SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 7, No. 2, 2024, 302-313
DOI: 10.35377/saucis...1414115

RESEARCH ARTICLE

Predictive Model for Incipient Faults in Oil-Filled Transformers

Efosa Igodan[1], Michael Osajeh[1], Linda Usiosefe[1]

[1] Unviersity of Benin, Benin City, Nigeria

Corresponding author:
Efosa Igodan, University of Benin, Benin City, Nigeria
charles.igodan@uniben.edu

Article History:
Received: 03.01.2024
Accepted: 01.07.2024
Published Online: 29.08.2024

ABSTRACT
The power transformer is an invaluable piece of device in the power system. To prevent catastrophic failures and the ensuing power outages, the status of a transformer linked to a system must be examined for any possible faults. Despite using DGA as a global tool for detecting faults, it is limited by the inability to accurately solve the problem associated with results variability due to the intrinsic nature of the IEC TC 10 database. This study proposed a data-driven fault/defect diagnostic model using four ensemble models with three base classifiers respectively. The base classifiers are comprised of SVM, C4.5 decision tree, and naive Bayes while the ensemble methods are comprised of stacking, voting, boosting and bagging respectively. The DGA dataset used comprises seven features and 168 instances split into training (i.e. 56%) and test (i.e. 44%) datasets respectively. The results indicate that C4.5 obtained a 98.33% accuracy while stacking obtained a 99.89% accuracy as the best-performing base and ensemble models respectively. The high classification performance accuracy achieved by our proposed models indicates its capacity for real-world applications. It can be applied to advance automation in mobile-based technology.

Keywords: Dissolve gas analysis, power transformer, Naive Bayes, support vector machine, C4.5, ensemble learning

1. Introduction

A power transformer is one of the costliest, intricate, and significant pieces of equipment in electrical power systems that is often susceptible to early failures when overworked. Its malfunctioning has the potential to cause considerable damage that can lead to economic and social loss. Therefore, early detection of this incipient fault and its precise identification are of immense importance towards averting any damage that may arise. The transformer's organic insulating materials and its oil, break down and release various gases due to the electrical and thermal stress when in operation. Some of these gases released are hydrogen H_2, acetylene C_2H_2, methane CH_4, ethylene C_2H_4, and ethane C_2H_6. While carbon dioxide CO_2 and carbon monoxide CO are formed due to the decomposition of the insulating paper [1-3] as part of the fault gases, nitrogen N_2 and Oxygen O_2 are the non-fault gases [4]. The primary fault categories that are reliably identifiable are partial discharges (PD), low-energy discharges (D1), high-energy discharges (D2), brownish paper-coloured thermal faults (T1), carbonized paper-coloured thermal faults (T2), and thermal faults above 700 ºC (T3) indicated by metal colouration, fusion, or oil carbonization. In recent times, fault-type prediction methods are either based on dissolved gas analysis (DGA) carried out by some conventional methods or machine learning-based (ML) methods and Artificial Intelligence (AI) or a combination of both methods [26]. The conventional methods are divided into ratio methods and graphical methods. The ratio methods include the Doernenburg ratio, Rogers' three and four-ratio [9-11,56], the gas production rate method [7,8], and the IEC 60599 code methods [26] among others. The graphical methods include the Duval triangle [5,9,12], Duval and Mansour pentagon [4,26], and Gouda heptagon [26] methods. These conventional methods are simple to implement. However, they have poor detection accuracy rates for power transformer fault types. Other limitations include their inability to deal with all data value ranges [59]. Also, because some ratio ranges lie outside the methods' parameters [14,15], it becomes difficult to classify the transformer's state correctly, causing variability in fault identification accuracy [17]. Furthermore, insufficient coding and strict coding limits constrain the three-ratio and enhanced three-ratio approaches [5]. Nevertheless, compared to traditional ratio and graphical methods, artificial intelligence-based methods have higher prediction accuracy for transformer problem diagnosis, which is required to guarantee the greatest degree of electrical grid reliability. Consequently, it was suggested that machine learning-based techniques, which are based on DGA, are better for power transformer diagnostics [26]. However, despite the progress in using these learning methods, they still have several limitations. Some of these limitations are model overfitting and underfitting, results variability issues, and the problem with bias-variance trade-off. While some are trapped in a local optimum problem [26], others may have difficulty in parameters tuning [59] resulting in the inefficiency of the use of single models to achieving optimality. The utilization of AI and ML-based methods include: naive Bayes, decision tree-based models [4], artificial neural network (ANN) [18], expert system [19], fuzzy theory [6,14,20], hybrid grey wolf optimization technique [3], grey system [5], support vector machine (SVM) [22], K-Nearest neighbour [23], Bayesian Neural Network [24], and other intelligent systems that infuses diversity in the models as reported in [59]. Therefore, to effectively address some of these challenges, an ensemble method is essential. It does this by utilizing the diversity of the predictions, lowering the risk of overfitting and underfitting, balancing the trade-off between bias and variance, and employing various subsets and features of the data to improve performance and robustness. However, there appears to be insufficient literature that has applied ensemble models of predict faults in power transformers to the best of the researcher's knowledge. Although, more research studies have been carried out in this area, however, there is still no one-fit-all model that can solve all problems [25]. While many diagnostic models underperform because they rely too heavily on the expert's knowledge, in other situations it can be difficult to find a suitable relationship between the input and output variables to support learning [27, 58]. This study proposes developing a power transformer defect prediction system based on six machine learning classification algorithms, motivated by the abundance of problems. Multilayer perceptron (MLP), Support Vector Machines (SVM), Naïve Bayes (NB), C4.5 decision trees, Logistic Regression (LR), and ensemble methods are the six machine learning algorithms. 168 dataset samples gathered from the literature were used to test the approaches. The collected samples have a high overlap degree among different fault types. To enhance the ML predicting accuracy, data preprocessing techniques were implemented. The prediction accuracy was compared among different existing classification methods from the literature. The various methods are implemented using Python Programming Language software on the Google Collab environment.

The organization of the next sections is as follows. Section two carries out the literature review, while section three introduces materials and methods. Section four shows the performance evaluation of the proposed method, while section five discusses the findings of the results and comparisons with existing results from the literature. Finally, section six presents conclusions and recommendations.

