Jurnal Kejuruteraan 38(5) 2026: 2221-2238
https://doi.org/10.17576/jkukm-2026-38(5)-11

Transformer Fault Classification Based on Dissolved Gas Compositions Using Microsoft Azure Machine Learning

Chee Ying Chan[a,b*], Salmey Abdul Halim[a], Nursyarizal Mohd Nor[b] & Nor Zaihar Yahaya[b]

[a]*Group Technical Solutions, Project, Technology & Health, Safety, Security, Environment Division,*

PETRONAS, Kuala Lumpur, Malaysia

[b]*Electrical & Electronics Engineering Department, Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia*

[*]Corresponding author: chancheeying@petronas.com

Received 12 July 2025, Received in revised form 25 February 2026
Accepted 25 March 2026, Available online 30 August 2026

ABSTRACT

This study presents a novel diagnostic methodology for oil-filled transformers, leveraging dissolved gas analysis (DGA) and a no-code machine learning framework. Conventional diagnostic methods such as Duval Triangle and gas ratio techniques are constrained by limited adaptability to modern transformer designs, susceptibility to misdiagnosis, and reliance on expert interpretation. Meanwhile, artificial intelligence (AI) approaches often demand significant programming expertise, large datasets, and computational resources. To address these challenges, the research introduced a unified machine learning model developed using Microsoft Azure Machine Learning, enabling field engineers to build and deploy diagnostic models without coding. The model utilized concentrations of 5 key gases: methane, ethylene, acetylene, ethane, and hydrogen, as input features to classify transformer conditions into 6 fault classes, including a no-fault state. 3 algorithms, namely Gradient Boosting, Random Forest, and Logistic Regression were evaluated, with Gradient Boosting consistently outperforming other models and Duval Triangle, achieving the highest accuracy of 81.3% with 2 false negatives. When the output labels were consolidated to 4 fault types, diagnostic accuracy improved to 95.6% without any false negative. Model performance was optimized through experiments involving train-test ratio variation and feature scaling. The model was further validated through field trials on actual transformers, demonstrating high reliability and alignment with forensic findings. This no-code approach empowers engineers to integrate domain expertise directly into model development, enhancing diagnostic precision and operational efficiency. The findings support scalable, interpretable, and proactive maintenance strategies, contributing to resilient infrastructure and sustainable energy systems.

Keywords: Classification; dissolved gas; fault diagnosis; machine learning; transformer

INTRODUCTION

The growing global demand for electricity is closely tied to economic progress, technological advancement, and improved quality of life. Ensuring reliable and sustainable power supply is essential for achieving the United Nations' Sustainable Development Goals (SDGs), particularly SDG 7, which promotes universal access to affordable and clean energy (Fonseca, Domingues, and Dima, 2020). Transformers are essential equipment in power generation and transmission systems, enabling the transfer of electrical energy from the source to endusers. They are considered critical equipment due to their complex internal design and the high cost involved in their manufacturing (Mohamed et al. 2024). This paper explores the use of machine learning (ML) to enhance transformer fault diagnosis through dissolved gas analysis (DGA). Traditional diagnostic methods often face limitations in accuracy and adaptability. By integrating machine learning, especially through no-code platforms, engineers can improve diagnostic precision, reduce downtime, and support proactive maintenance. This approach also aligns with SDG 9 by fostering innovation and resilient infrastructure. The study aims to demonstrate how data-driven techniques can contribute to a more efficient, dependable, and sustainable energy landscape.

INTERPRETATION OF DISSOLVED GAS COMPOSITION

Oil-filled transformers comprise cellulose insulated copper windings tightly wound on a ferromagnetic core and housed in a sealed tank of insulating fluid that also serves to cool the windings (Rampersad, Bahadoorsingh, and Sharma, 2018). The transformer insulating materials breakdown and release several different gases (E. Abbasi and O. P. Malik, 2016) as electrical and thermal stresses accumulate. DGA is a widely used and highly effective method for detecting early faults in oil-filled transformers (Zhang et al. 2022). DGA identifies the types and amounts of gases produced in a transformer, which reflect the nature and severity of internal faults. When gas concentrations or their rate of increase exceeds typical thresholds, it strongly indicates the presence of a fault.

Field engineers must analyze transformer DGA results from laboratories, comparing dissolved gas values against industry standards. Abnormal readings require proper interpretation to classify fault type and severity, guiding necessary actions. Accurate fault classification helps pinpoint fault locations, aiding monitoring and inspection efforts. Numerous interpretation methods are available for fault identification of transformer based on its dissolved gas composition. The conventional interpretation methods include the Key Gas Method, IEC Ratios, Rogers Ratios, Doernenburg Ratios, Duval Triangles and Duval Pentagons (IEEE Power & Energy Society, 2019). Despite being known for their simplicity and ease of implementation, the ratio methods may result in "no interpretation" due to incomplete ratio ranges of the methods, while the pictographic method by Duval, such as the Duval Triangle, requires pre-determination if a transformer has issue as there is no fault-free area in the triangles and pentagons. Consequently, the conventional methods raise persistent concerns regarding failure to diagnose and misdiagnosis. Expert judgment is crucial resulting in diverse interpretations and mitigations (Hydroelectric Research and Technical Services Group, 2005). Conventional methods also exhibit inflexible standards, struggling to adapt to evolving transformer designs, materials, and issues.

Due to the limitations of conventional fault diagnosis methods, researchers have increasingly turned to artificial intelligence (AI) techniques. This shift has led to significant advancements in transformer fault analysis, delivering high levels of accuracy and efficiency (Abdo et al. 2021). The automated creation of analytical models is widely known in data analysis as machine learning (ML). As a branch of AI, ML is founded on the principle that machines can learn from data, identify patterns, and make decisions independently. This reduces or eliminates the need for human involvement (Malla et al. 2024). These computational intelligence techniques incorporate elements of adaptation, learning, evolution, and evidence to formulate intelligent conclusions that address the discrepancies found in outcomes from conventional methods (Kaur, Bhalla, and Singh, 2023).

CHALLENGES AND MOTIVATION

Although machine learning has made impressive progress in recent years, several key challenges remain (Utmal, 2021). One major challenge is the need for large volumes of high-quality data to effectively train and validate machine learning models. Additionally, selecting, tuning, and interpreting these models often requires specialized domain knowledge, making the process complex and dependent on skilled programmers or data scientists. The computational complexity and resource requirements of some machine learning algorithms can present practical challenges, especially in real-time fault diagnosis applications.

Table 1 shows 30 research works published in the past 5 years, addressing the application of AI-based techniques in transformer fault diagnosis. The study of these related works reveals a common theme centered around the development of AI-based techniques for achieving high prediction accuracy. However, beyond prediction accuracy, there are 4 features which are deemed crucial for ensuring the models are relevant and practical for deployment in the industry: (1) Application of no-code machine learning approach, which allows technical domain experts, i.e. field engineers to develop fault diagnosis models without writing any code. This would minimize resources, including machine learning domain expertise, computing power, time and cost; (2) Mitigation of false negatives since failing to detect a fault may lead to transformer failure and catastrophic events. By mitigating false negatives, the model will be capable of consistently detecting faults; (3) Indentification of "no fault" condition helps to eliminate additional step required to confirm if the transformer is abnormal before executing fault classification; (4) Implementation of field trials, which involves deploying the model and validating it with actual field data of faulty transformers. This exercise aims to assess the practicality and robustness of the model in real-world condition.

Based on Table 1, no research work had been conducted using no-code machine learning approach. 15 models could differentiate heathy transformers from the faulty ones, while only 1 paper considered false negative.

7 papers implemented field trials with diversified number of samples and accuracy. 28 papers presented the accuracy of the models developed, ranging from 65.3% to 99.7%. The average accuracy is 91.4%.

PROPOSED SOLUTION AND CONTRIBUTION

This paper aims to: (1) Develop a single and unified machine learning model, using a no-code approach, to differentiate between healthy and faulty transformers, and accurately identify the fault type and severity; (2) Optimize the performance of the machine learning model, in terms of accuracy and false negative, using preprocessing techniques and variation in output labels; (3) Validate the performance of the machine learning model through field trials on actual transformers that have experienced faults during operation in real-world conditions.

The significance of this research lies in its potential to improve transformer fault diagnosis for field engineers. The adoption of no-code approach empowers field engineers to develop, train, and test the models themselves, without the need for significant computational resources and extensive programming knowledge, or reliance on specialized data scientists. This allows engineers to incorporate their insights, expertise, and experiences into the model, leading to more accurate and contextually relevant diagnostics. The integration of machine learning techniques with the capability to differentiate healthy and faulty transformers, minimise false negatives and successful validation with field trials, not only enhances the accuracy and reliability of fault detection, but also empowers engineers to take control of the diagnostic process, leading to improved maintenance strategies and overall system reliability.

