Abstract—This paper examines classification models using three classes of artificial neural networks (ANN). The first ANN uses Support Vector Machine activation functions. The second uses Multiple Layered Perceptron (MLP) activation functions with automatic relevance detection (ARD). And the third uses Radial Basis activation functions (RBF). In this work the decision is taken to remove or leave a bushing in service based on analysis of bushing parameters using RBF, SVM and MLP. The work finds that the RBF converges to a solution faster than both SVM and MLP. The MLP is the best tool of the three for analyzing large amounts of non-parametric non-linear data. MLP is the most accurate of the three networks. ARD reveals that Methane was the most common cause for action on bushings tested using DGA during the two years evaluation period.

Index Terms—multiple layered perceptron, radial basis, support vector machines, bushing, diagnosis, dissolved gas analysis.

V. INTRODUCTION

This paper evaluates data from dissolved gas analysis (DGA) tests and additional field services staff inputs using radial basis functions (RBF), support vector machines (SVM) and multiplayer perceptrons (MLP) with automatic relevance detection (ARD). This is because DGA is the standard and the most widely used diagnostic tests for transformer and bushing in industry [11]. Other diagnostic methods include On-line Partial Discharge Analysis (PDA), Infrared scanning, On-line Vibro-acoustic Analysis (VAA) and On-line Power Factor Monitoring (PFM). The application of an artificial neural network (ANN) to an integrated technique incorporating all of these methods increases the reliability of diagnosis and can serve a solid foundation for operational decisions.

The aim of the work is to analyze temperature, pressure as well as gaseous content from different types of bushings using two types of artificial neural networks to decide whether or not to remove a bushing from service. The data was taken over a period 2.5 years from bushings working at rated voltages of 132kV, 275kV and 400kV. The results from this work can be applied for on-line monitoring of bushings, thus allowing large amounts of data to be interpreted so that decisions can be taken onsite about whether or not a bushing needs to be replaced. The three neural networks compared in this classification analysis are RBF, SVM and MLP with ARD.

II. INPUT AND OUTPUT PARAMETERS

The input variables are measured by maintenance staff and generally analyzed manually according Cigre recommendations [23]. As it is customary in neural networks, the variables are normalized to fall within 0 and 1 bounds. Normalizing ensures that the input to the neurons lie in the range where the activation functions are defined. Normalizing also allows for accurate determination of relevance of a variable used in the MLP with ARD.

The fourteen input parameters were as follows:
Pressure before sampling; Bushing temperature; Bushing oil level before sampling; Bushing oil level after sampling; Bushing pressure after sampling; Hydrogen; Oxygen; Nitrogen; Methane; Carbon Monoxide; Carbon Dioxide; Ethylene; Ethane and Acetylene.

The output variables was binary with 1 indicating accept the bushing and 0 indicating reject the bushing.

A. Paper degradation indicators

The cellulose in the paper breaks down over time and releases glucose, carbon monoxide, carbon dioxide, water, and acids. The glucose further breaks down into furans [23]. Heat, water and oxygen are the primary causes of paper degradation. Evidence of combustion is the accumulation of carbon on the paper surface due to tracking, see Figure 1. Tracking is the process that produces a partially conducting path of localized deterioration on the surface of an insulating material as a result of the action of electric discharges on or close to an insulation surface. Damaged paper does not regenerate it must be replaced by rewinding the bushing. Damaged oil can simply be drained and and replaced at a lower cost.
B. Oil degradation indicators

The quantity and types of gases reflect the nature and extent of the stressing mechanism in the bushing [11]. Oil breakdown is shown by the presence of Hydrogen (H$_2$), Methane (CH$_4$), Ethane (C$_2$H$_6$), Ethylene (C$_2$H$_4$), and Acetylene (C$_2$H$_2$). High levels of only hydrogen show that the degeneration mechanism is due to partial discharge at temperatures below 300°C. High levels of hydrogen and acetylene show that the degeneration mechanism is due to arcing at temperatures above 700°C. High levels of hydrogen methane, ethane and ethylene show that the degeneration mechanism is due to thermal faults at temperatures between 300°C and 700°C.

C. Other indicators

A high moisture content, i.e. more than 23ppm [22], would also warrant the removal of the bushing from service because it can initiate PD activity inside the bushing and result in water treeing and electrical tree formation and growth.

The ambient temperature together with the transformer loading will affect the normal operating temperature of the bushing. Elevated t-emperatures will increase the rate of degradation of the bushing insulation.

The pressure of the bushing may reflect the extent of ingress of air and moisture into the bushing if the bushing pressure is less than the atmospheric pressure even if no leaks are visibly detectable.

Bushings are a critical component in electricity transportation. They are used in substation buildings, transformers, locomotives, and switchgear. Bushings cause more than 15% of transformer failures in Eskom [5]. Australasian reliability statistics on 2096 transformers over 1970-1995 concluded that bushings were second to tap changers as the component initially involved in failure and were amongst the top three contributors to costly transformer failures [6],[10]. Sokolov [7] found that more than 46% of transformer defects were found to be attributable to bushings, onload tap changers, and the cooling system. Another study from the 63 members of the Edison Electric Institute [8] listed bushings amongst the three most common transformer failure modes reported.