2. Related Literature Review

Recent developments in field-deplorable computer technology have reignited curiosity in using computers for routine tasks, particularly labour-intensive ones. Various researchers have conducted numerous studies to investigate cutting-edge methods for automated faultfinding on the electrical grid. The authors in [4] observed that DGA remains one of the best methods for transformer fault identification, however, it has the problem of results variability. The authors suggested using Naïve Bayes and decision trees to develop a fault identification system using 481 instances with nine distinct input vectors. While the decision tree obtained 83.75%, the naïve Bayes achieved 86.25% respectively. However, their study was unable to perform optimally. In [30], Fuzzy logic (FL) was used to identify and safeguard power system transformer problems. Significant maintenance and repair cost savings were achieved as a result of the investigation, however, the issue of variability of results was observed in the work and the inability of the model to learn. In [16], a data-driven method utilizing SVM, back-propagation neural networks (BPNN), and extreme learning machine-radial basis function (ELM-RBF) was proposed for a fault diagnosis system based on DGA data. In comparison to ELM-RBF and BPNN, SVM demonstrated the best performance in the proposed multistage fault diagnosis system. However, it was characterized by increased computation complexity for all stages. In [31], the authors presented a genetic algorithm-based model to choose the best-dissolved gas ratios for SVM-based power transformer problem diagnostics. The model was based only on the International Electro-Technical Commission (IEC) TC 10 database. The study only used three conventional techniques: IEC criteria, DGA data and IEC three key gas ratios with SVM, and back propagation neural network (BPNN) respectively with an 87.18% accuracy achieved. The works of [32], a probabilistic neural network (PNN) and bio-inspired optimizer, were applied to artificial intelligence to build a fault diagnostic model. The PNN's hidden layer smoothing factor was optimized by the bio-inspired optimizer known as the improved salp swarm algorithm (ISSA), which gradually enhances the PNN's classification performance. The PNN served as the fundamental classifier in the fault diagnostic model and achieved 99.65% accuracy higher compared to the traditional fault diagnosis techniques indicating that the technique has a powerful learning ability for data with high complexity. The authors in [34] proposed a neural network model based on traditional methods IEC and Roger's ratio for power transformer diagnosis. Most of the constraints were eliminated in their work, and the diagnostic outcomes increased for the IEC and Roger procedures, from 20% to 70% and 40% to 70% respectively. However, they observed that in certain samples, all approaches i.e. whether the conventional or the model-based on artificial neural networks were deceptive, providing incorrect diagnoses that put the power transformer's integrity at risk. The authors in [35] presented a fault diagnostic model for power transformers using machine learning algorithms and traditional methods. The results showed that the decision tree algorithm outperformed KNN and SVM with a 93.13% accuracy. The authors observed that the classification algorithm and the input data greatly affect the diagnostic accuracy. To determine the risk of transformer fault types, a unique DGA technique based on the Parzen window (PW) estimation was created in [36] using the quantities of five combustible hydrocarbon gases: hydrogen, ethylene, acetylene, methane, and ethane.

According to the results when compared, the PW method outperforms several AIs and ML techniques, including ANN, SVM, ELM, SaE-ELM, and NNCA, and performs noticeably better than traditional ratio-based diagnostic procedures. According to the experimental data, 94.82% of these challenging circumstances are correctly classified by the suggested model, with Duval's triangle providing an unclear classification. The drawbacks of the existing transformer fault diagnosis techniques in dissolved gas-in-oil analysis were addressed in [32] by proposing a transformer fault diagnostic model based on the three

DGA Ratios and PSO-SVM. With an 85.71% accuracy attained, the results showed that the suggested PSO-SVM strategy outperformed the SVM and GA-SVM approaches. However, their results were reported as suboptimal. In [59], a performance Assessment of the IEEE/IEC Method and Duval Triangle (DT) technique for Transformer Incipient Fault Diagnosis was proposed. While the DT methodology was found to perform better than the IEEE/IEC method, the scope of their investigation was confined to the consideration of only two DGA methods and two types of electrical faults: high energy discharge and low energy discharge. Furthermore, the authors in [5] suggested that intelligent algorithms should be combined for mutual complementation to form a hybrid fault diagnosis network to prevent local optimum problems. Their work offers insightful ideas and recommendations for studying intricate power systems, along with references and directions to help researchers select the best course of action for achieving DGA-based fault detection and selecting huge oil-immersed power transformers for electrical testing intended to be preventive. As a result of this strategy, some academics recommended tackling this problem with ensemble approaches. The authors in [33] observed that all the technologies applied in power system fault diagnosis are closely related to modern information technologies such as information system theory, machine learning integration, and information entropy. The authors posited that promoting practical research on modern information technologies is conducive to the high-quality prosperity of power system fault diagnosis, which would play an imperative role in promoting the whole intelligent process of modern society. Motivated by these limitations, three well-known supervised machine learning techniques - C4.5, SVM and naive Bayes, and four ML ensemble methods - Bagging, Boosting, Stacking and Voting methods were investigated to address some of the challenges associated with power transformer fault types towards building results confidence. This includes combining several intelligent algorithms to create a hybrid classifier that mutually complements each other to handle complex issues or combining them to form an ensemble (homogeneously or heterogeneously). Furthermore, the authors in [17] were motivated by the low accuracy of fault identification characterized by the usage of traditional transformer fault diagnostic methods. They suggested combining the XGBoost with an enhanced genetic algorithm (IA) to create a hybrid diagnostic network to locate power transformer problems. The transformer failure recognition problem was broken down and reconstructed into multiple smaller issues that the model could address by combining IGA and XGBoost. An accuracy of 99.2% higher than IEC ratios, dual triangles, SVM, and CVA was obtained. However, the model lacks generalization ability. The authors in [27] observed that the DGA interpretation highly depends on the technical personnel's competence, but it's not conclusive in determining the presence of incipient defects. The authors proposed a brand-new, decision tree-based multinomial classification model called KosaNet. According to the study, KosaNet outperformed the decision tree, k-NN, random forest, naive Bayes, and gradient boost, especially when it comes to classifying multinomial data, achieving an accuracy of 99.98%. The authors also intend to investigate the application of KosaNet for time series data in the context of deploying real-time IoT-enabled smart sensors for transformer observation. In contrast, six optimized machine learning classification algorithms (DT, DA, NB, SVM, KNN, and ensemble approaches) with four data transformation strategies were used in [38] to build a novel power transformer defect type diagnostic. An accuracy of 97.14% was achieved using the ensemble learning method.

3. Materials and Methods

This part contains the basic building block of the study as well as the overall procedure of our proposed model. The entire architecture of the suggested research work technique is shown in Figure 1.