TABLE 1. Investigation on accuracy and features of 30 published transformer fault diagnosis models

No.	Reference	Platform/ Programming Language	Accuracy (%)	Features			
				Apply No-code ML Approach	Identify "No Fault" Condition	Mitigate False Negative	Implement Field Trial
1	Misbahulmunir, Ramachandaramurthy, and Thayoob (2020)	MATLAB	74.0	No	No	No	Yes
2	Saravanan et al. (2020)	MATLAB, Python	81.4	No	No	No	No
3	Rao et al. (2021)	Not disclosed	90.3	No	No	No	Yes
4	Wang, Littler, and Liu (2021)	Not disclosed	99.6	No	Yes	No	Yes
5	Saha, Baruah, and Nayak (2021)	Python	83.0	No	No	No	No
6	Al-Katheri et al. (2021)	MATLAB	97.7	No	Yes	No	No
7	Taha, Ibrahim, and Mansour (2021)	MATLAB	97.4	No	No	No	No
8	Yaqin and Khayam (2021)	Not disclosed	94.9	No	No	No	No
9	Ghoneim et al. (2021)	Not disclosed	83.2	No	No	No	No
10	Abdo et al. (2021)	MATLAB	88.9	No	No	No	No
11	Patil et al. (2023)	Anaconda Navigator	99.5	No	Yes	No	No
12	(Ru et al. 2022)	Not disclosed	95.8	No	No	No	No
13	(Raghuraman and Darvishi, 2022)	Python	88.0	No	Yes	No	No
14	Paul et al. (2022)	Python	99.9	No	Yes	No	Yes
15	Verma and Chandel (2023)	MATLAB	96.0	No	No	No	No
16	Bharath et al. (2022)	Not disclosed	-	No	No	No	No
17	Poonnoy, Suwanasri, and Suwanasri (2022)	Not disclosed	99.0	No	Yes	No	No
18	Subalakshmi et al. (2022)	Not disclosed	98.4	No	Yes	No	No
19	Menezes et al. (2022)	SIPINA	99.4	No	No	No	No

continue...

...cont.

No.	Reference	Platform/ Programming Language	Accuracy (%)	Features			
				Apply No-code ML Approach	Identify "No Fault" Condition	Mitigate False Negative	Implement Field Trial
20	Wang et al. (2023)	Not disclosed	92.3	No	Yes	No	No
21	Prasojo et al. (2023)	Not disclosed	85.7	No	Yes	No	No
22	Kaur, Bhalla, and Singh, (2023)	Not disclosed	-	No	No	No	No
23	Jin et al. (2023)	Not disclosed	92.0	No	Yes	No	No
24	Malik, Sharma, and Naayagi (2023)	Not disclosed	100	No	Yes	No	Yes
25	Demirci, Gözde, and Taplamacioglu (2023)	Not disclosed	91.2	No	No	No	No
26	Das et al. (2023)	Not disclosed	98.84	No	Yes	Yes	Yes
27	Manoj et al. (2023)	MATLAB	84.9	No	Yes	No	No
28	Jing, Zhang, and Yang (2023)	Not disclosed	99.7	No	Yes	No	No
29	Jan et al. (2023)	MATLAB	82.1	No	Yes	No	No
30	Gao et al. (2023)	Not disclosed	65.3	No	No	No	Yes

METHODOLOGY

Figure 1 outlines a structured process for developing and deploying machine learning models for transformer fault diagnosis. The workflow is divided into 3 main activities: Model training, model testing and model deployment. Each step in the workflow is explained in detail in the following sections. Finally, the key limitations of the methodology are discussed, offering insights for future improvements.

MODEL TRAINING

This research utilized the Microsoft Azure Machine Learning in transformer fault diagnosis. Azure Machine Learning is a cloud-based service which empowers users to construct, train, deploy, predict outcomes, and cluster data through a user-friendly web browser (Data Science Team, 2020). One of the key advantages of Azure Machine Learning is its no-code web interface, which enables field engineers to build and train models without writing any code. This simplifies the development process and allows engineers to work directly with their data. Model training involves selecting appropriate input features and output labels, preparing datasets, choosing suitable algorithms, and fine-tuning hyperparameters to optimize performance.

Input features are the variables used to train a machine learning model and make predictions. In this study, the selected input features were derived from the concentrations of 5 key dissolved gases identified through DGA. These gases are measured in parts per million (ppm) and serve as the primary numerical data points for the machine learning model. Their specific roles and diagnostic significance are outlined in Table 2.

Output labels, also known as target variables, represent the values a machine learning model aims to predict based on the input features. In this study, the output labels corresponded to the transformer fault classes, which were determined by the concentrations of dissolved gases. The model categorized faults into 6 distinct classes, each representing a specific fault type or a no-fault condition, as shown in Table 3. To align with the available ground truth data for model training from Duval and DePablo (2001), T1 and T2 fault types were combined into a single class.

TABLE 2. Input features

No.	Dissolved Gas	Significance
1	Methane (CH_4)	Methane is typically produced during low-temperature thermal faults. Its concentration can indicate overheating of oil or paper.
2	Ethylene (C_2H_4)	Ethylene is generated during higher temperature thermal faults. Its presence suggests overheating of oil or paper. High levels of ethylene can indicate severe thermal stress.
3	Acetylene (C_2H_2)	Acetylene is created from arcing in oil or paper at very high temperatures and high-energy discharges. Its detection is crucial for identifying electrical faults such as arcing in oil or paper insulation.
4	Ethane (C_2H_6)	Ethane is produced during low to moderate temperature thermal faults. It often accompanies methane and ethylene, providing additional information about the thermal condition of the transformer.
5	Hydrogen (H_2)	Hydrogen is a general indicator of various fault types. Its presence in significant quantities can signal the occurrence partial discharge and electrical discharges.

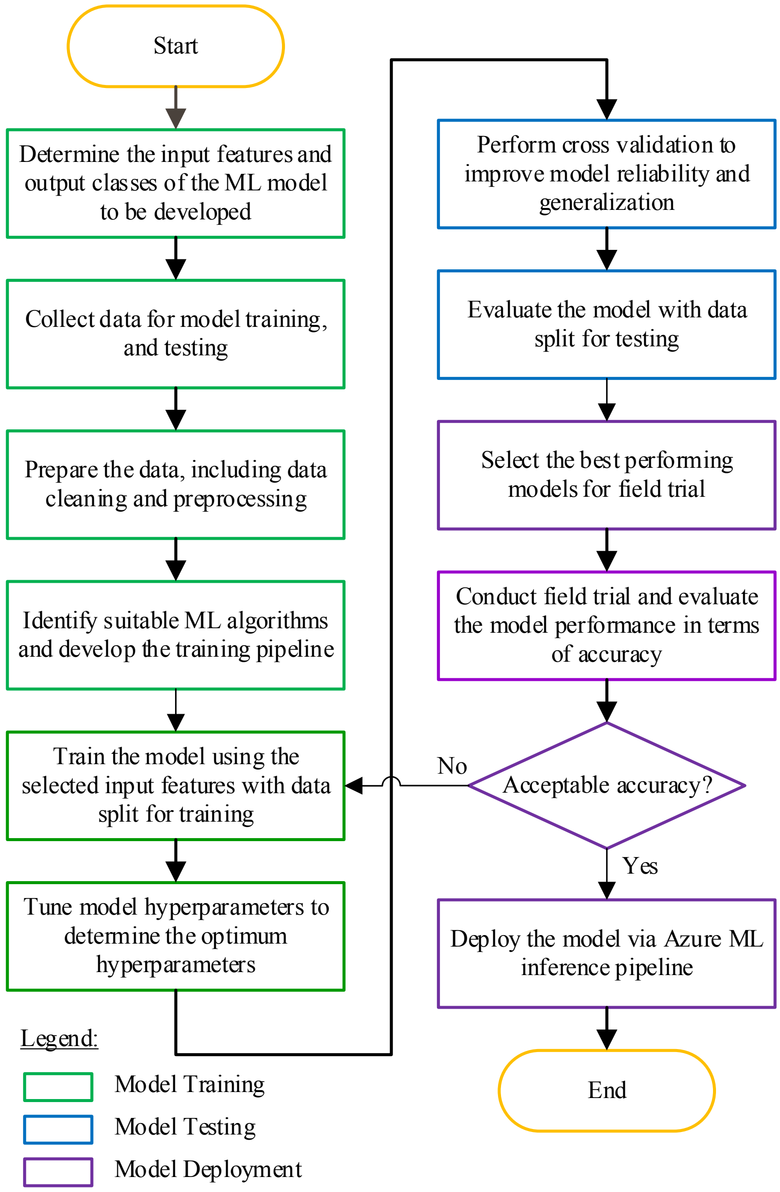


FIGURE 1. Development of machine learning models using Microsoft Azure Machine Learning

TABLE 3. Output labels

No.	Fault Class	Description
1	D1	Discharges of Low Energy
2	D2	Discharges of High Energy
3	PD	Partial Discharge
4	T1/T2	Thermal Faults below 300°C (T1) and between 300°C and 700°C (T2)
5	T3	Thermal Faults above 700 °C
6	NF	No Fault

Datasets are required for training and testing of the machine learning model. In this study, the machine learning models were trained and tested with exclusively with 100% of ground truth datasets. Ground truth data represents the actual outcomes, labels, or values that the model aims to predict. Ground truth data is crucial in supervised learning, where models are trained on input-output pairs. During training, the model analyzes patterns and relationships within the input data to make predictions. Ground truth data provides the correct output values for each input, enabling the models to learn and improve their predictive performance.

