the data suitable for development of classifiers, it is normalized to fall in the range 0 to 1. This is done to make the training easy and to avoid variables being treated as more important that others because of their order of magnitude. Function (15) is used for this purpose [10].

\[ X_{out} = \frac{X_{in} - X_{min}}{X_{max} - X_{min}} \]  

(15)

2) **Data discretization**

Since RS is a rule based classifier, it is necessary to transform continuous variables into discrete variable labels. This is done so that the rules are based on data categories which are countable, other than the set of real numbers which is uncountable. Equal Frequency Bin (EFB) is used for this purpose because of its simplicity and that it produces better results than the Equal Width Bin (EWB) [14].

The data is partitioned into 1400 training set, 600 validation set and 600 testing set. The classifiers are trained and tested in a Matlab development environment. They are evaluated in terms of classification accuracy, receiver operation characteristic (ROC) curve, the area under the ROC curve (AUC) and the training time.

Table II shows the result of the implementation.

<table>
<thead>
<tr>
<th></th>
<th>MLP</th>
<th>RBF</th>
<th>RS</th>
<th>Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TP</strong></td>
<td>307</td>
<td>306</td>
<td>297</td>
<td>307</td>
</tr>
<tr>
<td><strong>FP</strong></td>
<td>6</td>
<td>7</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td><strong>TN</strong></td>
<td>281</td>
<td>260</td>
<td>278</td>
<td>284</td>
</tr>
<tr>
<td><strong>FN</strong></td>
<td>6</td>
<td>27</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td><strong>AUC</strong></td>
<td>0.980</td>
<td>0.964</td>
<td>0.973</td>
<td>0.987</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>98.00</td>
<td>94.30</td>
<td>95.83</td>
<td>98.50</td>
</tr>
<tr>
<td><strong>Training time</strong></td>
<td>26</td>
<td>8</td>
<td>14</td>
<td>-</td>
</tr>
</tbody>
</table>

Table II. The characteristics of the MLP, RBF and RS. The table shows the confusion matrix, the area under the curve (AUC) accuracy of classification and the development time (s) of each classifier.

Figure 5 show the ROC curve (a) MLP ROC, (b) RBF ROC, (c) RS ROC and (d) Committee ROC.
Table III shows the measure of diversity of the committee of MLP, RBF and RS.

TABLE III. THE MEASURE OF DIVERSITY FOR THE MLP, RBF AND RS.

<table>
<thead>
<tr>
<th></th>
<th>MLP-RB</th>
<th>MLP-RS</th>
<th>RFB-RS</th>
<th>Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>N^1</td>
<td>563</td>
<td>570</td>
<td>552</td>
<td>-</td>
</tr>
<tr>
<td>N^00</td>
<td>9</td>
<td>7</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>N^01</td>
<td>3</td>
<td>5</td>
<td>23</td>
<td>-</td>
</tr>
<tr>
<td>N^01</td>
<td>25</td>
<td>18</td>
<td>14</td>
<td>-</td>
</tr>
<tr>
<td>Q_{ij}</td>
<td>0.9837</td>
<td>0.9718</td>
<td>0.8669</td>
<td>0.9400</td>
</tr>
<tr>
<td>P_{ij}</td>
<td>0.4341</td>
<td>0.3937</td>
<td>0.3333</td>
<td>0.387</td>
</tr>
<tr>
<td>Dis_{ij}</td>
<td>0.0467</td>
<td>0.0383</td>
<td>0.0617</td>
<td>0.0489</td>
</tr>
<tr>
<td>DF_{ij}</td>
<td>0.0150</td>
<td>0.0117</td>
<td>0.0183</td>
<td>0.015</td>
</tr>
</tbody>
</table>

V. RESULTS OBTAINED

The MLP performs better than the other classifiers in terms of accuracy. The fastest to train among the classifiers is the RBF, but it is the worst in terms of accuracy. RS takes second place in both accuracy and training time. The accuracy of the majority voting scheme outperforms all the stand-alone classifiers. The RS is desirable for its transparency in decision making. The advantage of the committee is that it contains both the classification accuracy of the MLP and the transparency of the RS.

The diversity measure shows that the MLP-RS is more diverse than the other combination. Q_{ij} measures similarity and Table II shows that the classifiers are able to recognize similar objects. Dis_{ij} is low, meaning the classifiers agree with each other while DF_{ij} is also low, meaning the classifiers do not commit double fault frequently. The diversity of the classifiers is low, but is still better that no diversity at all. The committee is diverse enough for the results to improve.

VI. CONCLUSION

Three classifiers were developed to monitor the condition of transformer bushings. These classifiers were implemented using the MLP, RBF and RS. The resultant classifiers are combined to form a committee of classifiers. The diversity of the classifiers is measured to evaluate the merit of combining them. The MLP classifier performs better than the RBF and the RS, but together they perform even better that they do individually. The diversity measure helps to understand the reason why the committee performs better that the stand-alone classifiers. For future development, an ensemble of more diverse classifiers can be implemented to investigate the diversity measure further.

REFERENCES