2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)

Transformer Health Index Monitoring using Supervised Prediction Model

Nazrul Fariq Makmor[1,*], Afzan Zamzamir[1], Ja'afar Adnan[1], Azharuddin Muktharuddin[1], Fakroul Ridzuan Hashim[1], Yulni Januar[2]

[1] Faculty of Engineering,
National Defence University of Malaysia,
Sg, Besi Camp, Kuala Lumpur, Malaysia

[2] Toshiba Transmission & Distribution,
System Asia Sd Bhd,
Petaling Jaya, Selangor, Malaysia

* Email: nazrulfariq@upnm.edu.my

***Abstract*— Dissolve gas analysis (DGA) is a method used to distinguish between transformers that are in optimal condition and those that require scheduled maintenance. The primary objective of DGA is to accurately identify issues caused by different gas forms in the transformer. The key gas method (KGM) analysis is a frequently employed approach in DGA. KGM is utilised to classify the health index of the transformer based on the development of gases within the transformer. Additionally, the prediction models used for health index classification include K-Nearest Neighbours (KNN), Discriminant Analysis, Principal Component Analysis (PCA), and Decision Tree as decision-making tools. The resulting outcome is subsequently compared to alternative prediction models to determine the ideal performance based on accuracy, precision and recall prediction. The results demonstrate that the KNN prediction model surpasses other models with an accuracy of 94.27%, precision of 94.12% and recall 92.44%.**

Keywords*— *Transformer; Dissolve gas analysis; Key gas method; MSE.

I. Introduction

Power transformers are the most important components in a power grid and make up a significant part of the total investment in a power delivery system. Due to their crucial significance, any breakdown in these transformers that causes power disruptions can result in significant financial losses for plant owners, as they are unable to generate energy. Hence, conducting regular inspections and doing periodic maintenance are crucial proactive actions to minimise such dangers [1-2]. Moreover, the failure of transformers can give rise to a range of dangerous circumstances, such as noise pollution, loss of power, and potentially even combustion, all of which can lead to the release of smoke from the transformers [3]. It is essential to ensure the optimal operation of a transformer to prevent such problems. Furthermore, power transformers are currently running at or very close to their maximum capacity due to the continuous rise in demand for electricity, hence heightening the probability of probable malfunctions or failures.

II. Literature Review

A. *Transformer Health Index*

The transformer health index is a quantitative measure that assesses the condition and efficiency of the transformer. Typically, the process involves examining many factors such as temperature, oil quality, insulation condition, winding resistance, and further diagnostic tests [4]. Through the examination of these features, researchers and experts may assess the general well-being of a transformer and determine how much longer it can be effectively used. This allows them to detect possible malfunctions and, consequently, make arrangements for maintenance or replacement procedures. The transformer Health Index (HI) is a comprehensive tool that combines data from operating observations, field inspections, and laboratory testing to generate accurate asset management decisions for transformers [5]. As stated by reference, this decision can be utilised to maintain transformers in optimal operational condition.

A considerable number of researchers have also focused on identifying the characteristics that have the most important influence on the transformer health index. The data source and diagnostics levels are responsible for replicating the primary factors contributing to uncertainty. The data sources that are accessible are affected by measurement uncertainties, which manifest as a variety of errors including measurement and quantization inaccuracies. These imperfections can be identified in the readily available data sources. Modelling uncertainties affect the diagnostic models, regardless of whether there is accurate knowledge or random uncertainty. This holds true regardless of the specific form of uncertainty that is there. Figure 1 depicts the framework of the transformer health index that was constructed and put into operation.

Several uncertainty sources investigated across transformer subsystems also affect the transformer health index. This improves final decision-making and makes it suitable for systems with measurement and process uncertainty. Additionally, it is suitable for unclear decision-making procedures. DGA is often used to test power transformers [6]. Gas concentrations dissolved in transformer insulating oil must be measured and studied. The transformer's internal defects and irregularities produce these gases. Device faults and anomalies are inside. The DGA is used to frequently diagnose transformers [7]. Previous experience has shown that the DGA works [8].