A total of 639 ground truth datasets were collected from published sources, including (Duval and DePablo, 2001),(Hooshmand, Parastegari, and Forghani, 2012), (Pattanadech et al. 2019), (A.S. Monzaid, 2018) and expert networks. These datasets represent transformers that were either confirmed healthy for at least one year following oil sampling or had experienced operational faults and were subsequently taken out of service. Each case was validated through visual inspections conducted by experienced engineers and maintenance specialists, ensuring accurate fault classification. The distribution of fault classes is illustrated in Figure 2.

The selection of machine learning algorithms for this research was based on 4 key criteria. Firstly, the algorithms needed to be capable of handling multiclass classification tasks, as the goal was to accurately diagnose various fault types in oil-filled transformers. Secondly, the algorithms must be robust and able to handle imbalanced datasets, which is common in fault diagnosis scenarios. Thirdly, the computational efficiency and scalability of the algorithms were considered, ensuring they could be effectively implemented within the Azure Machine Learning environment. Lastly, the interpretability of the models was also a factor, as it is important for field engineers to understand and trust the diagnostic results.

Using these criteria, 3 multiclass classification algorithms were selected for this research. They were Random Forest (RF), Gradient Boosting (GB), and Logistic Regression (LR). Supervised learning was implemented, which involved training models using labeled data, where each data point was associated with a known label or outcome. This approach allowed the model to learn the relationship between input features and the target labels, enabling it to make predictions on new, unseen data.

The "Tune Model Hyperparameter" component was employed to optimize the hyperparameters of machine learning models to achieve optimal performance. Hyperparameter tuning, identified the best set of hyperparameters that result in the highest model performance. Random sweep was selected as the sweeping mode, with maximum 200 of runs in measuring the accuracy of the model.

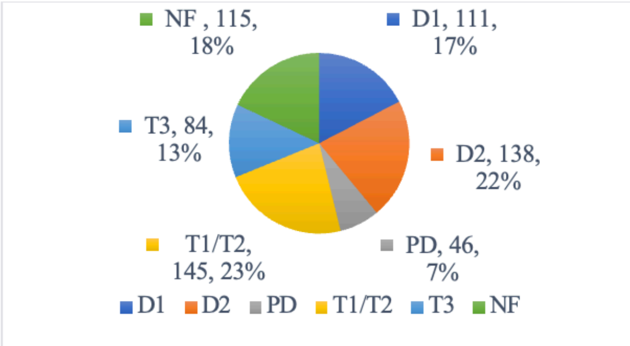


FIGURE 2. Distribution of Fault Classes for Ground Truth Datasets (6 Fault Classes)

MODEL TESTING

The performance of the machine learning models was evaluated using several key metrics, including accuracy, false negatives and confusion matrix.

Accuracy measures the proportion of correctly predicted instances relative to the total instances in the dataset. It offers a clear indication of the model's overall effectiveness and is calculated as the ratio of correct predictions to total predictions. In this study, model accuracy was not the only metric of comparison. Unnoticed faults (false negatives) were considered as the costliest to business and shall be mitigated as much as possible. A false negative occurs when the model fails to identify a condition or class that is present. In fault diagnosis, this is especially critical, as missed fault detections can have serious consequences. Confusion matrix provides a detailed visual summary of the model's predictions by contrasting actual labels with predicted labels. It breaks down performance into 4 categories, i.e. true positives, true negatives, false positives, and false negatives.

To ensure robust generalization to unseen data, this research employed k-fold cross-validation. The dataset was divided into k equally sized subsets (folds). In each of the k iterations, the model was trained on k-1 folds and evaluated on the remaining fold. This process was repeated such that each fold served as the test set once, and the final performance metrics were averaged across all iterations to provide a comprehensive estimate. To further enhance evaluation reliability, stratified k-fold cross-validation was used. This technique maintains the same distribution of fault classes in each fold as in the original dataset, ensuring balanced representation across splits. For this study, 9 folds were used, with 64 examples in each fold.

MODEL DEPLOYMENT

After training and testing, the machine learning models were deployed to a production environment for inference, where new input data was processed through the model or pipeline to produce outputs (predictions). The deployment process, along with model retraining, was managed using inference and training pipelines within Azure Machine Learning, as illustrated in Figure 3. This setup supports scalable prediction delivery and facilitates automated model updates based on newly available data. The field trials in this study were conducted using the inference pipeline developed.

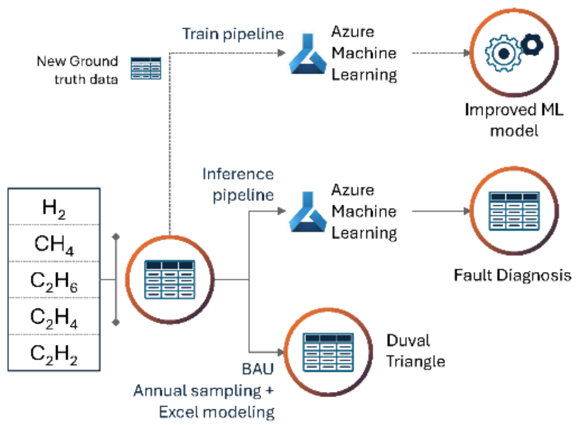


FIGURE 3. Deployment of machine learning model using Microsoft Azure Machine Learning

Establishing structured feedback loops is essential to maintaining the long-term effectiveness of AI systems. Ongoing performance reviews, together with insights from relevant stakeholders, allow organizations to pinpoint areas that require refinement and ensure the AI continues to operate at an optimal level (Hafiz et al. 2025). To continuously evaluate the performance of the machine learning model, dissolved gas composition data can be routinely uploaded to the Azure Machine Learning inference pipeline. This enables fast fault diagnosis and immediate comparison with the conventional Duval Triangle. Additionally, incorporating fresh ground truth data into the training pipeline supports ongoing model retraining and refinement. By regularly updating the model with real-world data, the system adapts to changing fault patterns and enhances its diagnostic precision over time. This iterative learning cycle ensures that the model remains robust and dependable, delivering consistent, high-quality fault detection essential for transformer maintenance.

LIMITATIONS

One of the main challenges in this study was obtaining high-quality data. Machine learning models need large, reliable datasets for effective training and validation. Ground truth data is ideal but often difficult and costly to acquire. In this study, ground truth datasets were collected from published sources and expert networks, but they were limited in quantity and diversity. Due to these constraints, it was not feasible to develop a model capable of distinguishing arcing faults in both paper and oil insulation.

Another major issue was data imbalance. Some fault types were underrepresented, which risked biasing the model toward more common classes. Oversampling techniques were used to address this, but they occasionally reduced model accuracy. Achieving a balanced dataset without introducing noise or overfitting remained a delicate task.

Field trials offered useful insights into real-world performance, but they were limited by the small number of transformers tested. While results were encouraging, a larger sample size is needed to confirm the model's robustness. Additionally, the trials were conducted in a specific operational setting (tropical, saline environment, continuous high loading), which may not reflect the full range of conditions found across different industries and geographical locations.

RESULTS AND DISCUSSION

This section provides a comprehensive evaluation of the machine learning models developed for diagnosing faults in oil-filled transformers using Azure Machine Learning. It began by comparing the performance of the baseline model with 2 experimental setups incorporating different data preprocessing techniques.

Next, the analysis explored the impact of output label variations, followed by results from field trials that validated the models' effectiveness under real-world operating conditions.

In each case, the performance of the models was assessed by evaluating their accuracy and false negatives. These results were then compared with the diagnostic accuracy of the Duval Triangle. Duval Triangle method has been proven to be accurate and dependable over many years since it was first introduced in 1960s (Hydroelectric Research and Technical Services Group, 2005). It is regarded as one of the most successful conventional diagnostic methods. This comparison served as a crucial benchmark to determine if the machine learning models could surpass the performance of the Duval Triangle.

BASE CASE

The base case served as a reference scenario against which variations and alternative configurations were compared. It functioned as a control benchmark to assess the impact of changes or interventions in model performance. Figure 4 illustrates the performance of 3 selected machine learning models: Random Forest, Gradient Boosting, and Logistic Regression, under the base case conditions.