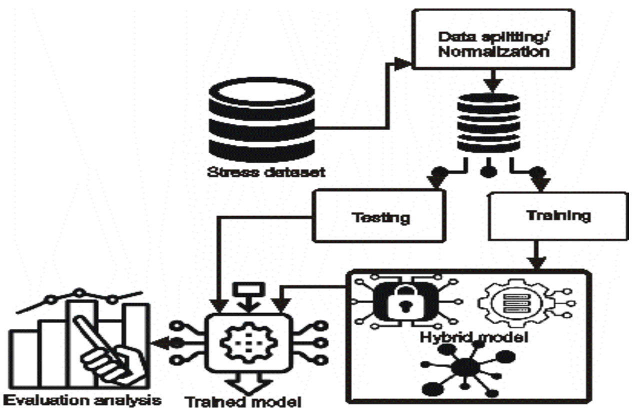

Figure 1. Methodology of the Proposed System Architecture

3.1. Transformer Fault Diagnosis System

Analysis of the dissolved gas in insulation oil sheds light on the thermal and electrical stressors that the oil-immersed power transformer experiences, which cause the oil and paper insulator to break down in the transformer. These stresses release gases as they break down the insulating materials. Whilst the paper insulator produces CO and CO_2, the oil decomposition releases H_2, CH4, C_2H_2, C_2H_4 and C_2H_6 [16]. The fault type can be determined by type and amount [1]. The incipient fault types used in this proposed framework are based on the new IEC publication 60599 in the IEC TC 10 database, simplified into five categories as depicted in Figure 2 as our proposed methodology for this research.

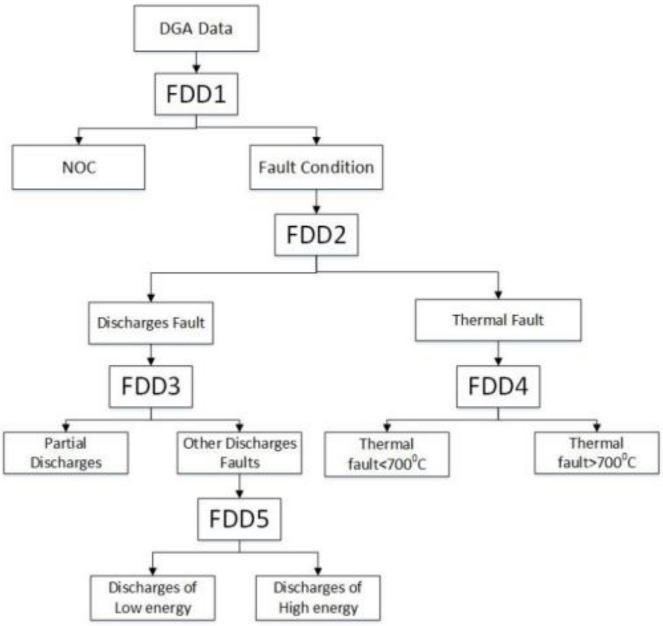

Figure 2. Data-Driven Fault Diagnosis System [16]

3.2. Data Description

The proposed fault diagnosis framework was implemented using 168 DGA data, consisting of 50 normal operating conditions (NOC) and 118 fault data with seven attributes originating from IEC TC 10 databases. The fault type distribution is shown in Table 1. A ratio of 80:20 was used to split the data into training and testing respectively.

Table 1. Distribution of DGA Instances

Fault type	PD	D1	D2	T1&T2	T3	Normal	Total
Train	6	15	28	10	10	25	94
Test	5	10	20	6	8	25	74

3.2.1. Data preprocessing: This is a crucial step in preparing data for training, ensuring it is in the right format and quality. It involves various procedures, such as data cleansing, balancing, imputing, normalizing, encoding, augmenting, and bias mitigation, to make the data suitable for further analysis and modelling.

3.2.2. Data normalization: Normalizing data prevents bias, improves algorithm convergence and speed and stabilizes variance. The normalization enhances model performance, interpretability, and the reliability of statistical analyses, by bringing all features to a common scale [33,39,40]. Min-Max normalization was used to standardize and transform the dataset using Equation 1.

$$f(x) = \frac{x_i - X_{min}}{(X_{max} - X_{min})} \qquad (1)$$

where $f(X')$ is the normalized value, x_i the original values are X_{min} and X_{max} i.e. min and max values of all original values.

3.3. Classification algorithms

In the following subsection, the classifiers and their ensembles used to evaluate our proposal are briefly discussed. Five well-known classifiers of different families were used. Creating a composite global model with accurate and diverse estimates is the primary objective of the ensemble methodology over a single model. The original problem is solved by every model in the ensemble collection as it lowers the generalization error [39,40].

3.3.1. Multilayer Perceptron (MLP): One well-known "black box" neural network model is the MLP, which adjusts propagated error to reach an arbitrary level of accuracy through the use of a back-propagation algorithm [25,41]. The input vector x_i of the MLP is multiplied by a weight vector w_i, and summed with the bias b, to produce an output $\hat{y}$ using the following Equations 2-5:

$$y_i = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \qquad (2)$$

where n stands for an input-output pairs, f stands for an activation function shown as:

$$f = \frac{1}{1+exp^{-x_i}} \quad (3)$$

$$E(\hat{y}, y) = \frac{1}{2}\sum_{i=1}^{n}(\hat{y} - y)^2 \quad (4)$$

where E is the error function.

$$\delta_i = \frac{dE}{w} \quad (5)$$

Where δ_i and w are the gradient descent and weight respectively. One hidden layer is used.

3.3.2. Support Vector Machines (SVMs): In a binary classification or multi-class scenario, the Support Vector Machine (SVM) finds the hyperplane that can maximize the margin between distinct classes using a one-to-one or one-to-many technique [42]. The hyperplane is described in Equation 6. Equations 7 and 8 depict the dual issue of the objective using the Lagrange multiplier approach.

$$\hat{y}(x) = w^T.x + b \quad (6)$$
$$\max_{a} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j \quad (7)$$
$$s.t. \sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, n \quad (8)$$
$$w = \sum_{i=1}^{N_{train}} \alpha_i y_i x_i \quad (9)$$

Once α_i is calculated, w can be obtained using Equation 9. The radial basis function is used as the kernel function and the regularization parameters as C and σ which are set to 100 and 10 respectively.

