IJIE

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

Vol. 18 No. 1 (2026) 349-367
https://publisher.uthm.edu.my/ojs/index.php/ijie

Development and Performance Evaluation of Machine Learning Model for Fault Diagnosis of Mineral Oil-Filled Transformer Using No-Code Approach

Chee Ying Chan[1,2]*, Nursyarizal Mohd Nor[2], Nor Zaihar Yahaya[2]

[1] *Group Projects and Engineering, Projects, Engineering and Technology Division, PETRONAS, Level 20, Tower 3, KLCC, Kuala Lumpur, 50088, MALAYSIA*

[2] *Electrical & Electronics Engineering Department, Universiti Teknologi PETRONAS, Persiaran UTP, Seri Iskandar, Perak, 32610, MALAYSIA*

*Corresponding Author: chancheeying@petronas.com
DOI: https://doi.org/10.30880/ijie.2026.18.01.020

Article Info

Received: 30 March 2026
Accepted: 1 April 2026
Available online: 20 April 2026

Keywords

Classification, dissolved gas, duval triangle, fault diagnosis, machine learning, oil-filled, transformer

Abstract

Interpretation of dissolved gas composition is crucial in fault diagnosis of oil-filled transformer. Conventional interpretation methods have limitations in terms of accuracy and adaptability to new fault patterns, while artificial intelligence techniques applied in this field require extensive programming knowledge, data volumes and computational resources. This study presents the development and performance evaluation of machine learning models for transformer fault diagnosis using a no-code approach. Utilizing Microsoft Azure Machine Learning, trained on 639 ground truth datasets, the Gradient Boosting model consistently outperformed other models and Duval Triangle, achieving the highest accuracy of 81.25% with 2 false negatives. Experiments were conducted to optimize the model's performance, including varying train-test ratios, applying oversampling and outlier treatment. The model was validated through field trials on 4 transformers, successfully predicting faults for all transformers, consistent with forensic findings. Feature importance analysis results were consitent with established dissolved gas–fault relationships, which provided transparency into the model's fault classification and supported user confidence. This research highlighted the practical applicability and robustness of the machine learning model, offering a singular diagnostic method for field engineers to distinguish between healthy and faulty transformer, and accurately identify the fault type and severity. Leveraging the no-code approach, field engineers, being the subject matter experts, could harvest the advantages of artificial intelligence and impart their knowledge and data, without vast computing knowledge and resources.

1. Introduction

The 21st century is marked by an ever-increasing reliance on electricity, driving economic development, technological innovation, and the improvement of the quality of life worldwide. The availability of reliable and sustainable electricity is essential for human well-being and plays a paramount role in achieving the United Nations' Sustainable Development Goals (SDGs) [1]. Transformers serve as the backbone of electricity distribution networks, making their dependable operation pivotal in realizing the vision of universal electricity access outlined in SDG 7. The integration of machine learning with fault diagnosis of transformer presents an innovative and efficient approach to transformer health monitoring. This approach not only enhances the availability of power distribution networks, but also aligns with the goals of SDG 9 by promoting resilient infrastructure and technological innovation. This present work explores the integration of machine learning techniques in transformer fault diagnosis based on dissolved gas composition, aiming to contribute to more efficient, dependable, and sustainable energy landscape. By leveraging the power of data analytics and automation, fault diagnosis accuracy can be enhanced, enable proactive maintenance strategies, and contribute to the achievement of SDGs.

1.1 Transformer Fault Diagnosis via Dissolved Gas Composition

Transformers are considered critical assets because of their complex internal construction and the significant cost associated with their manufacture [2]. Hence, transformer health assessment, in the form of a series of field tests are conducted regularly to identify potential problems. It ranges from routine visual inspection, online measurements and monitoring to minor and major shutdown maintenance and overhaul [3]. [4] proposed a range of maintenance actions, which include condition assessment actions, which include electrical testing, oil testing and inspections, and maintenance actions to improve condition, such as dry outs and oil treatment. However, transformers are subjected to multifactorial stresses which involve mechanical, chemical, electrical and environmental factors [5]. Hence, despite their rigorous maintenance regimes, transformers are susceptible to various faults, which can compromise their performance and lead to operational failures. Swift and efficient fault diagnosis of transformers is crucial for enhancing power grid reliability, preventing cascading failures, and ensuring overall resilience. Historical events like the 2003 North America blackout, the 2006 Europe interconnected grid blackout, and the 2009 Brazil blackout, highlight the importance of this [6]. It minimizes downtime and economic losses by enabling proactive maintenance and early detection of potential failures, reducing repair costs by 75% and mitigating revenue loss by 60% [7]. Additionally, it prevents catastrophic failures, reducing risks of severe disruptions and safety hazards. Effective fault diagnosis also optimizes maintenance strategies, lowers costs, and aligns sustainable practices by extending transformer longevity [8].