The failure mechanisms result in a sudden and catastrophic failure [9] often followed by fire, tank rupture, violent explosion of the bushing propelling large broken pieces of porcelain several of metres at velocities enough to embed the material in concrete walls. Figure 2 shows that the result of collateral damage and risk of personnel injury is a major concern warranting the development of new and improvement of existing affordable, reliable, and flexible diagnostics technologies to permit asset owners to detect impending failures in a timely manner.

The cost of a 20 MVA transformer is approximately R2 million. The replacement cost of adjoining substation equipment may be several hundred thousand. The transportation costs to replace the transformer may be R500 000. Time to install and commission may be two months, which equates to a loss in revenue from unsold energy. If labour, testing, investigating and commissioning costs are added the total cost due to the bushing failure could easily reach R10 million if there is no loss of human life. Clearly, on-line monitoring equipment to monitor the integrity of bushings by means of unbalanced neutral current, leakage current, power factor, or dissolved gas content would add value and reduce risk on power transformers. Figure 3 shows a transformer that was damaged severely as a result of fire caused by an exploded bushing [22].

EPRI work done by Lindgren [12] used online monitoring DGA gasses trends to successfully diagnose transformer conditions using an instrument called a Serveron. Lau [13] also supports using DGA as an online diagnostics tool for bushings. Dominelli and Lua [14] were successfully able to interpret the fault condition that caused transformer failure by using DGA data and DGA criteria as presented in IEEE C57.104.
using an instrument called Vector. Collaborative research is currently being sought for sponsorship by EPRI to develop tools that make use of the bushing power factor for online monitoring [15].

This paper is arranged as follows. Section 2 presents the source of data. Section 3 describes the inputs to the ANN. Section 4 looks at the MLP. Section 5 analyses the data using the RBF. Section 6 looks at SVM. Section 7 compares the three ANN and makes concluding remarks.

III. BUSHING SAMPLE

A total of 1256 bushings were tested. The tests were done over a period of two and a half years from September 2001 to March 2004. Each bushing oil was checked using a DGA (dissolved gas analysis) test at intervals between six months and one year. For each bushing 14 input variables were considered. If each variable is one input \( x_i \) and the number bushings was 1256, then the number of data points is \( n = 17584 \). This represents a lot of time if a decision taken after manual consideration of data point. Application of ANN accelerates the decision process and allows the network to update criteria based on the decisions taken for other bushings. The input can change if different materials and voltages are used on the bushing. Also the experience of maintenance staff in the field can motivate for a change in the decision criteria. Their observations of non-quantifiable information such as “oil filling plugs were found to be rusty and the surfaces inside of the terminal caps were coated with white powdery material indicating signs of moisture ingress”, or “oil was dripping out of the capacitance tap continuously when the cap was removed” or “oil was black” or “oil had a burnt smell” or “passed the PD test”, “change in the Dissipation Factor measured before and after the AC withstand test” or combinations of these and other information can be used as an input into an ANN.

IV. MULTIPLE LAYERED PERCEPTRON

A. General Neural networks

The basis of all NN architecture is the neuron, which approximates the human brain neuron. Each neuron can have multiple inputs but only one output. The input to the neuron is defined as \( x \). The neuron is an activation function. The extent to which each neuron reacts to an input depends on the weights of that input. The activation function or neuron is equivalent to the transfer function in control systems. The output \( y \) is related to the input by the activation function. Neural networks learn when they are used to solve problems. They are not programmed. Learning is done by modifying weights, not by changing the number of neurons or the activation functions of each neuron. Training is the procedure by which the NN learns. Figure 3 shows the architecture for the MLP [21][27] that was applied. To prevent over training of the NN the number of neurons is limited to equal less than the
number of inputs.

\[ a_j = \sum_{i=1}^{d} w_{ij} \cdot x_i + b_j \]  
\( (1) \)

\[ a_j = \sum_{i=1}^{d} x_i + b_j \]  
\( (2) \)

\[ z_j = \tanh(a_j) \]  
\( (3) \)

\[ a_k = \sum_{j=1}^{m} w_{kj} \cdot z_j + b_k \]  
\( (4) \)

\[ y_k = \frac{1}{1 + \exp(-a_k)} \]  
\( (5) \)

Fig. 3 MLP architecture

B. Multiple Layered Perceptron

The MLP used for the classification problem has 2 layers of neurons, i.e. one hidden layer and one output layer. There were 7 neurons in the hidden layer, corresponding to one neuron per input. One outer neuron was used because only one number is required as an output. The input is a vector with 14 variables. Weights were used only between inner and output layer. The network uses back propagation with a momentum function to reduce the probability of the NN sticking in a local minimum. Bias terms are also used, even though none of the inputs is zero. The bias terms allow the inputs to influence the output even when the inputs are zero.