3.3.3. The Bayes Theorem: Is a probabilistic ML model based on the foundation of the Naive Bayes classification algorithm. The algorithm uses probability theory to predict the class of an input instance by calculating the conditional probabilities of each feature given its class [39,40].

$$P(y|X) = \frac{P(y)P(X|y=C_l)}{P(X)} = \frac{P(y)\prod_{i=1}^{n} P(x_i|y=C_l)}{P(X)} \quad (10)$$

Where, X is given data instance which is represented by its feature vectors $(x_1, x_2, \dots, x_n)$, y is a class target (normal or fault types). $P(y)$ and $P(X|y)$ are the prior and conditional probability of the outcome, while $P(X)$ represent the probability of the predictor values.

3.3.4. C4.5 Decision trees: To generate decision tree classifiers, the C4.5 methods were applied. Based on the idea of information gain, the C4.5 algorithm builds decision trees where the decisions made in each classification are connected to the target classification [4,43,44].

$$\mathrm{InfoGain(T, X) = Entropy(T) - Entropy(T, X)} \quad (11)$$

$Entropy$ is the amount of uncertainty in the randomness of elements, and it is used to measure impurity.

$$Entropy(T, X) = -\sum_{i=1}^{c=7} p_i log_2 (P_i) \quad (12)$$

3.3.5. Logistic Regression: Is a statistical technique for assessing the assumed relationships between the independent factors x and the dependent variable y. Our decision on logistic regression is that it is recommended for meta-level learning, and is used to combine the base learners [45].

$$y = \left(\frac{p(x)}{1-p(x)}\right) = \frac{e^{(\beta_0+\beta_1 X)}}{1+e^{(\beta_0+\beta_1 X)}} \quad (13)$$

Where β_0 is the bias intercept term, β_1 is the coefficient for an input value, x is the input values, and y is the predicted output.

3.4. Ensemble Methods

Ensemble methods' [40] main goal is to create a composite global model with more accurate and reliable decision estimates than the single base models. The idea is to combine the results obtained from the multiple classifiers to increase accuracy and reduce generalization errors.

3.4.1. Adaboost: Adaboost uses the iterative ensemble approach to merge weak classifiers into a powerful strong classifier. The basic idea behind Adaboost is to use Equation 14 to illustrate how to build classifier weights and use the average majority vote [46] to train data samples to predict a class target of a given data instance with two classes.

$$\sum_{t=1}^{T} w_t d_{t,j}(x) = max_{j=1}^{C} \sum_{t=1}^{T} w_t d_{t,j} (x) \quad (14)$$

where $d_{t,j}\ (x)$ represents support given by the t^{th} classifier to the j^{th} class for the instance x, w_t is the weight of classifier t and T is the total number of classifiers.

3.4.2. Bagging: Bagging, a technique developed from bootstrap aggregation, is the most straightforward yet efficient independent ensemble method for improving the accuracy of unstable learning algorithms. The datasets are split up among many bootstrap replicates during bagging. Every replication is made from the original dataset, which comprises, on average, 63.2% of the original data. The sluggish learner must go through multiple bootstraps repeatedly as part of the process. With every iteration, the weak learner's classifier is fused into a strong composite classifier, yielding better accuracy than any single component classifier could achieve [60]. The plurality voting method sometimes referred to as the majority voting system was used in this study and shown in Equation 15. This system is then utilized to calculate the total of all base learners.

$$\sum_{t=1}^{T} d_{i,j} = max_{j=1}^{C} \sum_{t=1}^{T} d_{t,j}(x) \qquad (15)$$

where the decision of the t^{th} classifier is defined as $d_{t,j} \in \{0,1\}$, $t = 1, \ldots, T$ *and* $j = 1, \ldots, C$. T represents the size of the classifiers, and C represents the size of the classes. If t^{th} chooses ω_j, then $d_{t,j} = 1$, otherwise *0*.

3.4.3. Stacking: Stacking is an ensemble strategy to integrate heterogeneous models using a meta-classifier. SVM, C4.5, NB, and MLP are the five basis classifiers trained to extract the final outputs for predicting outcomes from the base classifier. Next, to avoid the overfitting issue brought on by the ensemble's base models, logistic regression is used as the meta-classifier [43].

3.4.4. Voting: Voting is decisions ensemble method in machine learning where multiple independent models are trained heterogeneously or homogeneously on the same dataset and their predictions are used to make final decisions by choosing the most frequent prediction. The combining of the predictions of the classifiers can proceed in multiple ways either using majority voting or weighted voting [48,49]. This research study adopted the majority voting as in Equation 16.

$$class(x) = \underset{c_i \in\ dom(y)}{\arg\max} \sum_k g(y_k(x), c_i) \qquad (16)$$

where $y_k(x)$is the classification of the k^{th} classifier, and $g(y_k(x))$ is an indicator function defined as:

$$g(y_k(x)) = \begin{cases} 1 & y = c \\ 0 & y \neq c \end{cases} \qquad (17)$$

4. Performance Evaluation Methods

To evaluate the behavior of models concerning the applicability and performance, several evaluation measures need to be defined. Measures including classification accuracy, F-Measure, precision, and recall are frequently employed to highlight reducibility power of the classification models [50–52]. The classification results, which are frequently recorded in a matrix format called a Confusion Matrix summarizes the outcome of the algorithm, and is used to determine these metrics [53] with the following four outcomes as illustrated in Table 2 [54], and represented in equations 18 through 22.

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \qquad (18)$$

$$Precision = \frac{TP}{TP+FP} \qquad (19)$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad (20)$$

$$Specificity = \frac{TP}{TP+FN} \qquad (21)$$

$$F1 - score = 2\frac{(P*Sn)}{(P+Sn)} \qquad (22)$$

Table 2. Confusion Matrix

Confusion matrix		***Classifier***	
		Negative	***Positive***
Actual class	*Negative*	*TP*	*FP*
	Positive	*FN*	*TP*

5. Results and Discussions

This section discusses the results of the proposed model about the standard metrics applied for evaluation. The default parameters of the models were applied in this study.

5.1. Results

The results obtained from the analysis of each data split – training and testing datasets, are presented in Tables 3, 4, and 5 respectively. Figures 3 to 11 depict the graphical visualizations of the confusion matrices of both the single and ensemble models.

5.2. Discussion

Table 3 shows the decision tree (C4.5) classifier obtained the highest accuracy of 98.33% compared to the naïve Bayes and SVM classifiers with an accuracy of 91.67% and 85.00% respectively. C4.5 performed better than others because it mitigates issues like overfitting, and incomplete data, and can handle discrete and continuous data respectively. In Table 4, the stacking ensemble model shows significant improvement over others by obtaining an accuracy of 99.89%, while the bagging, boosting and voting ensemble obtained an accuracy of 83.33%, 96.95%, and 93.33% respectively. The stacking ensemble performed better because it is a heterogeneous ensemble model comprising the three base classifiers of SVM, C4.5, and naïve Bayes respectively.