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 [5]. As electrical and thermal stresses accumulate, the transformer insulating materials will break down and release several different gases [9]. Dissolved Gas Analysis (DGA) is a very effective diagnostic strategy that is widely used to diagnose early faults in oil-immersed transformers [10]. DGA determines the types and quantities of gases generated, which depend on the nature and intensity of the fault. If the dissolved gasses are above typical concentration values and/or rates of increase, an actual fault in the transformer is probable.

Numerous interpretation methods (or "fault identification" method as per [11]) 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. The key gasses or ratio, diagnosable fault and limitations of each conventional interpretation methods for dissolved gas are discussed [11], [12], [13]. Key Gas Method is one of the main and simplest methods. Apart from that, in general, the conventional methods can be divided into 2 distinct types, i.e. the ratios method and the pictographic method [14]. The ratios method is based on pre-determined values of gas ratios to classify the fault. The most well-known ratios methods are elaborated in [11]. They are the IEC Ratios, Rogers Ratios and Doernenburg ratios methods. The pictographic method includes the Duval Triangle and Duval Pentagon, which classify faults based on relative percentages of gasses plotted on each side of the polygon and lines drawn across the polygon for each gas. The zone where the lines intersect indicates the fault type [13]. The ratio methods may result in "no interpretation" due to incomplete ratio ranges of the methods, while the pictographic method by Duval requires pre-determination if a transformer has issue as there is no fault-free area in the triangles and pentagons. Consequently, the diagnosis results are influenced more by engineers' knowledge and engagement than by strict mathematical computations. There is a significant probability of obtaining different conclusions when analyzing one sample using various techniques. These divergent conclusions may result in inaccuracies and errors in fault identification. Under extreme conditions, these errors could potentially lead to mishandling and unnecessary cost management [15].

Int. Journal of Integrated Engineering Vol. 18 No. 1 (2026) p. 349-367

Due to challenges associated with conventional methods, researchers have redirected their focus toward artificial intelligence (AI) techniques, leading to a revolutionary development in fault analysis with notable results characterized by high accuracy and efficiency [16]. Computational intelligence incorporates elements of adaptation, learning, evolution, and evidence to formulate intelligent conclusions that address the discrepancies found in outcomes from conventional methods [17].

1.2 Challenges and Motivation

Field engineers, from either technical services or maintenance department, need to review and analyze the transformer DGA results submitted by the laboratories. The resultant value of each dissolved gas is compared against standards and guidelines. If there is any abnormal reading, the whole result set needs to be interpreted properly. The engineers need to classify the fault type and severity in order to determine the next course of action. Typically, minor fault may require more frequent oil sampling and close monitoring, while major faults may require immediate shut down for inspection and rectification. From the fault classification, the location(s) of the fault can be narrowed down to facilitate monitoring and inspection.

Conventional interpretation methods raise persistent concerns regarding failure to diagnose and misdiagnosis. Expert judgment is crucial resulting in diverse interpretations and mitigations [13]. Conventional methods also exhibit inflexible standards, struggling to adapt to evolving transformer designs, materials, and issues. Their rigidity, set until formal intervention, hinders responsiveness to industry changes, necessitating tedious updates through organizations like CIGRE.

Despite the remarkable advancements in machine learning in recent years, several significant challenges persist [18]. A key challenge lies in the necessity for a substantial volume of high-quality data for effective training and validation. The necessity for domain expertise in selecting, tuning, and interpreting machine learning models further complicates the process, often requiring the involvement of specialized programmers and data scientists. This reliance on external expertise not only incurs additional costs but also extends project timelines, encompassing phases such as bidding, evaluation, contract award, requirements gathering, data collection, model development, and multiple review and validation sessions. These limitations hinder field engineers from independently developing, training, and validating machine learning models, despite being the subject matter experts with comprehensive knowledge and data. Consequently, the current approach results in inefficiencies and delays, as engineers must rely on external resources to perform tasks that could otherwise be managed internally.