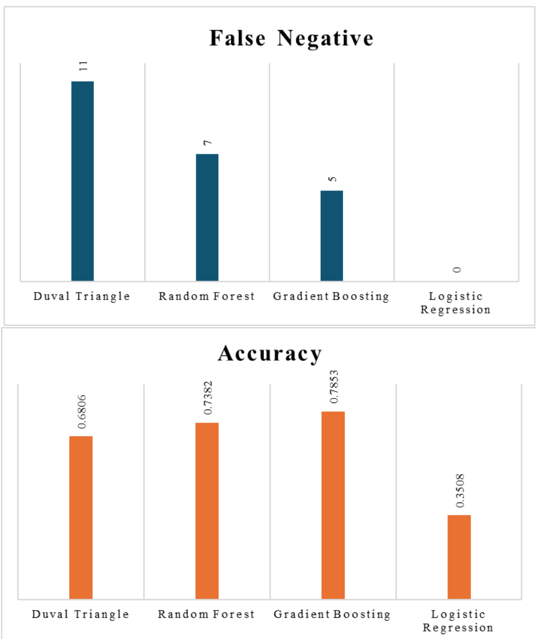


FIGURE 4. Model performance at base case (train-test ratio 70:30)

For this setup, a 70:30 train-test split was employed, and no additional data preprocessing techniques were applied. The Gradient Boosting model achieved the highest accuracy, recording 78.53% with 5 false negatives. This was closely followed by the Random Forest model, which attained 73.82% accuracy with 7 false negatives. Although the Logistic Regression model produced no false negatives, its overall accuracy was considerably lower at 35.08%, indicating limited generalization performance. For comparison, Duval Triangle's accuracy was 68.06% with 11 false negatives.

A 70:30 train-test split was used as the base case configuration. This ratio is widely recognized as a standard practice in machine learning due to its effectiveness across diverse applications. It provides a balanced division, allocating 70% of the data for model training and 30% for testing, offering sufficient data for the model to learn meaningful patterns while preserving enough data to evaluate its generalization to unseen inputs. Moreover, this setup helps mitigate overfitting, where the model may memorize noise in the training set instead of learning generalizable features. However, while commonly used, the 70:30 ratio is not universally optimal. Therefore, the impact of varying the train-test split was explored in the next section to assess its influence on model performance.

VARIATION IN TRAIN-TEST RATIO

To assess the effect of data allocation on model performance, various train-test ratios were explored. The objective was to strike a balance between providing sufficient data for training and enough for reliable testing. Higher training proportions (e.g., 90:10) enable models to learn more from the dataset, potentially enhancing accuracy and robustness. However, this comes at the expense of having less data for performance evaluation. On the other hand, lower training proportions (e.g., 60:40) offer a more rigorous evaluation by increasing the test set size but may limit the model's learning capacity. By experimenting with these split ratios, the study aimed to identify an optimal configuration that maximizes predictive performance while ensuring trustworthy evaluation.

Figure 5 illustrates the results for 3 machine learning models across different split ratios: 60:40, 70:30, 80:20, and 90:10. Interestingly, increasing training data did not consistently improve accuracy across all models. Each model exhibited differing degrees of sensitivity to the size of the training dataset. Ensemble methods like Random Forest and Gradient Boosting were more complex models capable of leveraging larger datasets effectively, which helped mitigate overfitting and improve generalization. In contrast, simpler models such as Logistic Regression gained limited benefit from additional data once the core patterns in the dataset had been captured. In this experiment, the highest accuracy was achieved by the Gradient Boosting model, reaching 81.25% with only 2 false negatives when using a 90:10 train-test split.

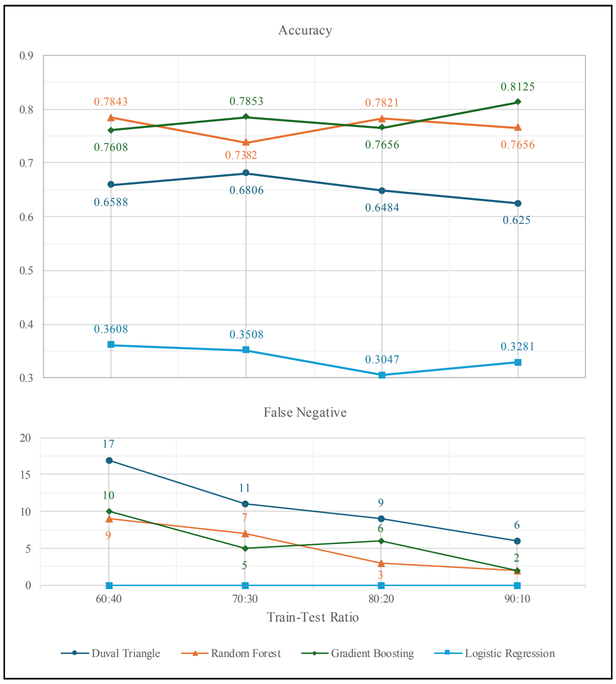


FIGURE 5. Model performance in terms of accuracy and false negative with varied train-test ratio

FEATURE SCALING

Feature scaling is a crucial preprocessing step in machine learning that involves adjusting the range and distribution of input variables to ensure comparability across features. By normalizing or standardizing these inputs, models can achieve better performance, faster convergence during training, and reduced sensitivity to variations in feature magnitude. In this study, Min-Max normalization and LogNormal transformation were applied for data normalization, while Z-score normalization was used to standardize the data distribution.

Figure 6 illustrates the performance of 3 machine learning models: Random Forest, Gradient Boosting, and Logistic Regression, with and without various feature scaling techniques. The results indicated that Min-Max and Z-score normalization had negligible impact on the predictive accuracy of all 3 models. This suggested that these models, particularly the ensemble methods, were inherently robust to feature scaling. However, a notable improvement was observed in Logistic Regression when LogNormal normalization was applied. Accuracy surged from a low 32.81% (without scaling) to 79.69%, revealing the model's high sensitivity to input feature magnitude. Without appropriate normalization, Logistic Regression struggled to define an effective decision boundary, likely due to disproportionate feature influence and scale imbalances. On the other hand, the Gradient Boosting model consistently delivered the highest performance, achieving 81.25% accuracy and only 2 false negatives across all scaling scenarios. As an ensemble-based method that aggregated multiple weak learners, Gradient Boosting effectively captured complex feature interactions and was less affected by differences in feature scale.

These findings highlighted the importance of selecting suitable preprocessing techniques based on model characteristics. While complex models like Gradient Boosting exhibited resilience to raw data scales, simpler models such as Logistic Regression required careful normalization to perform optimally.

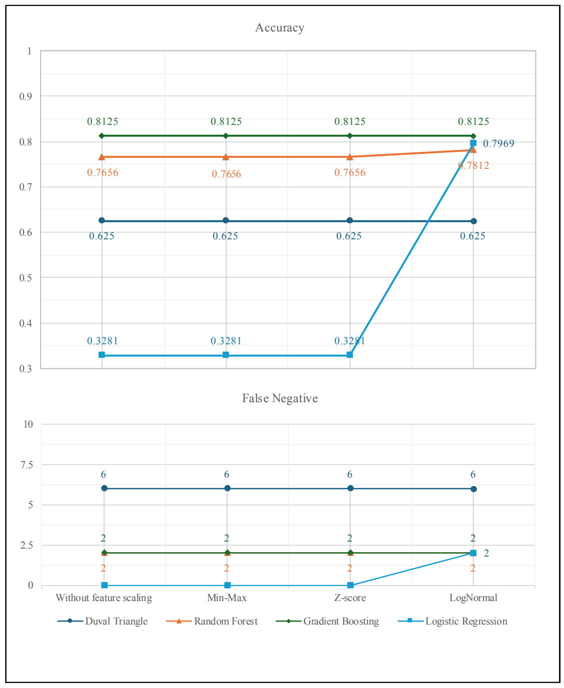


FIGURE 6. Model performance with and without feature scaling

VARIATION IN OUTPUT LABELS

This section intends to investigate the impact of varying the output labels. In the base case, the machine learning models classified the fault into 1 of the 6 output fault classes. In this section, the performance of the machine learning models was compared with another scenario using 4 output fault types. In both study scenarios, the train-test ratio used was 90:10 and no other pre-processing technique was applied.

In the first scenario, the models were trained to classify the transformers into 6 output fault classes, namely D1, D2, T1/T2, T3, PD and no fault. With data split for training and testing of 90:10, 575 datasets were used for training and 64 for testing. Table 4 shows the distribution of fault classes in the testing datasets for 6 fault classes. The distribution was skewed and unequal, with an average of 16.6%, a median of 17.2%, and a standard deviation of 5.8.

The evaluation results of 3 distinct machine learning models are presented in Table 5, together with the performance of Duval Triangle. From Table 5, Gradient Boosting outperformed other models with an accuracy of 81.25%, followed by Random Forest at 76.56%. Both models exhibited 2 false negatives, demonstrating their superiority over the Duval Triangle method, which had an accuracy of 62.5% and 6 unnoticed faults. Logistic Regression, while having the lowest accuracy at 32.81%, did not produce any unnoticed faults.