Table 3. Performance for Single Classification Models

Models		***F-Score%***	***Precision%***	***Recall%***	***Accuracy%***
SVM	*Training*	*83.65*	*84.97*	*85.00*	*85.50*
	Testing	*84.80*	*86.21*	*85.00*	*85.00*
naive Bayes	*Training*	*90.05*	*92.49*	*91.67*	*91.30*
	Testing	*92.10*	*93.29*	*91.67*	*91.67*
Decision Trees	*Training*	***98.90***	*99.98*	*99.92*	*99.95*
	Testing	***97.80***	***98.52***	***98.33***	***98.33***

Table 4. Ensemble Models Performance

Models		**F-Score%**	**Precision%**	**Recall%**	**Accuracy%**
Bagging (SVM)	*Training*	82.05	83.75	83.75	83.75
	Testing	82.65	83.33	83.33	83.33
Boosting (NB)	*Training*	96.07	97.68	97.50	97.50
	Testing	95.89	96.95	96.67	96.95
Stacking	*Training*	**99.05**	99.92	99.83	99.97
	Testing	**99.50**	**99.98**	**99.95**	99.89
Voting	*Training*	94.64	95.50	95.00	95.00
	Testing	93.65	94.20	93.33	93.33

The visualization of Figures 3 to 9 depicts the various models' misclassification. Out of the 60 samples presented, the SVM misclassified 9, naïve Bayes misclassified 5, and C4.5 misclassified 1. From the single models, C4.5 proves to be the best model with the least misclassification error. The bagged SVM misclassified 10 samples, the boosting of naïve Bayes misclassified 1 sample, the voting ensemble misclassified 4 samples and the stacking ensemble misclassified 0 samples respectively. The stacking ensemble was able to achieve this feat because it can reduce the bias and variation of the models merged and in turn increase the overall predictive performance significantly. Finally, our proposed model is compared with existing models as shown in Table 5. While our model obtained a 99.89% accuracy, others were 86.25%, 93.13%, 97.14%, 99.20%, and 87.18% respectively. Our model shows to be better than the five existing works except the work of [27], with a 0.09 difference, thereby justinfging the need for future improvement of the models.

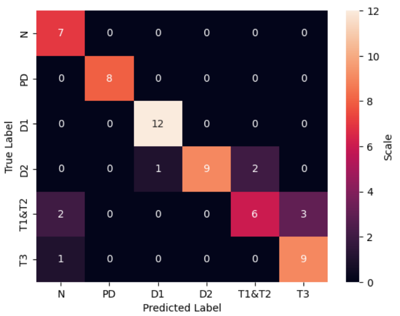

Figure 3. SVM Classifier

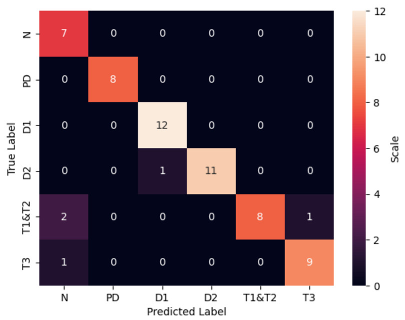

Figure 4. naive Bayes

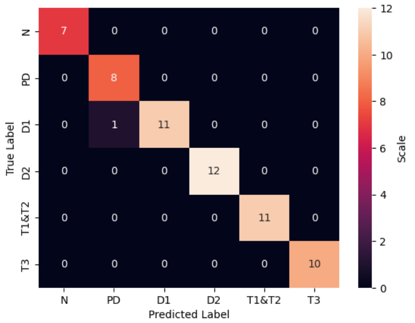

Figure 5. C4.5 DT

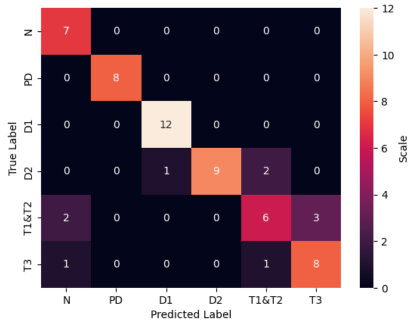

Figure 6. Bagging (SVM)

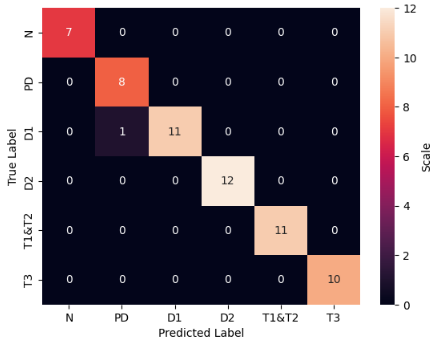

Figure 7. Boosting (naive Bayes)