979-8-3503-8969-2/24/$31.00

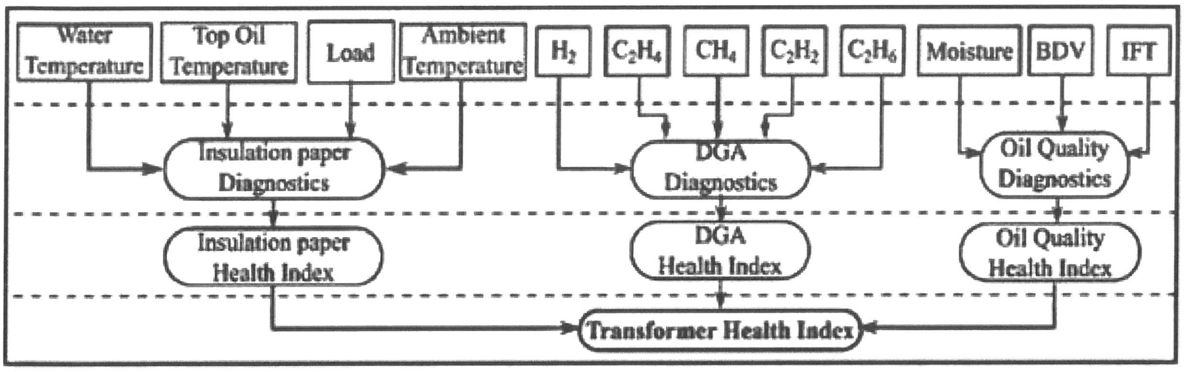

Fig. 1 Transformer health index soft computing framework [8]

B. K-Nearest Neighbours (KNN)

The K-Nearest Neighbours (KNN) algorithm has become common and basic in the field of machine learningIt is frequently used for activities associated with regression and classification. KNN is viewed as a non-parametric method that relies on specific instances. KNN assesses the likeness between a new data point and the current data points in a dataset for making predictions [9]. Euclidean distance is commonly used to measure similarity, but other metrics like Manhattan or Minkowski distances can also be applied. The value K represents the number of nearest neighbours considered. Choosing an appropriate value for K is hard as it has a substantial impact on the efficiency of the algorithm.

Raising the K value can lessen the model's sensitivity to local patterns, while lowering the K value can make the model more susceptible to noise. The average of the target values of the nearest neighbors is computed by the K-nearest neighbors (KNN) method for regression purposes, and for classification tasks, a majority vote is conducted. In the classification procedure, the newly found data point receives the most common label from its K closest neighbors. Regression involves calculating the goal value by taking the average of the values from the neighbouring data points.

C. Discriminant Analysis

Linear Discriminant Analysis (LDA) assumes a stable covariance matrix for every class, implying that only the class means differ, while the variability of features remains the same for all classes. Linear Discriminant Analysis (LDA) is a robust technique that effectively categorises several classes by employing linear decision boundaries [10]. However, Quadratic Discriminant Analysis (QDA) relaxes this criterion by allowing each class to have its own distinct covariance matrix. QDA can capture complex relationships among the properties of several categories. QDA surpasses LDA in versatility due to its ability to depict non-linear decision boundaries. However, to accurately forecast additional attributes, this technique requires a greater amount of computational power and data [11].

The decision to use QDA or LDA depends on the dataset's characteristics and the assumptions made about the data. QDA is recommended when there are noticeable changes in variance between classes and linear decision boundaries are inadequate. However, LDA is more efficient in terms of processing and produces better outcomes when the assumption of equal variance holds true. Both Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are frequently employed in supervised classification tasks, which aim to predict the category or class of a data point by considering its attributes. Additionally, KNN, a straightforward and widely used machine learning technique, is frequently applied in regression and classification tasks.

D. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a commonly used technique in data analysis, machine learning, and statistics. By transforming the original attributes into a new set of variables called principal components, it aids in reducing the dimension's number in a dataset. The fundamental purpose of PCA is to reduce the number of dimensions while preserving the most crucial information. PCA employs mean cantering and scaling to normalise the data, ensuring that all features are measured on a consistent scale. Due to the process of standardisation, PCA can compute the covariance matrix of the standardised data, providing a thorough understanding of the relationships between different variables. The covariance matrix displays the degree of correlation or independence between pairs of features [12].

The principal components are represented by the eigenvectors, while the eigenvalues represent the quantity of variance associated with each principal component. The arrangement of the eigenvectors and eigenvalues is in descending order, defining the principal components. Researchers have the option to choose a particular quantity of principal components according to the amount of variance they want to keep. These selected principal components are then used to project the original data, resulting in a reduced-dimensional representation of the data.

E. Tree Decision

Machine learning often use the decision tree classification method to categorise data points into separate groups. This supervised learning strategy is capable of successfully completing both binary classification (with two classes) and multi-class classification (with more than two classes). The dataset employed in this method contains labelled instances, with each data point assigned a distinct class label. This method creates a hierarchical structure that resembles a tree. The technique partitions the data into different subsets at each node of the tree by choosing the attribute from the dataset that provides the most valuable information. The objective is to choose the property that minimises ambiguity regarding the class labels. Different measurements like Gini impurity, information gain, and entropy are employed to evaluate how efficient a division is [13].

A decision tree is constructed using a recursive approach. The information is divided at every node until a certain stopping condition is reached. This criterion could be based on factors such as the tree's depth, the number of data points at a node, or other relevant variables. The node is given a specific class label and turns into a leaf node whenever the stopping condition is met. The label at the leaf node is typically determined by the majority class of the data points within that node. Starting at the root node of the decision tree, a fresh and unseen data point is classified. The decision criteria at each node are compared with the feature values of the data point to determine the next best course of action. The data point moves through the tree until it arrives at a terminal node, where a prediction is made based on the class label associated with that leaf node.

III. METHODOLOGY

A. Data Collection

The research methodology utilised is referred to as the Integration of the KGM into a system for assessing transformer data obtained from DGA. The conditions specified in Table I have functioned as the guiding framework for arranging this data. The conditions are calculated by conducting the TDCG calculation, which entails summing the gas concentrations shown in Table II. Subsequently, the TDCG employs the data itself to determine the specific criteria for structuring the data. After completing this stage, the values obtained from the TDCG are used in the data sorting procedure.

B. Pre Processing Dataset

The TDCG limit concentration identified in Table II is used as a benchmark to categorise the obtained data into four distinct categories. The term "Condition 1" denotes a transformer that complies with all specifications and is free from any defects. Condition 2 denotes a state that is deemed acceptable despite the existence of small flaws, and may still be effectively employed, but with a few of these concerns present. The item is in a deteriorated condition, and issues requiring repair or maintenance have been identified that must be addressed within the next six months. This situation is indicated by the term "Condition 3." Lastly, "Condition 4" pertains to a pressing situation where significant flaws must be promptly rectified within the next three months or at the earliest opportunity.

TABLE I. GASSES INVOLVED IN KGM

Type of Gas	*Formula*
Hydrogen	H_2
Methane	CH_4
Acetylene	C_2H_2
Ethylene	C_2H_4
Ethane	C_2H_6
Carbon Monoxide	CO
Carbon Dioxide	CO_2