Additionally, the computational complexity and resource requirements of certain machine learning algorithms may pose practical challenges, particularly in real-time fault diagnosis applications. As outlined by [19], Artificial Neural Networks (ANN) present challenges such as a high computational load, slow convergence speed, and susceptibility to local minima, necessitating ongoing research for optimization and enhancement. In comparison, Support Vector Machines (SVM) exhibit greater efficiency than ANN, yet challenges persist in determining kernel function parameters and regularization parameters, potentially resulting in misjudgments and omissions due to the non-divisibility of certain regions. Despite the strong generalization ability of deep learning models, they still lack comprehensive mathematical theoretical support.

In summary, there is absence of a singular diagnostic method for field engineers to distinguish between healthy and faulty transformer, and accurately identify the fault type and severity, due to limitations of conventional interpretation methods. It is also a challenge for field engineers to harness the advantages of artificial intelligent techniques without necessitating machine learning domain expertise, extensive data volume and computational resources.

1.3 Proposed Solution and Contribution

This research aims to develop a machine learning model using a no-code or low-code approach, which is capable of distinguishing healthy and faulty transformers and identify the fault type and severity, based on dissolved gas composition. Attempts will be made to optimize and assess the performance of the machine learning model using various preprocessing techniques and compare the performance with the conventional method in terms of accuracy and false negative. Lastly, performance of the machine learning model will be validated 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 reform and elevate transformer fault diagnosis for field engineers. By developing a single, unified method to differentiate between healthy and faulty transformers and accurately identify fault types and severity, this research addresses the critical need in the industry. The adoption of no-code or low-code approach in developing the machine learning model empowers field engineers to develop, train, and test the models themselves, without the need for extensive programming knowledge or reliance on specialized data scientists. This not only saves time and money but also allows engineers to incorporate their insights, expertise, and experiences into the model, leading to more accurate and contextually relevant diagnostics. The no-code or low-code approach also addresses the challenge of requiring significant computational resources. This research offers significant benefits to field engineers and industry by providing a more accurate, efficient, adaptable and accessible method for transformer fault diagnosis. The integration of machine learning techniques 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.

The rest of this paper is organized as follows: The Methodology section details the use of Microsoft Azure Machine Learning for developing multiclass classification models, including data collection, preprocessing, model training and evaluation, as well model deployment for fault prediction. The Results and Discussion section presents experimental results, analyzing model performance, the impact of preprocessing techniques, comparison with Duval Triangle, the results of field trials, the comparison with recent AI-based DGA diagnosis methods, the rankings of feature importance and limitations of the study. Finally, the Conclusion summarizes key findings, discusses implications, and suggests future research directions.

2. Methodology

A detailed and structured approach to developing and deploying machine learning models for fault diagnosis of transformers is presented in Figure 1. The detailed implementation of each step is presented in the subsequent sections.