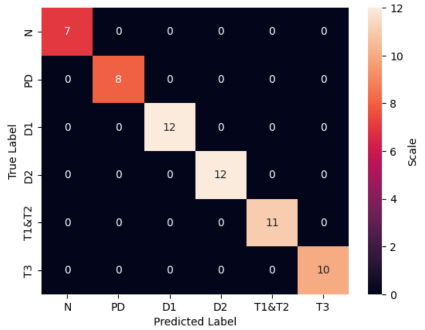

Figure 8. Stacking

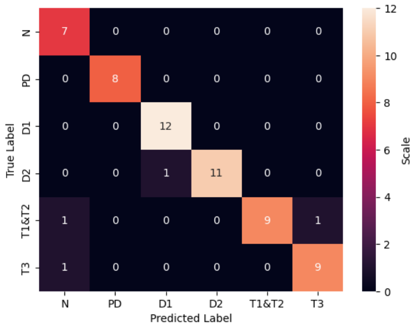

Figure 9. Voting

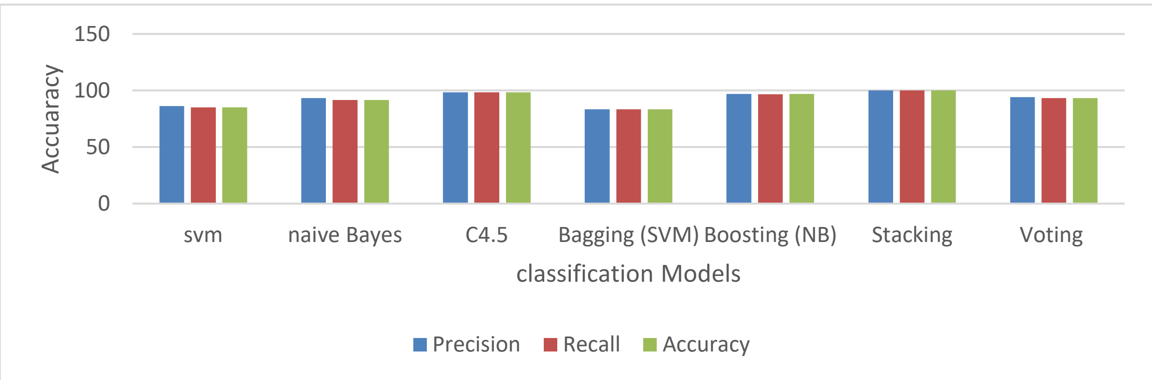

Figure 10. Performance Metrics of all Classification Models

Table 5. Comparison of Different Model Accuracies

Reference	Research methods	Accuracy (%)
Proposed model	Stacking	99.89
[4]	*naive Bayes*	86.25
[17]	*IGA-XGBoost*	99.20
[27]	*KosaNet*	99.98
[28]	*Ensemble (6) classifiers*	97.14
[29]	*SVM*	86.18
[35]	*Decision tree*	93.13

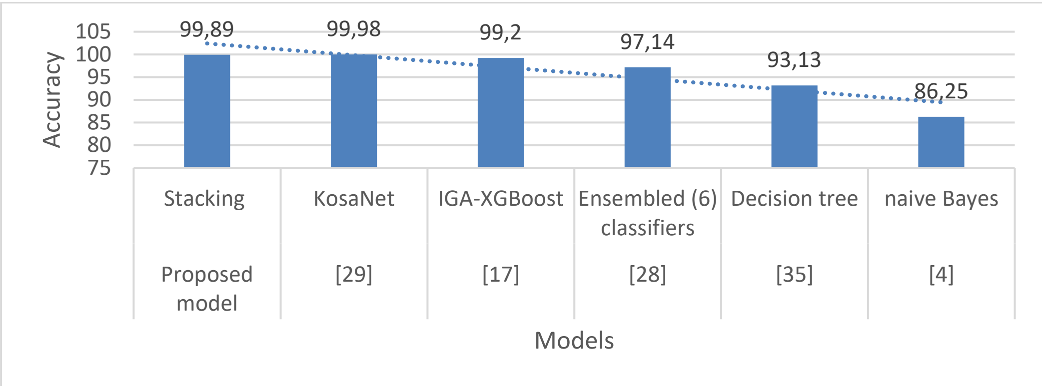

Figure 11. Comparison of Different Model Accuracy

6. Conclusion

In this study, we used seven classification methods and normalization approaches to create a power transformer fault types diagnostic model on 168 DGA data points with seven features obtained from IEC TC 10 databases. The seven classification models used were C4.5, SVM naive Bayes, and four ensemble methods: voting, stacking, bagging and boosting. The transformer fault detecting accuracies of the C4.5 decision tree classifier and the stacking ensemble models showed better performances than others with an accuracy of 98.33% and 99.89% respectively. The proposed model was compared with other existing works validating the superiority and adequacy of our proposed model with prospects. In future work, the authors hope to extend the input space, introduce other data transformation techniques like the Synthetic Minority Oversampling Technique (SMOTE) to handle data imbalance problems, introduce other DGA datasets from different domains, with improved machine learning algorithms and also introduce the concept of model interpretability. Also, we intend to extend the ensemble variants using other machine-learning techniques.

References

[1] Duval M, dePablo A (2001) Interpretation of Gas-In-Oil Analysis Using New IEC Publication 60599 and IEC TC 10 Databases. IEEE Electrical Insulation Magazine, vol. 17, no. 2, pp. 31–41.

[2] Izzularab MA, Aly, GEM, Mansour DA (2004) On-line diagnosis of incipient faults and cellulose degradation based on artificial intelligence methods", IEEE Int. Conf. on Solid Dielectrics, Toulouse, France.

[3] Hoballah, A., Mansour, J. G. and Taha IBM, (2020) Hybrid grey wolf optimizer for transformer fault diagnosis using dissolved gases considering uncertainty in measurements. IEEE Access, vol. 8, pp. 139176–139187.

[4] Mahamdi Y, Boubakeur A, Mekhaldi A, Benmahamed Y (2022) Power Transformer Fault Prediction using Naive Bayes and Decision tree based on Dissolved Gas Analysis. ENP Engineering Science Journal, Vol. 2, No. 1. Digital Object Identifier (DOI): 10.53907/enpesj.v2i1.63.

[5] Cheng L, Yu T (2018) Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: A survey. Energies 11, 913. doi:10.3390/en11040913

[6] Fu Wan W, Weigen Chen W, Xiaojuan Peng X, Jing Shi J (2012) Study on the Gas Pressure Characteristics of Photoacoustic Spectroscopy Detection for Dissolved Gases in Transformer Oil. In 2012 International Conference on High Voltage Engineering and Application. IEEE, 286–289. doi:10.1109/ICHVE.2012.6357108.

[7] Chen Xi, Chen W, Gan D (2010) Properties and Gas Production Law of Surface Discharge in Transformer Oil-Paper Insulation. In 2010 Annual Report Conference on Electrical Insulation and Dielectric Phenomena. IEEE, 1–4. doi:10.1109/CEIDP.2010.5724049.

[8] Zeng W, Yang Y, Gan C, Li H, Liu G (2011) Study on Intelligent Development of Power Transformer On-Line Monitoring Based on the Data of DGA. In 2011 Asia-Pacific Power and Energy Engineering Conference. IEEE, 1-4. doi:10.1109/appeec.2011.5749107

[9] Jiang XQ, Gong Y, Han S, Zhou K (2014) Application of the Improved Three-Ratio Method in Chromatographic