TABLE II. LIMIT ASSIGNED AS INPUT

TDCG Limit	*Condition*
720	1
721-1920	2
1921-4630	3
>4630	4

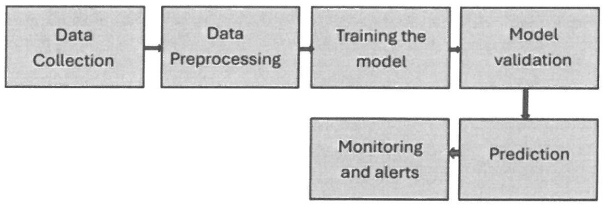

Fig. 2 Process flow diagram

To monitor the transformer health index using supervised prediction methods, data from Tenaga Nasional Berhad (TNB) will be used as input. This dataset includes various parameters like operational metrics, condition monitoring indicators, and historical performance records. The K-Nearest Neighbors (KNN) algorithm, a simple yet effective instance-based method, will classify new data points by identifying the 'K' nearest neighbors in the training data using a distance metric such as Euclidean distance. The algorithm then assigns the new point to the most common class among these neighbors, leveraging proximity for prediction. Alongside KNN, other supervised prediction models will be employed and simulated to ensure the accuracy and reliability of predictions.

Following simulation and validation, the performance of each model will be assessed based on accuracy and other relevant metrics. Five different models, including KNN, Support Vector Machines (SVM), Random Forest, Decision Tree, PCA, LDA and QDA will be compared to determine the most effective one. The best-performing model will be selected for deployment, integrating it into a real-time monitoring system to continuously evaluate the transformer health index. This model will provide ongoing assessments and generate alerts for potential issues, facilitating timely maintenance and enhancing transformer reliability.

IV. RESULT AND DISCUSSION

The implementation of the DGA is approaching, and the objective of this study is to determine the most efficient approach to achieve it. The aim of this study is to evaluate and compare the effectiveness of different training algorithms based on the morphology of the parameters being considered. To assess the efficiency of the network, measurements such as accuracy, precision and recall. Furthermore, this project will explore other approaches of diagnose DGA. The anticipated output values are presented in Table III as a percentage of the accuracy, precision and recall of various classifiers, including KNN, Decision Tree, PCA, LDA and QDA [14-15]. The goal of the study is to identify the best prediction model for classifying cardiac anomalies based on input of parameters by comparing several models, including KNN, Decision Tree, PCA, LDA, and QDA. Table III presents the results of the comparison, providing a comprehensive analysis of each model's performance in accuracy terms, precision and recall. This allows readers to compare and evaluate the relative efficacy of the models.

TABLE III. PERFORMANCE OF SEVERAL PREDICTION MODEL.

Prediction Model	*Accuracy, %*	*Precision, %*	*Recall, %*
KNN	93.28	94.12	92.44
Decision Tree	92.44	93.71	91.17
PCA	91.32	91.89	90.75
LDA	90.24	90.42	90.06
QDA	87.42	88.12	86.72

With a remarkable accuracy rate of 93.28%, Table III demonstrates that the KNN model outperformed all other models in terms of prediction accuracy. Furthermore, the precision and recall value of the KNN model showed with 94.12% and 92.44%, respectively. This implies that there were not many significant prediction errors. The decision tree classifier, on the other hand, came in second among the models with an accuracy rating of 92.44%. On top of that, the decision tree classifier's precision and recall performance are 93.71% and 91.17%, respectively. The results show KNN model performed better than the decision tree classifier, and it ranked second behind KNN model.

PCA achieves a commendable accuracy, precision and recall performance of 91.32%, 91.89% and 90.75%, respectively. The results placed the PCA model as the third best performing method. The categorization of transformer health index using LDA achieved an accuracy of 90.24%, precision of 90.42% and 90.06%. The LDA can surpass the QDA in terms of performance but, the model is unable to outperform others prediction models. The QDA prediction model demonstrates the capacity to accurately classify transformer health index with an accuracy rate of 87.42%. The performance followed with precision and recall measurement with 88.12% and 86.72%, respectively. Nevertheless, the outcomes fail to surpass existing prediction models in classifying the transformer health index.

V. CONCLUSION

The objective of evaluating classifiers for its accuracy, precision and recall performance in predicting transformer health index to several class and categories. All the classifier or prediction models capable to perform with high accuracy, precision and recall performances. However, the KNN prediction model outperforms others prediction model with 93.38% on the accuracy performance, 94.12% on the precision performance and 92.44% on the recall performance. On top of that, QDA prediction model with the lowest with 87.428% on the accuracy performance, 88.12% on the precision performance and 86.72% on the recall performance. Results shown by KNN model and QDA model are not significant difference and conclusion can be made that all classifier or prediction model capable to provide high classification performance. Nevertheless, using others prediction model such as multilayer perceptron (MLP), radial basis function (RBF) and/or Support Vector Machine (SVM) can be performed to increase the performance. More dataset can also be provided in order to increase the prediction performance dan generalization of the classifiers.