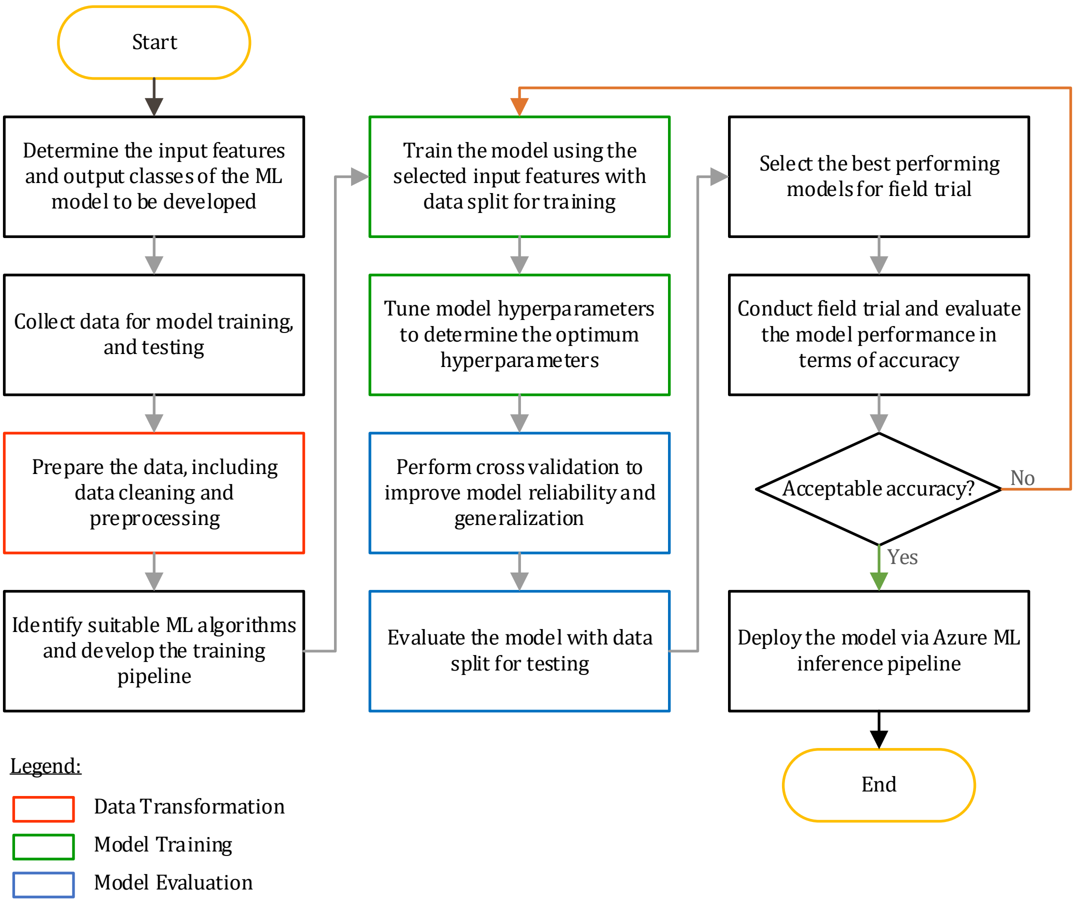


Fig. 1 *End-to-end workflow for developing and deploying machine learning model*

2.1 Input Features and Output Labels

Input features are the variables used to train a model and predict a target. In this work, the input features selected were derived from the concentrations of 5 key dissolved gases obtained through DGA. The specific gases and their significance are detailed in Table 1. The concentrations of these gases are measured in parts per million (ppm) and served as the primary data points in numerical features for the machine learning model.

Output labels or target variables are values that a model aims to predict based on the input features. In this work, the output label was the predicted transformer fault class based on the concentrations of the dissolved gases. The model classified the fault into 1 of the 6 categories, each representing a specific type of fault or a no-

Int. Journal of Integrated Engineering Vol. 18 No. 1 (2026) p. 349-367

fault condition. The output labels are presented in Table 2. T1 and T2 faults were combined into one fault class to accommodate the fault classes of the ground truth data available, specifically from [20].

Table 1 *Input features selected for the machine learning model*

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.

Table 2 *Output labels selected for the machine learning model*

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

2.2 Datasets

Datasets are required for training and testing of the machine learning model. In this work, the machine learning models were exclusively trained and tested 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 essential in supervised learning, where models are trained on input-output pairs. During training, the model learns patterns and relationships in the input data to make predictions. Ground truth data provides the correct answers or target values for the input data, allowing machine learning models to learn and improve their predictive capabilities.

A total of 639 ground truth datasets were sourced from [20], [21], [22], [23] and expert networks. These transformers were either healthy for a year since the oil samples were taken or have encountered faults during operation and have subsequently been taken out of service. They underwent visual inspection by experienced engineers and maintenance experts, leading to the clear identification of the fault classes. The distribution of fault classes is shown in Figure 2. The datasets were imbalance, for example, only 7% of the datasets were for PD, where there was 23% for T1/T2. There was an attempt to address this issue via "oversampling" - a data preprocessing technique, and the results were presented and discussed in the Section 3.3.

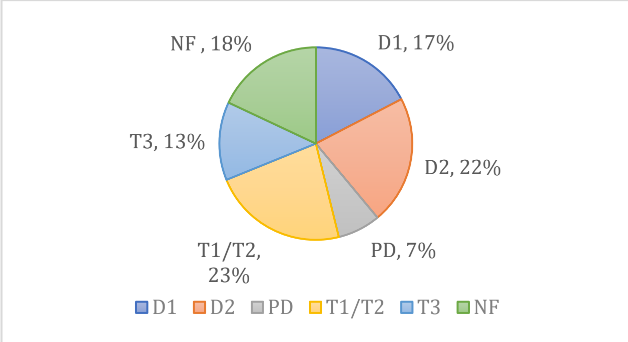


Fig. 2 *Distribution of fault classes for ground truth datasets*

2.3 Data Preparation

Data preparation involved transforming raw data into a suitable format for analysis and model training. There were 2 steps in data preparation, first was data cleaning and transformation, second was data preprocessing. 5 data cleaning and transformation components in Microsoft Azure Machine Learning were applied to ensure that the datasets meet the requirements of downstream components in the training pipeline in terms of data quality and usability. Table 3 explains the data cleaning and transformation performed. Data preprocessing prepared the data in a way that could enhance the performance of machine learning models. The data preprocessing applied in this work were splitting data, oversampling and outlier treatment. The details are discussed in Section 3.