Analysis of Locomotive Transformer Oil. Amr 1030-1032, 29–33. doi:10.4028/www.scientific.net/amr.1030-1032.29
[10] Dhote N, Helonde J (2012) Diagnosis of Power Transformer Faults Based on Five Fuzzy Ratio Method. WSEAS Trans. Power Syst. 7, 12.
[11] Liu Z, Song B, Li E, Mao Y, Wang G (2015) Study of "code Absence" in the IEC Three-Ratio Method of Dissolved Gas Analysis. IEEE Electr. Insul. Mag. 31, 6-12. doi:10.1109/MEI.2015.7303257
[12] Shang H, Xu J, Zheng Z, Qi B, Zhang L (2019) A Novel Fault Diagnosis Method for Power Transformer Based on Dissolved Gas Analysis Using Hypersphere Multiclass Support Vector Machine and Improved D-S Evidence Theory. Energies 12, 4017. doi:10.3390/en12204017
[13] Singh S, Bandyopadhyay M (2010) Dissolved Gas Analysis Technique for Incipient Fault Diagnosis in Power Transformers: A Bibliographic Survey. IEEE Electr. Insul. Mag. 26, 41–46. doi:10.1109/MEI.2010.5599978
[14] Bhalla D, Bansal RK, Gupta HO (2011) Transformer Incipient Fault Diagnosis Based on DGA using Fuzzy Logic. IEEE, pages 1-5.
[15] Yadaiah N, Ravi N (2011) Internal Fault Detection Techniques for Power Transformers. Appl. Soft Comput. 11, 5259–5269. doi:10.1016/j.asoc.2011.05.034
[16] Dhini A, Faqih A, Kusumoputro B, Surjandari I, Kusiak A (2020) Data-driven fault Diagnosis of Power transformers using Dissolved Gas Analysis (DGA). International Journal of Technology. 11(2), 388-399. http://ijtech.eng.ui.as.id
[17] Wu Z, Zhou M, Lin Z, Chen X, Huang Y (2021) Improved Genetic Algorithm and XGBoost Classifier for Power Transformer Fault Diagnosis. Front. Energy Res. 9:745744. doi: 10.3389/fenrg.2021.745744.
[18] Yi J-H, Wang J, Wang G-G (2016) Improved Probabilistic Neural Networks with Self-Adaptive Strategies for Transformer Fault Diagnosis Problem. Adv. Mech. Eng. 8, 168781401562483. doi:10.1177/ 1687814015624832
[19] Mani G, Jerome J (2014) Intuitionistic Fuzzy Expert System Based Fault Diagnosis Using Dissolved Gas Analysis for Power Transformer. J. Electr. Eng. Tech. 9, 2058–2064. doi:10.5370/JEET.2014.9.6.2058
[20] Parihar VR, Nimkar DS, Warudkar S, Deshmukh R, Thakare M (2017) Power Transformer Protection using Fuzzy Logic based Controller. International Journal of Engineering Research. Volume No.6, Issue No.7, pp: 366-370
[21] Fan J, Wang F, Sun Q, Bin F, Liang F, Xiao X (2017) Hybrid RVMANFIS Algorithm for Transformer Fault Diagnosis. IET Generation, Transm. Distribution 11, 3637–3643. doi:10.1049/iet-gtd.2017.0547
[22] Yin J, Zhu Y, Yu G (2011) Power Transformer Fault Diagnosis Based on Support Vector Machine with Cross Validation and Genetic Algorithm. In 2011 International Conference on Advanced Power System Automation and Protection. IEEE, 309–313. doi:10.1109/APAP.2011.6180419
[23] Benmahamed Y, Teguar Y, Boubakeur A (2017) Application of SVM and KNN to Duval Pentagon 1 Transformer Oil Diagnosis. IEEE Trans. Dielect. Electr. Inst., 24, 3443–3451, 2017
[24] Aizpurua JI, Catterson VM, Stewart BG, McArthur SDJ, Lambert B, Ampofo B, Pereira G, Cross JG (2018) Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing. IEEE Trans. Dielectrics Electr. Insul., vol. 25, no. 2, pp. 494–506, Apr. 2018. DOI: 10.1109/TDEI.2018.006766.
[25] Igodan CE, Ojietohamen SE, Izevbizua RI (2023) Cervical cancer prediction using ensemble models, NIPES Conference proceedings. The 2nd International conference 15th-17th Feb. 2023, pp.71-83. www.nipesjournals.org.ng
[26] Yuan F, Guo J, Xiao Z, Zeng B, Zhu W, Huang S (2019) A Transformer Fault Diagnosis Model Based on Chemical Reaction Optimization and Twin Support Vector Machine. Energies 12, 960. doi:10.3390/en12050960
[27] Li J, Zhang Q, Wang K, Jianyi W (2015) Optimal Dissolved Gas Ratios Selected by Genetic Algorithm for Power Transformer Fault Diagnosis Based on Support Vector Machine. IEEE Transactions on Dielectrics and Electrical Insulation Vol. 23, No. 2. DOI: 10.1109/TDEI.2015.005277.
[28] Taha IBM, Mansour DEA (2021) Novel Power Transformer Fault Diagnosis Using Optimized Machine Learning Methods. Intelligent Automation & Soft Computing. IASC, 2021, vol.28, no.3. DOI:10.32604/iasc.2021.017703
[29] Odongo G, Musabe R, Hanyurwimfura DA (2021) Multinomial DGA Classifier for Incipient Fault Detection in Oil-Impregnated Power Transformers. Algorithms, 14, 128. https://doi.org/10.3390/a14040128.
[30] Dapshima BA, Essa YC, Chaturvedi S (2023) Fault Detection and Protection of Power Transformer Using Fuzzy Logic. International Journal for Research in Applied Science and Engineering Technology (IJRASET). Volume 11 Issue I. www.ijraset.com
[31] Li J, Zhang Q, Wang K, Jianyi W (2015) Optimal Dissolved Gas Ratios Selected by Genetic Algorithm for Power Transformer Fault Diagnosis Based on Support Vector Machine. IEEE Transactions on Dielectrics and Electrical Insulation Vol. 23, No. 2. DOI: 10.1109/TDEI.2015.005277.
[32] Tao L, Yang X, Zhou Y, Yang LA (2021) Novel Transformers Fault Diagnosis Method Based on Probabilistic Neural Network and Bio-Inspired Optimizer. Sensors, 21, 3623. https://doi.org/10.3390/s21113623.
[33] Huang X, Yuan Y, Li J (2023) A Review of Transformer Fault Diagnosis Based on Information System Theory and Machine Learning. Preprints, pp.1-12. doi:10.20944/preprints202305.0036.v1.
[34] Bouchaoui L, Hemsas KE, Mellah H, Benlahneche S (2021) Power Transformer faults diagnosis using undestructive methods (Rogers and ICE) and artificial neural network for dissolved gas analysis applied on the functional transformer in the Algerian north-eastern: A comparative study. Electrical Engineering & Electromechanics 2021(4):3-11. DOI: 10.20998/2074-272X.2021.4.01.
[35] Demirci M, Gozde H, Taplamacioglu MC (2021) Fault Diagnosis of Power Transformers with Machine Learning Methods using Traditional Methods Data. International Journal on Technical and Physical Problems of Engineering.