The confusion matrix for Gradient Boosting in predicting 6 fault classes is shown in Figure 7. The prediction accuracy for all classes exceeded 80%. The highest accuracy was for predicting No fault (90.9%), with 1 out of 11 healthy transformers misclassified as having a T1/T2 fault, resulting in a false positive. This was followed by T3 fault (87.5%), where 1 out of 8 T3 fault was incorrectly labeled as D2 fault. The prediction accuracy for D2 fault was 85.7%, with 2 out of 14 faults predicted as D1 fault. The model was observed to have the lowest success rate in predicting D1 fault with 72.7% success rate. However, it is worth noting that the wrong predictions were within the same fault types, i.e. discharge fault, where 3 out of 11 D1 fault (27.3%) were predicted as D2 fault.

A significant concern with this model was the false negatives, where 13.3% (2 out of 12) of T1/T2 faults were incorrectly predicted as No Fault. Figure 8 presents the 2 datasets where the model failed to detect a fault, together with the threshold for fault detection. A deep dive into these false negatives revealed that the dissolved gas concentrations for all 5 types of gases were considerably low and did not meet the gas threshold for fault detection. The threshold was determined based on the midpoint of the ranges of typical gas concentration values observed in power transformer in (IEC TC10, 2022), except for C_2H_2. This was because C_2H_2 is an indication of sparking discharges and arcs, which is not normally present. As such, a low concentration would require attention and further investigation. Both the Gradient Boosting model and conventional plotting of Duval Triangle diagnosed these transformers as "No Fault," whereas, in reality, T1/T2 faults were observed when the transformers were de-energized and inspected.

The Gradient Boosting model failed to detect faults in 2 datasets because the dissolved gas concentrations were too low to meet the detection thresholds. There could be several possible reasons for the discrepancies: (1) Fault Progression, where it was possible that the faults in the transformers were in an early stage during the oil sampling, making them less apparent. Over time, these faults might have progressed and became more evident during subsequent inspections; (2) Sensitivity of diagnostic method, as different diagnostic tools and techniques had varying levels of sensitivity. Duval Triangle might not have been sensitive enough to detect the T1/T2 faults, especially at early stage. Duval Triangle 1 is known to be more sensitive to ethylene and T3 fault; (3) External Factors, where elements such as the presence of Dibenzyl Disulfide (DBDS) in the oil, the age of the equipment, the manufacturers of both the equipment and the oil, environmental conditions (e.g., temperature, humidity), and operational loading could introduce variability into the diagnostic process, potentially resulting in erroneous diagnoses.

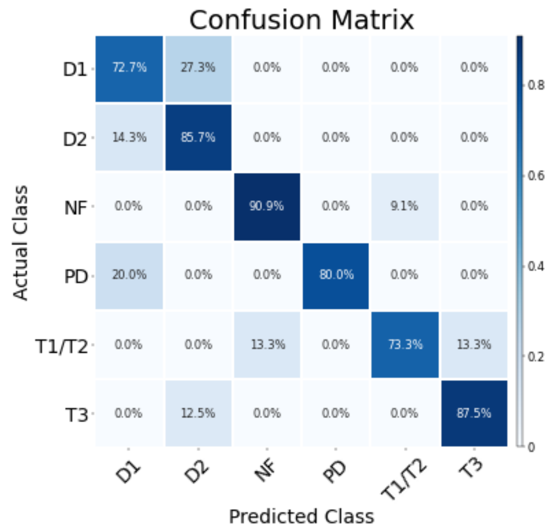


FIGURE 7. Confusion Matrix for Gradient Boosting Model (6 Output Fault Classes)

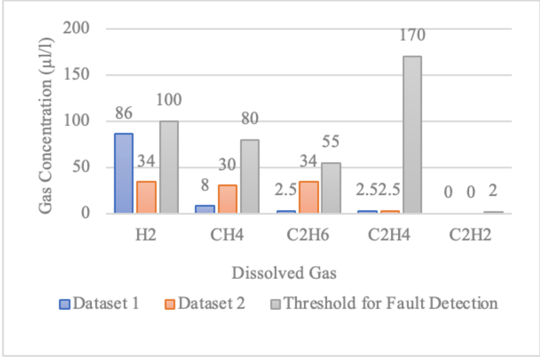


FIGURE 8. Datasets of False Negatives Predictions

Next, the models were trained to classify the transformers into 4 output fault types, namely discharge fault (D1/D2), thermal fault (T1/T2/T3), partial discharge (PD) and no fault. Similarly, the data split of training and testing was 90:10. 611 datasets were used for training and

68 for testing. The number of datasets slightly increased from the previous scenario as the full ground truth datasets include 38 rows classified as T1/T2/T3 and 2 rows labeled as D1/D2. These 40 rows were excluded from the train/test database for 6 fault classes, as they cannot be properly classified for supervised learning. However, these rows were usable for 4 fault types. Table 6 presents the distribution of fault classes in the testing datasets for 4 fault types. The distribution was skewed and unequal, with an average of 25.0%, a median of 26.5%, and a standard deviation of 15.7.

Table 7 presents the evaluation results of 3 machine learning models alongside the performance of the Duval Triangle method. By reducing the output labels from 6 to 4, the accuracy of the conventional Duval Triangle method increased to 79.41%. Among the machine learning models, Gradient Boosting achieved the highest accuracy at 95.59%, followed by Random Forest at 94.12%. Although Logistic Regression remained the least accurate model, its prediction accuracy improved to 61.76%. Importantly, none of the machine learning models exhibited any unnoticed faults.

The confusion matrix for the Gradient Boosting model is shown in Figure 9. The prediction accuracy for all classes exceeded 80%. The highest accuracy was achieved in predicting T1/T2/T3 faults (100%), with all thermal faults correctly classified. This was followed by D1/D2 (96%), where 1 out of 25 transformers with discharge fault was misdiagnosed as a thermal fault. NF followed with an accuracy of 90.9%, where 1 healthy transformer was misdiagnosed as having a thermal fault. The model had the lowest success rate in predicting the PD fault, with 20% of these instances misclassified as D1 faults. This was likely due to the limited amount of training and testing data available for the PD fault type. As shown in Figure 10, only 7% of the total datasets were classified as PD faults. Initially, fault classes such as T3, T1/T2, and D1 also had low percentages of datasets. However, by combining all thermal faults (T1, T2, T3) into a single fault type and all discharge faults (D1, D2) into another, the data sizes for these two fault types had increased significantly. This increase has proven sufficient for the effective training and testing of the model. A significant achievement of this model was the absence of undetected faults or false negatives. All faults were successfully detected, with 65 correct predictions and only 3 misclassifications. There was a false positive where a healthy transformer was predicted to have thermal fault. While reducing the number of classification labels from 6 to 4 led to an overall increase in model accuracy, this improvement came at the expense of diagnostic granularity. The simplified labeling scheme allowed the model to broadly classify faults as either thermal or discharge-related, improving clarity and reducing classification ambiguity. However, this approach eliminated the ability to assess the severity levels within each fault type, such as distinguishing between low and high-energy discharge events or varying thermal degradation stages. Such specific classification is often critical in predictive maintenance scenarios, where understanding the progression or urgency of a fault can significantly influence operational decisions. As a result, while the streamlined label set enhanced general classification performance, it limited the model's ability to provide detailed, actionable insights for more targeted maintenance planning.

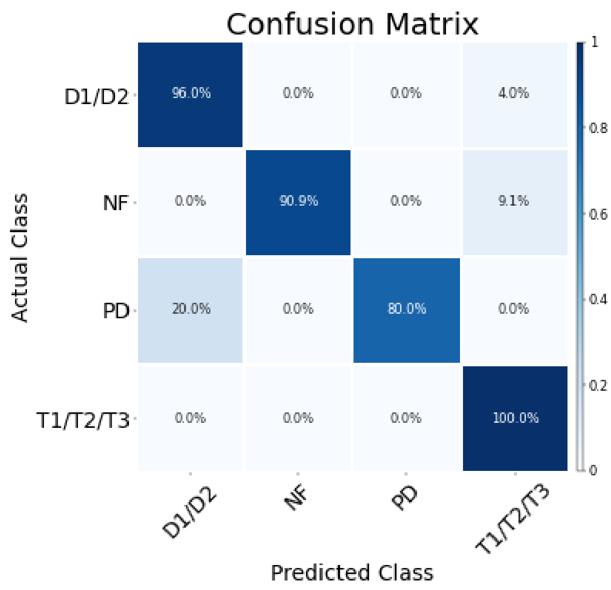


FIGURE 9. Confusion Matrix for Gradient Boosting Model (Grouped into Four Output Fault Types)