Table 3 *Data cleaning and transformation*

No.	Component	Description
1	Edit metadata	Mark columns as features and indicate column that contains the target variable
2	Convert to dataset	Replace diagnosis column indicating T1 and T2 fault classes with T1/T2 fault class to be consistent with the selected output label
3	Apply SQL transformation	Remove rows containing T1/T2/T3, D1/D2 and DT as diagnosis in the datasets, as these fault classes are not selected as output labels for the machine learning model
4	Remove duplicate rows	Remove rows which are entirely identical to ensure that each row is unique
5	Clean missing data	Remove rows with missing values

2.4 Machine Learning Model Selection

This research utilized the Microsoft Azure Machine Learning in transformer fault diagnosis. Microsoft 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 [24]. A very important feature of Microsoft Azure Machine Learning is its no-code web user interface, allowing business domain experts to train models on their data without the need to write a single line of code. Multiclass classification was employed to identify transformer fault classes based on dissolved gas composition. Random Forest, Gradient Boosting, and Logistic Regression were selected due to their proven effectiveness in classification problems with tabular data.

2.5 Model Training

Training a model in Microsoft Azure Machine Learning began with setting up the workspace, which acted as a central hub for managing all machine learning resources. The next step involved preparing the data by uploading it to the workspace and ensuring it was clean and properly formatted for training. Once the data was ready, an experiment was created in Microsoft Azure Machine Learning. A compute target was then selected for running the pipeline steps. Each step could include data preprocessing, model training, and evaluation. The pipeline was then submitted to Microsoft Azure Machine Learning, initiating the training process. Throughout this process, progress could be monitored and logs viewed in Microsoft Azure Machine Learning to ensure everything runs smoothly.

The "Tune Model Hyperparameters" component in Azure Machine Learning was employed in this study to systematically optimize the hyperparameters of model performance. It identified the combination of hyperparameters that yields the best model performance. Table 4 presents the parameters selected to execute tuning. Random sweep was selected as the sweeping mode with a maximum of 200 runs. Accuracy is the metric selected for evaluating the classification models. The default value of 9666613 was applied to seed the random number generator.

Table 4 *Parameters for hyperparameter optimisation*

Parameter	Selection for Execution
Sweeping Mode	Random Sweep
Maximum Number of Runs on Random Sweep	200
Random Seed	9666613
Metric for Measuring Performance for Classification	Accuracy

2.6 Model Evaluation

The performance of the machine learning models was evaluated using several key metrics, including accuracy, false negatives, and the confusion matrix, with particular emphasis on false negatives due to their implications for transformer operational risk. Stratified k fold cross validation was applied to ensure robust performance assessment across imbalanced fault classes. The number of folds applied was 9 and the number of examples in each fold was 64.

2.7 Model Deployment

Once machine learning models were trained and tested, they were deployed to production for inference. Inference involved applying new input data to a machine learning model or pipeline to generate outputs, commonly referred to as "predictions”. Deployment and retraining of the machine learning model via inference and training pipelines in Microsoft Azure Machine Learning is illustrated in Figure 3.

Structured feedback mechanisms are critical to sustaining the long-term performance of AI systems. Regular performance assessments, combined with input from relevant stakeholders, facilitate the identification of necessary refinements and support continued optimal operation [25]. To continuously benchmark the performance of the machine learning model, dissolved gas composition data can be systematically uploaded into the Microsoft Azure Machine Learning inference pipeline. This data can be used to rapidly diagnose faults and allow for immediate comparison with conventional diagnostic tools such as the Duval Triangle. Furthermore, incorporating additional ground truth data into the Microsoft Azure Machine Learning training pipeline enables the continuous retraining and improvement of the model. By regularly updating the model with new and accurate data, the system adapts to evolving fault patterns and improves its diagnostic accuracy over time. This iterative learning process ensures that the model remains robust and reliable, thereby providing consistent and high-quality fault diagnosis for transformer maintenance.