REFERENCES

[1] M. V. A. J. Wang, A. J. Vandermaar, and K. D. Srivastava, "Review of condition assessment of power transformers in service," IEEE Electrical Insulation Magazine, vol. 18, no. 6, pp. 12-25, Nov./Dec. 2002.

[2] A. E. Abu-Elanien and M. M. A. Salama, "Asset management techniques for transformers," Electric Power Systems Research, vol. 80, no. 4, pp. 456-464, Apr. 2010.

[3] A. Christina, M. A. Salam, Q. M. Rahman, F. Wen, S. P. Ang, and W. Voon, "Causes of transformer failures and diagnostic methods–A review," Renewable and Sustainable Energy Reviews, vol. 82, pp. 1442-1456, Feb. 2018.

[4] G. C. Jaiswal, M. S. Ballal, P. A. Venikar, D. R. Tutakne, and H. M. Suryawanshi, "Genetic algorithm–based health index determination of distribution transformer," International Transactions on Electrical Energy Systems, vol. 28, no. 5, e2529, May 2018.

[5] A. Jahromi, R. Piercy, S. Cress, J. Service, and W. Fan, "An approach to power transformer asset management using health index," IEEE Electrical Insulation Magazine, vol. 25, no. 2, pp. 20-34, Mar./Apr. 2009.

[6] S. S. Ghoneim and I. B. Taha, "A new approach of DGA interpretation technique for transformer fault diagnosis," International Journal of Electrical Power & Energy Systems, vol. 81, pp. 265-274, Oct. 2016..

[7] K. D. Patekar and B. Chaudhry, "DGA analysis of transformer using Artificial neural network to improve reliability in Power Transformers," in Proc. 2019 IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Nov. 2019, pp. 1-5.

[8] J. I. Aizpurua, B. G. Stewart, S. D. McArthur, B. Lambert, and J. G. Cross, "Towards a comprehensive DGA health index," in Proc. 2018 IEEE 2nd International Conference on Dielectrics (ICD), July 2018, pp. 1-4.

[9] L. Sun, Y. Liu, B. Zhang, Y. Shang, H. Yuan, and Z. Ma, "An integrated decision-making model for transformer condition assessment using game theory and modified evidence combination extended by D numbers," Energies, vol. 9, no. 9, p. 697, Sep. 2016.

[10] [10] N. A. Salim, J. Jasni, H. Mohamad, and Z. M. Yasin, "Transformer health index prediction using feedforward neural network according to scoring and ranking method," International Journal of Advanced Technology and Engineering Exploration, vol. 8, no. 75, p. 292, Jun. 2021.

[11] A. K. Al Mhdawi and H. S. Al-Raweshidy, "A smart optimization of fault diagnosis in electrical grid using distributed software-defined IoT system," IEEE Systems Journal, vol. 14, no. 2, pp. 2780-2790, Jun. 2019.

[12] A. H. F. S. A. Jamil, J. A. Kadir, J. M. Jamil, F. R. Hashim, S. Shaharuddin, and N. F. Makmor, "Multilayer Perceptron Optimization of ECG Peaks for Cardiac Abnormality Detection," in Proc. 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), Oct. 2022, pp. 37-41.

[13] S. Tangirala, "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm," International Journal of Advanced Computer Science and Applications, vol. 11, no. 2, pp. 612-619, Feb. 2020.

[14] F. R. Hashim, S. H. F. S. A. Jamil, J. A. Kadir, N. S. Hasan, B. Mustapha, and Y. Januar, "Tansig Based MLP Network Cardiac Abnormality," in Proc. 2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Nov. 2019, pp. 199-203.