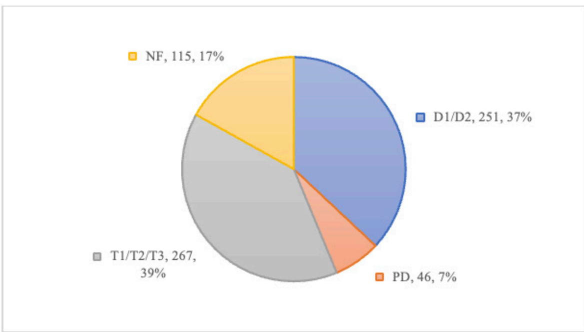


FIGURE 10. Distribution of Fault Classes for Ground Truth Datasets (4 Fault Types)

TABLE 4. Distribution of Fault Classes for Testing Dataset (6 Fault Classes)

Fault Class	Total Samples	Percentage (%)	Prediction	
			Correct	Wrong
D1	11	17.8	8	3
D2	14	21.9	12	2
PD	5	7.8	4	1
T1/T2	15	23.4	11	4
T3	8	12.5	7	1
NF	11	17.2	10	1
Total	64	100	52	12

TABLE 5. Comparison of Prediction Accuracy and False Negatives (6 Fault Classes)

Fault Class	Total Samples	Prediction		Accuracy (%)	False Negatives
		Correct	Wrong		
Duval Triangle	64	40	24	62.5	6
Random Forest	64	49	15	76.56	2
Gradient Boosting	64	52	12	81.25	2
Logistic Regression	64	21	43	32.81	0

TABLE 6. Distribution of Fault Classes for Testing Dataset (4 Fault Types)

Fault Class	Total Samples	Percentage (%)	Prediction	
			Correct	Wrong
D1/D2	25	36.8	24	1
PD	5	7.4	4	1
T1/T2/T3	27	39.7	27	0
NF	11	16.2	10	1
Total	68	100	65	3

TABLE 7. Comparison of Prediction Accuracy and False Negatives (4 Fault Types)

Fault Class	Total Samples	Prediction		Accuracy (%)	False Negatives
		Correct	Wrong		
Duval Triangle	68	54	14	79.41	4
Random Forest	68	64	4	94.12	0
Gradient Boosting	68	65	3	95.59	0
Logistic Regression	68	42	26	61.76	0

FIELD TRIALS

A field trial involves deploying the machine learning model in a real-world setting and validating its performance using actual operational data that includes known fault conditions. This step is essential for evaluating the model's practicality, robustness, and real-time inference capability, as well as for testing the ease with which the model can be retrained and improved when new data becomes available.

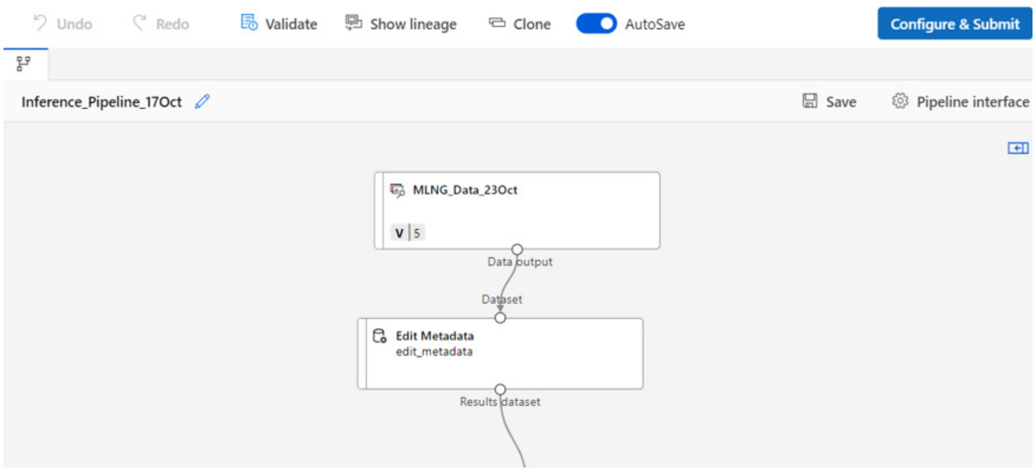
For this study, the best-performing Gradient Boosting model, configured with a 90:10 train-test split, and without feature scaling, was selected for deployment in field conditions. The field trials were conducted using DGA datasets obtained from 4 failed transformers located at different operational sites of the company. The datasets were fed into the inference pipeline depicted in Figure 3. Figure 11 presents the user interface and steps in Azure Machine Learning to make predictions with new data via the inference pipeline. First, open the inference pipeline and select the datasets of the 4 transformers to be used for prediction. Then, run the inference pipeline using the

Gradient Boosting model. Once the computation is completed, download the prediction results.

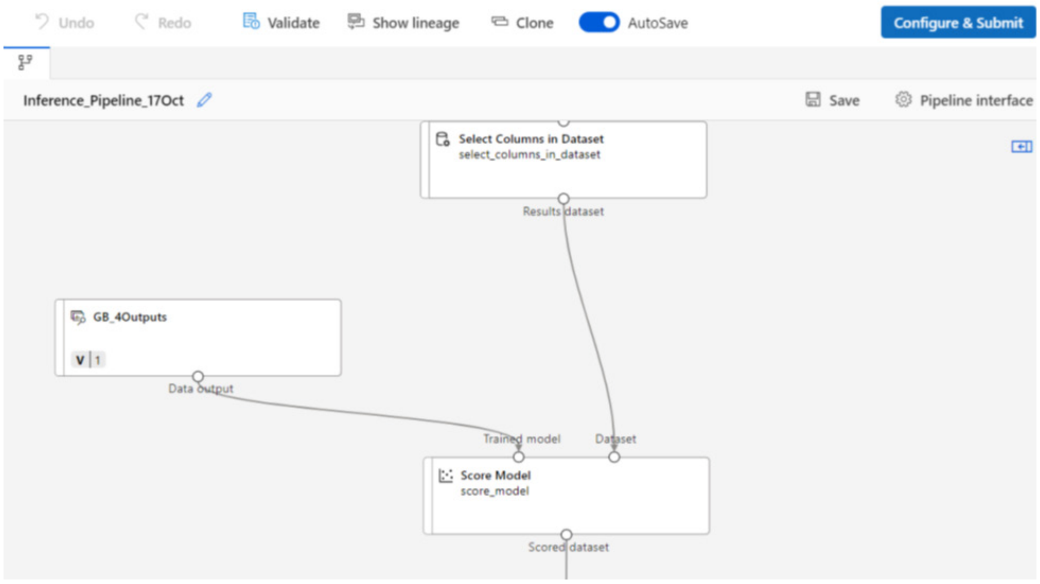
Table 8 provides the details of DGA inputs, prediction results of the Gradient Boosting model and the corresponding forensic findings. For benchmarking, fault predictions based on the Duval Triangle method were also included. The comparison between the Gradient Boosting model predictions and the forensic findings revealed a high level of accuracy. The model's predictions were consistent with the actual fault conditions observed during the inspections. For example, the model's prediction of a high-temperature thermal fault (T3) for TR-1 was confirmed by the presence of burnt connections and carbon particles in the transformer. Similarly, the prediction of a high-energy discharge (D2) for TR-2 was validated by the discovery of arcing at the copper winding. The model's predictions were also consistent with the popular Duval Triangle diagnosis, while eliminating the extra step required by Duval Triangle to first identify if a fault exists in the closed system where no area is designated as normal or no-fault.

The successful field trials highlighted the model's significance for integration into real-world transformer health monitoring systems. These trials have demonstrated the model's capability to provide accurate and timely diagnostics, which are crucial for the proactive maintenance of transformers. By enabling early detection of potential issues, the model facilitates timely interventions, thereby significantly reducing the likelihood of transformer failures.

The simplicity and ease of use of the inference pipeline in Azure Machine Learning further strengthened the adoption potential of the developed ML models in the real-world setting.



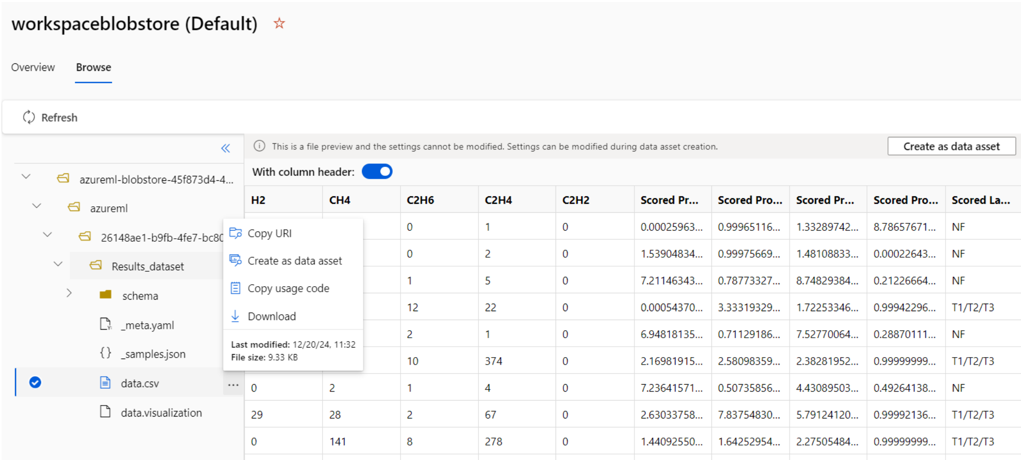
(a) Open the inference pipeline and select data to be used for prediction



(b) Run the inference pipeline with the selected model

continue...