For long-term deployment in industrial settings, it is essential to establish a robust model lifecycle plan. This includes implementing a periodic retraining schedule, such as quarterly or bi-annually, based on newly acquired DGA data to ensure the model adapts to evolving fault patterns. A data governance strategy should be enforced to maintain data quality, consistency, and traceability. Additionally, integrating drift detection mechanisms can help monitor model performance over time and trigger retraining when significant deviations in prediction accuracy or data distribution are detected. These practices ensure the model remains reliable, accurate, and aligned with operational realities.

To contextualize the ease of use and adoption potential of the developed models in Microsoft Azure Machine Learning, Figure 4 presents the user interface in Microsoft Azure Machine Learning to make prediction with new data via the inference pipeline.

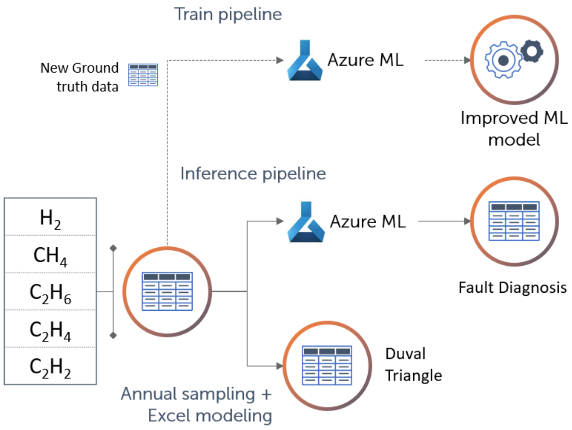
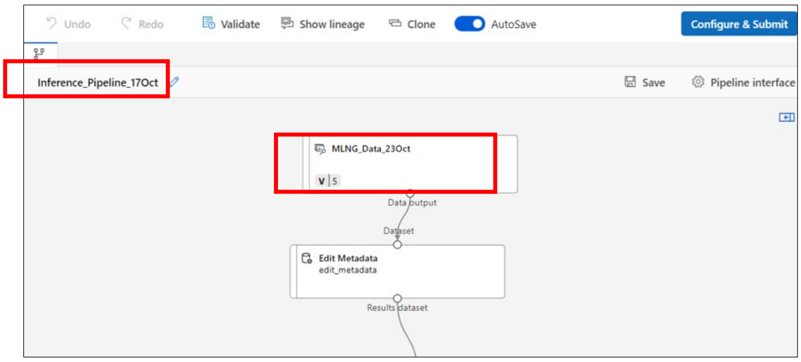
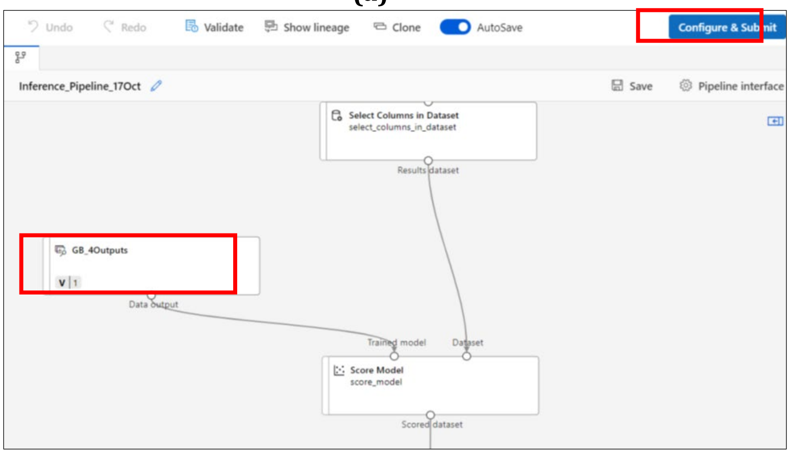