A logistic activation function for the output neuron \( y_k \) for a classification problem produces a more repeatable result than a softmax or linear, so the format of the other NN components is as follows:

\[ a_j = \sum_{i=1}^{d} w_{ij} \cdot x_i + b_j \]  
\( (1) \)

Figures 5 & 6 shows the errors in the simulation. An accuracy of 99% was achieved with 462 randomly selected bushings, while 99% accuracy was obtained for 67 bushings.

Methane had the highest relevance among the normalized input variables, indicating that most maintenance interventions on the bushings were due to high methane levels. Ethane was 2nd most important followed by nitrogen. Indicating that the bushings often need to be pressurized with nitrogen. Hydrogen and Ethylene were ranked 8th and 12th respectively in terms of importance in the accept/reject decision.

\[ r = \| x_i - x_m \| = \sqrt{x_i^2 - x_m^2} \]  
\( (6) \)
The activation function for the thin spline RBF is given by:
\[ \phi_j(x) = r^2 \log_e(r) \]  
(7)

The output in this example is a linear function, given by:
\[ y_k = \sum_{j=1}^{m} w_{kj} \cdot \phi_j(x) + w_{ko} \]  
(8)

\( w_{ko} \) is the weight when all the inputs are zero, which did not apply in this example.

As with MLP a solution is obtained by adjusting the weights until all the inputs give the target output using the same weights. The activation functions are not changed during training. The results of the RBF simulation are shown in Figure 8 & 9.

The SVM was developed specifically for classification problems. SVM are based on the structural risk minimization principle that was shown by Gunn [25] to be superior to empirical risk minimization principle used in conventional neural networks.

The SVM may include RBF or MLP as kernel functions nested within the activation function. Other kernel functions used are polynomial, Gaussian RBF, exponential RBF, Fourier series, splines, b-splines, or sum or product combinations of these. The theoretical criteria that are recommended for selecting a suitable function is the bootstrapping method, another is the cross validation method [3]. For this NN the author chose the bspline because it did not result in any errors. It has optimal properties when the underlying distribution of the data is not known (non-parametric data).

The SVM network has the form shown in Figure 10. If the plane separating data is not linear the SVM maps the input data onto a higher dimensional plane (feature space) where an optimal hyper plane can be defined. SVM work best if the number support vectors are equal to the inputs. In this example eleven support vectors were used with 14 inputs.

The SVM summarize the data into a limited number of support vectors. The support vectors will lie on the margins of the hyper plane if the data is linearly separable, then the number of support vectors needed is very small and the calculation requires less resources.

The SVM was able to classify the data with no errors in its training. The time period for learning with 462 data points was 14 minutes compared with 8 seconds for RBF and 45 seconds for MLP with ARD. The results of the SVM simulation are shown in Figure 11.

The classification inner activation function is given by:
\[ f(x) = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) \cdot k(x_i) \]  
(9)

Where \( k(x_i) \) is the kernel function and \( \alpha_i, \alpha_i^* \) are the Lagrange multipliers. The hyper plane (h) that optimally separates the data is derived by minimizing the Lagrangian \( \Phi \) with respect to the weights (w), bias (b) and \( \alpha \) [23],
\[ \Phi = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{l} \alpha_i [y_i (w \cdot x_i + b)] - 1 \]  
(10)
The multipliers are constrained in the range, $0 \leq \alpha_i, \alpha_i^* \leq C$. C is a misclassification tolerance, which can be chosen to represent the noise in the data. C can be a value from 1 to infinity [3]. The data points where the Lagrange multipliers are non-zero are called support vectors.

During the simulation 462 unique bushings were evaluated and a decision taken on whether to keep or replace them by training an ANN. A new set of bushings was then analyzed using the trained ANN with the following results: The SVM neural network processed the 462 bushing data in 14 minutes, with 88% accuracy. RBF crashed twice during the analysis and processed the 462 bushings data in 8 seconds with 60% accuracy. The MLP with ARD processed 462 bushing data in 45 seconds with 99% accuracy. SVM processed the 67 bushing data in 13 seconds with 100% accuracy. The RBF neural network processed the 67 bushing data in 1.5 seconds, with 95% accuracy. RBF crashed four times during the analysis, due to singularity in the determinant of the radial basis function. The MLP processed 67 bushing data in 14 seconds with 99% accuracy, SVM processed the 67 bushing data in 13 seconds with 100% accuracy. All the networks demonstrated a high level of accuracy in learning to discriminate between a rejected bushings and an acceptable bushing. The exercise was repeated 3 times for each network with the result that the SVM produced repeatable results. The work demonstrated that SVM processing time increases exponentially as the volume of data increases as a result SVM should not be used where time is a major constraint. MLP is most suitable for online applications. MLP is superior to RBF and SVM in terms of accuracy, repeatability, stability and speed.

**REFERENCES**


[16] S.Happe, H. Kranz, Practical application of machine intelligent partial discharge pulse suppressing system neurotekII.