...cont.



(c) Download the prediction result

FIGURE 11. Making fault prediction with dataset for field trials via the inference pipeline of Azure Machine Learning

TABLE 8. Field Trials Datasets and Results

Tag	Gas Concentration (µl/l)					Fault Class			Forensic Findings
	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	Ground Truth	GB	Duval	
TR-1	2241	9841	3724	10443	0	T3	T3	T3	TR-1 had a burnt connection at tap 4 connection 3 of the de-energized tap changer (DETC), with carbon particles present. This was suspected to be due to a loose connection, which resulted in high resistance and caused a high-temperature thermal fault. This coincides with a T3 fault
TR-2	85	62	8	64	86	D2	D2	D2	TR-2 showed signs of arcing at the copper winding between high voltage winding layers 20 and 21. Sulphur was detected on both the copper winding and the insulation paper at layer 20. It is suspected that copper sulphide weakened the insulating paper, resulting in electrical conductivity between the windings and ultimately causing a turn-to-turn fault, which fits the description of a D2 fault.
TR-3	689	89	104	8	248	D1	D1	D1	TR-3 exhibited visible burn marks on its paper insulation after the primary winding was removed. The secondary winding showed slight deformation and burn marks on the frame, indicating D1 fault
TR-4	156	508	129	424	0	T2	T1/ T2	T2	TR-4 experienced an internal flashover due to the reaction between corrosive sulphur and a hotspot at a loose connection in the red phase of the de-energized tap changer (DETC). This resulted in a high-temperature thermal fault and the formation of carbon particles at the DETC, signifying T2 fault.

CONCLUSION

This paper demonstrated a significant advancement in transformer fault diagnosis by introducing a unified no-code machine learning framework that enables field engineers to develop, deploy, and validate diagnostic models without programming expertise. The proposed approach addressed critical industry needs, including the ability to identify no-fault conditions, minimize false negatives, simple implementation and maintain diagnostic robustness under real operating conditions.

The experimental results showed the effectiveness of the proposed Gradient Boosting model. Under the optimized 90:10 train–test ratio, the model achieved 81.25% accuracy with 6 fault classes, outperforming the Duval Triangle method by nearly 19% while reducing unnoticed faults from 6 to 2. When the fault classes were consolidated into 4 key fault types, the model delivered a remarkable accuracy of 95.59% accuracy with zero false negative, demonstrating its dependability in identifying all actual fault cases. These performance gains highlight the model's capability not only to exceed conventional diagnostic accuracy but also to address high-impact failure risks that traditional methods frequently overlook.

The significance of this research is further reinforced by the successful field trials, in which the machine learning model correctly diagnosed all 4 real transformer failures, with predictions fully aligned with forensic inspections. Such consistency under real-world conditions validates the practical applicability, reliability, and readiness of the proposed approach for integration into operational maintenance workflows.

By empowering field engineers to build and deploy models without deep technical expertise, this research bridged the gap between advanced analytics and day-to-day engineering practice. The no-code methodology allows field engineers to apply their domain expertise directly into model development, enabling rapid retraining, continuous improvement, and scalable deployment across fleets. The resulting model supports timely, accurate, and interpretable diagnostics, enabling more informed maintenance decisions and contributing to the overall reliability of power systems.

Future research could focus on acquiring larger, high-quality, and diverse datasets with validated ground truth to enhance the reliability and generalizability of machine learning models. With access to richer data, the models can be further refined to detect more granular fault types, such as arching faults occurring in paper insulation versus those in transformer oil. Additionally, addressing data imbalance remains a critical challenge. Advanced resampling techniques, such as Adaptive Synthetic Sampling (ADASYN) and hybrid resampling strategies, could be investigated to improve model performance on minority fault classes and reduce bias in predictions.

ACKNOWLEDGEMENT

This work was supported in part by Project, Technology & Health, Safety, Security, Environment Division, PETRONAS, Malaysia.

DECLARATION OF COMPETING INTEREST

None.

REFERENCES

Abdo, A., H. Liu, H. Zhang, J. Guo, Q. Li. 2021. A new model of faults classification in power transformers based on data optimization method. *Electric Power Systems Research* 200. https://doi.org/10.1016/j.epsr.2021.107446.

Al-Katheri, H.H., M.F.M. Yousof, H.A. Illias, M.A. Talib. 2021. Classification of fault and stray gassing in transformer oil using svm, nb and knn algorithms. *Proceedings of the IEEE International Conference on Properties and Applications of Dielectric Materials*. Institute of Electrical and Electronics Engineers Inc. pp. 326–329. https://doi.org/10.1109/ICPADM49635.2021.9493941.

A.S. Monzaid. 2018. UK GT F main transformer buchholz tripped (11th jan 18, 15.22) rep. inv/oeims/pgb/2018/0005/1. Paka, Terengganu.

Bharath, L., P.A. Vahini, T. Sukumar, M.S. Kumar, S.M. Krishnan, B. Vigneshwaran. 2022. Recognition of power transformer fault based on dissolved gas analysis using fuzzy logic approach. *4th International Conference on Circuits, Control, Communication and Computing, I4C 2022*. Institute of Electrical and Electronics Engineers Inc. pp. 113–117. https://doi.org/10.1109/I4C57141.2022.10057869.

Das, S., A. Paramane, S. Chatterjee, U.M. Rao. 2023. Accurate identification of transformer faults from dissolved gas data using recursive feature elimination method. *IEEE Transactions on Dielectrics and Electrical Insulation* 30(1): 466–473. https://doi.org/10.1109/TDEI.2022.3215936.

Data Science Team. 2020. Azure ml studio-overview. citizen analytics (level 3). Kuala Lumpur, Malaysia.

Demirci, M., H. Gözde, M.C. Taplamacioglu. 2023. Improvement of power transformer fault diagnosis by using sequential kalman filter sensor fusion. *International Journal of Electrical Power and Energy Systems* 149. https://doi.org/10.1016/j.ijepes.2023.109038.

Duval, M., A. DePablo. 2001. Interpretation of gas-in-oil analysis using new iec publication 60599 and iec tc 10 databases. *IEEE Electrical Insulation Magazine* 17(2): 31–41. https://doi.org/10.1109/57.917529.

E. Abbasi, O. P. Malik. 2016. Failure rate estimation of power transformers using inspection data. *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. https://doi.org/10.1109/PMAPS.2016.7764150.

Fonseca, L.M., J.P. Domingues, A.M. Dima. 2020. Mapping the sustainable development goals relationships. *Sustainability* 12(8): 3359. https://doi.org/10.3390/SU12083359.

Gao, Z., X. Wang, F. Yi, W. Xu, M. Zhu. 2023. Transformer dga diagnostic method based on fuzzy logic function method. *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. IEEE. pp. 1639–1642. https://doi.org/10.1109/ITAIC58329.2023.10408761.

Ghoneim, S.S.M., K. Mahmoud, M. Lehtonen, M.M.F. Darwish. 2021. Enhancing diagnostic accuracy of transformer faults using teaching-learning-based optimization. *IEEE Access* 9: 30817–30832. https://doi.org/10.1109/ACCESS.2021.3060288.

Hafiz, M.I.A.H., E.D.S.K. Azmi, M.I.N.N. Azlie, M.H. Zaki, M.A.H. Mazilan, M.S.M. Mazuki, et al. 2025. Application of artificial intelligence in electronics and semiconductor industries. *Jurnal Kejuruteraan* 37(7): 3245–3253. https://doi.org/10.17576/jkukm-2025-37(7)-11.

Hooshmand, R.A., M. Parastegari, Z. Forghani. 2012. Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers. *IEEE Electrical Insulation Magazine* 28(5): 32–42. https://doi.org/10.1109/MEI.2012.6268440.

Hydroelectric Research and Technical Services Group. 2005. Transformers: basics, maintenance, and diagnostics. U.S. Department of the Interior - Bureau of Reclamation, Denver, Colorado.

IEC TC10. 2022. Mineral oil-filled electrical equipment in service – guidance on the interpretation of dissolved and free gasses analysis, iec 60599.

IEEE Power & Energy Society. 2019. IEEE guide for the interpretation of gases generated in mineral oil-immersed transformers, c57.104-2019. IEEE.

Jan, B., O. Rahman, S. Parveen, S.A. Khan. 2023. Multi-stage binary classification technique for incipient fault diagnosis of oil immersed power transformers based on anfis. *2023 International Conference on Power, Instrumentation, Energy and Control, PIECON 2023*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/PIECON56912.2023.10085900.