Issue 49, Volume 13, Number 4, pp. 225-230
[36] Islam MM, Lee G, Hettiwatte, SN (2018) Application of Parzen Window estimation for incipient fault diagnosis in power transformers. The institute of engineering and technology. doi: 10.1049/hve.2018.5061.
[37] Ma H, Zhang W, Wu R, Yang C (2018) A Power transformers fault diagnosis model based on three DGA ratios and PSO optimization SVM. IOP Conf. Series: Materials Science and Engineering 339 012001. doi:10.1088/1757-899X/339/1/012001.
[38] Taha IBM, Mansour DEA (2021) Novel Power Transformer Fault Diagnosis Using Optimized Machine Learning Methods. Intelligent Automation & Soft Computing. IASC, 2021, vol.28, no.3. DOI:10.32604/iasc.2021.017703
[39] Igodan EC, Obe O, Thompson A.F-B. and Owolafe., O. (2022a). Erythemato Squamous Disease Prediction using Ensemble Multi-Feature Selection Approach. International Journal of Computer Science and Information Security (IJCSIS), Vol. 20, No. 2, pp. 95-106.
[40] Igodan EC, Obe, O., Thompson, AF-B, Owolafe O (2022b) Prediction of erythemato Squamous-disease using ensemble learning framework. The Institute of Engineering and Technology. In Explainable Artificial Intelligence in Medical Decision Systems, pp.197-228.
[41] Igodan CE, Ukaoha KC (2019) Using Multilayer Perceptron and Deep Neural Networks for the Diagnosis of Breast Cancer Classification", 2019 IEEE AfriCon, pp. 1-7, doi:101109/AFRICON46755.2019.9133873.
[42] Vapnik CV, (1995) Support-vector networks. Machine Learning 20, 3, 1995, 273. doi:10.1007/BF00994018.
[43] Liu J, Ning B, Shi D (2019) Application of Improved Decision Tree C4.5 Algorithms in the Judgment of Diabetes Diagnostic Effectiveness. IOP Conf. Series: Journal of Physics: Conf. Series 1237 (2019) 022116. doi:10.1088/1742-6596/1237/2/022116.
[44] Xuanyuan S, Xuanyuan S, Yue Y (2022) Application of C4.5 Algorithm in Insurance and Financial Services Using Data Mining Methods. Mobile Information Systems, vol. 2022, Article ID 5670784, 8 pages, 2022. https://doi.org/10.1155/2022/5670784.
[45] Witten IH, Frank E, Hall M A, Pal CJ (2017) Data Mining: Practical Machine Learning Tools and Techniques. Fourth Edition, Cambridge, U.S.A.
[46] Seijo-Pardo B, Porto-Diaz I, Bolon-Canedo V, Alonso-Betanzos A (2017) Ensemble Feature Selection: Homogeneous and Heterogeneous Approaches. Knowledge-Based System. 2017; 118:124-139.
[47] Liu J, Dong X, Zhao H, Tian Y (2022) Predictive Classifier for Cardiovascular Disease Based on Stacking Model Fusion. Processes. 10(4):749. https://doi.org/10.3390/pr10040749 [Assessed 27th March 2022
[48] Rokach L (2019) Ensemble Learning Pattern Classification Using Ensemble Methods. Second Edition. Volumn 85. ISSN: 1793-0839. https://dokumen.pub/ensemble-learning-pattern-classification-using-ensemble-methods-2nbsped-9811201951-9789811201950.html
[49] Tattar PN (2018) Hands-On Ensemble Learning with R: A beginners' guide to combining the power of machine learning algorithms using ensemble techniques. https://dokumen.pub/hands-on-ensemble-learning-with-r-a-beginners-guide-to-combining-the-power-of-machine-learning-algorithms-using-ensemble-techniques-1788624149-9781788624145.html
[50] Villacampa O (2015) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis. Doctoral Dissertation Nova Southeastern University. Retrieved from NSUWorks, College of Engineering and Computing. https://nsuworks.nova.edu/gscis_etd/63.
[51] Bolon-Canedo V, Sanchez-Marono N, Alonso-Betanzos A (2015) Feature Selection for High-Dimensional Data. Springer.
[52] Brownlee J (2019) Statistical methods for machine learning: Discover how to transform data into knowledge with Python. https://machinelearningmastery.com/statistics_for_machine_learning/
[53] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16 (2002) 321-357.
[54] Shehzad K, Zhenhua T, Shoukat S, Saeed A, Ahmad I, Sarwar Bhatti S, Chelloug, SAA (2023) Deep-Ensemble-Learning-Based Approach for Skin Cancer Diagnosis. Electronics 12, 1342. https://doi.org/10.3390/electronics12061342.
[55] Witten IH, Frank E (2000) Data Mining: Practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, San Francisco.
[56] Ghoneim SSM, Taha IBM, Elkalashy NI (2016) Integrated ANN Based Proactive Fault Diagnostic Scheme for Power Transformers Using Dissolved Gas Analysis. IEEE Trans. Dielect. Electr. Insul. 23, 1838–1845. doi:10.1109/TDEI.2016.005301
[57] Li J, Chen X, Wu C (2009) Application of Comprehensive Relational Grade Theory in Expert System of Transformer Fault Diagnosis. In 2009 International Workshop on Intelligent Systems and Applications. IEEE, 1-4. doi:10.1109/iwisa.2009.5072742
[58] Zarkovic M, Stojkovic Z (2017) Analysis of Artificial Intelligence Expert Systems for Power Transformer Condition Monitoring and Diagnostics. Electric Power Syst. Res. 149, 125–136. doi:10.1016/j.epsr.2017.04.025
[59] Zhu X, Xiong J, Liang Q (2018) Fault Diagnosis of Rotation Machinery Based on Support Vector Machine Optimized by Quantum Genetic Algorithm. IEEE Access 6, 33583–33588. doi:10.1109/ACCESS.2018.2789933
[60] Zhou Z-H (2012) Ensemble methods: foundation and algorithms. CRC Press.

[61] Manisha, Kaur, K, Sharma NK, Singh J, Bhalla D (2022) Performance Assessment of IEEE/IEC Method and Duval Triangle technique for Transformer Incipient Fault Diagnosis

Conflict of Interest
Author declares that there are no potential conflicts of interest.

Funding
This research did not receive a specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethical Approval
It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefited from are stated in the bibliography

Plagiarism Statement
This article has been scanned by iThenticate ™.