Jin, Y., H. Wu, J. Zheng, J. Zhang, Z. Liu. 2023. Power transformer fault diagnosis based on improved bp neural network. *Electronics* 12(16): 3526. https://doi.org/10.3390/electronics12163526.

Jing, Z., H. Zhang, P. Yang. 2023. Analysis method of dissolved gas in power transformer oil based on fuzzy reinforcement learning. Institute of Electrical and Electronics Engineers (IEEE). pp. 148–152. https://doi.org/10.1109/icpet59380.2023.10367721.

Kaur, K., D. Bhalla, J. Singh. 2023. Fault diagnosis for oil immersed transformer using certainty factor. *IEEE Transactions on Dielectrics and Electrical Insulation*. https://doi.org/10.1109/TDEI.2023.3307513.

Malik, I.M., A. Sharma, R.T. Naayagi. 2023. A comprehensive and practical method for transformer fault analysis with historical data trend using fuzzy logic. *IEEE Transactions on Dielectrics and Electrical Insulation* 30(5): 2277–2284. https://doi.org/10.1109/TDEI.2023.3286795.

Malla, M.A., O.H. Al-Beak, D.M. Hameed, M.M.M. Al-Hatab, R.R.O. Al-Nima, M.S. Jarjees, et al. 2024. Adopting machine learning to automatically identify a suitable surgery type for refractive error patients. *Jurnal Kejuruteraan* 36(4): 1749–1757. https://doi.org/10.17576/jkukm-2024-36(4)-37.

Manoj, T., C. Ranga, S.S.M. Ghoneim, U.M. Rao, S.A.M. Abdelwahab. 2023. Alternate and effective dissolved gas interpretation to understand the transformer incipient faults. *IEEE Transactions on Dielectrics and Electrical Insulation* 30(3): 1231–1239. https://doi.org/10.1109/TDEI.2023.3237795.

Menezes, A.G.C., M.M. Araujo, O.M. Almeida, F.R. Barbosa, A.P.S. Braga. 2022. Induction of decision trees to diagnose incipient faults in power transformers. *IEEE Transactions on Dielectrics and Electrical Insulation* 29(1): 279–286. https://doi.org/10.1109/TDEI.2022.3148453.

Misbahulmunir, S., V.K. Ramachandaramurthy, Y.H.M.D. Thayoob. 2020. Improved self-organizing map clustering of power transformer dissolved gas analysis using inputs pre-processing. *IEEE Access* 8: 71798–71811. https://doi.org/10.1109/ACCESS.2020.2986726.

Mohamed, R., S. Abd Halim, S. Sahrani, M.A. Talib. 2024. Accelerated heating paper pressboard and transformer oil for transformer aging assessment. *Jurnal Kejuruteraan* 36(5): 1875–1883. https://doi.org/10.17576/jkukm-2024-36(5)-08.

Patil, S.D., A.J. Patil, M. Dharme, R.K. Jarial. 2023. DGA based ensemble learning approach for power transformer fault diagnosis. *2023 International Conference in Advances in Power, Signal, and Information Technology, APSIT 2023*. Institute of Electrical and Electronics Engineers Inc. pp. 722–727. https://doi.org/10.1109/APSIT58554.2023.10201755.

Pattanadech, N., K. Sasomponsawatline, J. Siriworachanyadee, W. Angsusatra. 2019. The conformity of dga interpretation techniques: experience from transformer 132 units. *Proceedings - IEEE International Conference on Dielectric Liquids*. IEEE Computer Society. https://doi.org/10.1109/ICDL.2019.8796588.

Paul, D., A.K. Goswami, R.L. Chetri, R. Roy, P. Sen. 2022. Bayesian optimization-based gradient boosting method of fault detection in oil-immersed transformer and reactors. *IEEE Transactions on Industry Applications*. Institute of Electrical and Electronics Engineers Inc. pp. 1910–1919. https://doi.org/10.1109/TIA.2021.3134140.

Poonnoy, N., C. Suwanasri, T. Suwanasri. 2022. Neural network approach to dissolved gas analysis for fault analysis in power transformers. *Proceedings of the 2022 International Electrical Engineering Congress, iEECON 2022*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/iEECON53204.2022.9741595.

Prasojo, R.A., H.F.W. Azizi, I. Kurniawan, Suwarno, H. Sutikno, I. Ridzki. 2023. Naïve bayes model for combining results from multi-methods fault identification of power transformer. *Proceedings of 2023 4th International Conference on High Voltage Engineering and Power Systems, ICHVEPS 2023*. Institute of Electrical and Electronics Engineers Inc. pp. 199–203. https://doi.org/10.1109/ICHVEPS58902.2023.10257469.

Raghuraman, R., A. Darvishi. 2022. Detecting transformer fault types from dissolved gas analysis data using machine learning techniques. *Proceedings of the 2022 IEEE Dallas Circuits and Systems Conference, DCAS 2022*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/DCAS53974.2022.9845611.

Rampersad, R.M., S. Bahadoorsingh, C. Sharma. 2018. Multifactorial frameworks modelling linkages of power transformer failure modes.

Rao, U.M., I. Fofana, K.N.V.P.S. Rajesh, P. Picher. 2021. Identification and application of machine learning algorithms for transformer dissolved gas analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* 28(5): 1828–1835. https://doi.org/10.1109/TDEI.2021.009770.

Ru, F., L. Zhang, X. Yang, H. Zou, Y. Lu, X. Xu. 2022. Transformer fault diagnosis based on dissolved gas analysis in oil and ensemble learning. *Proceedings of 2022 IEEE 5th International Electrical and Energy Conference, CIEEC 2022*. Institute of Electrical and Electronics Engineers Inc. pp. 3174–3178. https://doi.org/10.1109/CIEEC54735.2022.9846076.

Saha, C., N. Baruah, S.K. Nayak. 2021. Implementation of self-organizing map and logistic regression in dissolved gas analysis of transformer oils. *Proceedings of the IEEE International Conference on Properties and Applications of Dielectric Materials*. Institute of Electrical and Electronics Engineers Inc. pp. 131–134. https://doi.org/10.1109/ICPADM49635.2021.9493899.

Saravanan, D., A. Hasan, A. Singh, H. Mansoor, R.N. Shaw. 2020. Fault prediction of transformer using machine learning and dga. *2020 IEEE International Conference on Computing, Power and Communication Technologies, GUCON 2020*. Institute of Electrical and Electronics Engineers Inc. pp. 1–5. https://doi.org/10.1109/GUCON48875.2020.9231086.

Subalakshmi, J., J.J. Janet, J.J.S. Lakshmi, T. Sukumar, B. Vigneshwaran. 2022. Recognition of power transformer defect identification based on dissolved gas analysis using support-vector machine approach. *6th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2022 - Proceedings*. Institute of Electrical and Electronics Engineers Inc. pp. 104–108. https://doi.org/10.1109/ICECA55336.2022.10009258.

Taha, I.B.M., S. Ibrahim, D.E.A. Mansour. 2021. Power transformer fault diagnosis based on dga using a convolutional neural network with noise in measurements. *IEEE Access* 9: 111162–111170. https://doi.org/10.1109/ACCESS.2021.3102415.

Utmal, Dr.M. 2021. Machine learning its applications, challenges & tools: a review. *International Journal of Computer Science and Mobile Computing* 10(3): 32–38. https://doi.org/10.47760/ijcsmc.2021.v10i03.004.

Verma, N., A.K. Chandel. 2023. Incipient fault diagnosis of power transformer based on duval pentagon using medium neural network. Institute of Electrical and Electronics Engineers (IEEE). pp. 1–5. https://doi.org/10.1109/stpes54845.2022.10006647.

Wang, L., T. Littler, X. Liu. 2021. Gaussian process multi-class classification for transformer fault diagnosis using dissolved gas analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* 28(5): 1703–1712. https://doi.org/10.1109/TDEI.2021.009470.

Wang, T., J. Yang, J. Shao, P. Wu, Y. Ma, J. Liang. 2023. Improvement of power transformer faults diagnosis using random search with classification and regression tree. *2023 5th Asia Energy and Electrical Engineering Symposium, AEEES 2023*. Institute of Electrical and

Electronics Engineers Inc. pp. 653–658. https://doi.org/10.1109/AEEES56888.2023.10114139.

Yaqin, E.N., U. Khayam. 2021. Improvement of application cognitive artificial intelligence based on doernenburg ratio method for dissolved gas analysis interpretation. *Proceedings of the International Conference on Electrical Engineering and Informatics*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICEEI52609.2021.9611122.

Zhang, Y., Y. Tang, Y. Liu, Z. Liang. 2022. Fault diagnosis of transformer using artificial intelligence: A review. *Frontiers in Energy Research* 10. https://doi.org/10.3389/fenrg.2022.1006474.