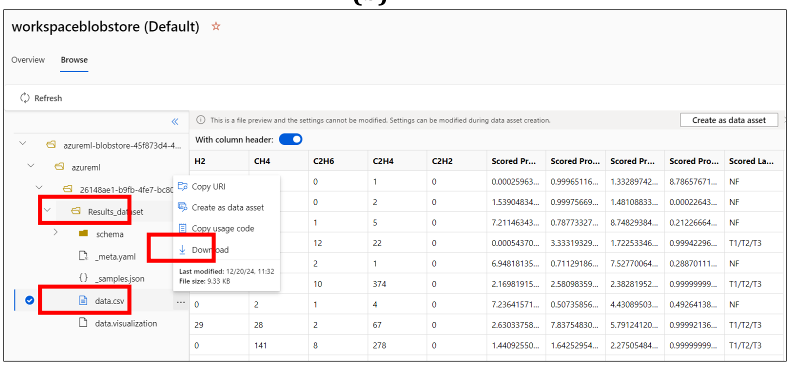
Fig. 3 *Deployment and retraining of the machine learning model*



(a)



(b)



(c)

Fig. 4 *Making prediction with new data via inference pipeline (a) Open the inference pipeline and select data to be used for prediction; (b) Run the inference pipeline with the selected model; (c) Download the prediction result*

Int. Journal of Integrated Engineering Vol. 18 No. 1 (2026) p. 349-367

3. Results and Discussion

This section presents a comprehensive analysis of the machine learning models developed for diagnosing faults in oil-filled transformers using Microsoft Azure Machine Learning. The performance of the base case and 3 experiments involving different data preprocessing methods are presented. The analysis of false negatives is discussed, followed by the results of field trials which validated the models' applicability in real-world conditions and the rankings of feature importance analysis which measured how much each input feature contributed to the model's predictive performance. Finally, the limitations encountered in the study are presented.

3.1 Base Case

Base case refers to a standard or reference scenario against which other scenarios or variations are compared. It serves as a control or benchmark to understand the impact of changes or interventions. Figure 5 presents the performance of the 3 selected machine learning models, i.e. Random Forest (RF), Gradient Boosting (GB), and Logistic Regression (LR), at base case. The train-test ratio used was 70:30 and no other pre-processing technique was applied.

The Gradient Boosting model demonstrated the highest accuracy of 78.53%, resulting in 5 false negatives. This performance was closely followed by the Random Forest model, which achieved an accuracy of 73.82% with 7 false negatives. The Logistic Regression model did not have any false negative, but the accuracy was very low at 35.08%. Duval Triangle's accuracy is 68.06% with 11 false negatives. 70:30 train-test ratio was selected for the base case as it was a common practice and industry standard in machine learning development due to its effectiveness in various applications. It provided a good starting point for most machine learning tasks, widely accepted and understood by practitioners. This ratio provided a balanced allocation of data between training and testing, where the model had enough data (70%) to learn the underlying patterns and relationships and sufficient data (30%) to evaluate the model's performance and generalization ability. This ratio could detect and mitigate overfitting, where the model learned the training data too well, including noise and outliers, which negatively impacted its performance on new, unseen data. However, while the 70:30 ratio was common, it was not a one-size-fits-all solution. In the subsequent section, the impact in varying the train-test ratio was investigated.

3.2 Varying Train-Test Ratio

The first preprocessing executed was splitting the data into training and testing using various ratios. The train-test ratios were varied to investigate the impact on the performance of the machine learning models. The choice of these split ratios was driven by the need to balance between having sufficient data for training the model and enough data for testing to ensure reliable evaluation of its performance. A higher proportion of training data (e.g. 90:10) allowed the model to learn more from the dataset, potentially improving its accuracy and robustness. However, it also reduced the amount of data available for testing, which could lead to less reliable performance evaluation. Conversely, a lower proportion of training data (e.g. 60:40) provided more data for testing, offering a more robust evaluation but potentially limiting the model's learning capability. By experimenting with different split ratios, the research aimed to identify the optimal balance that maximizes model performance while ensuring reliable evaluation.

Figure 6 presents the performance of the 3 machine learning models with different train-test ratios. Increasing the training data did not uniformly improve accuracy across all models. Different models had varying sensitivity to the amount of training data. For instance, ensemble methods like Random Forest and Gradient Boosting could leverage more data to improve their predictions, while Logistic Regression might not see the same level of improvement due to its simpler nature. Random Forest and Gradient Boosting were more complex and could handle larger datasets better, reducing overfitting. However, simpler models like Logistic Regression might not benefit as much from additional data if they were already capturing the underlying patterns well. Duval Triangle's performance consistently took the third place after Gradient Boosting and Random Forest. In this experiment, the best accuracy was achieved by the Gradient Boosting model when the train-test ratio was 90:10 with an accuracy of 81.25% and 2 false negatives. This train-test ratio was used for the next 2 experiments.

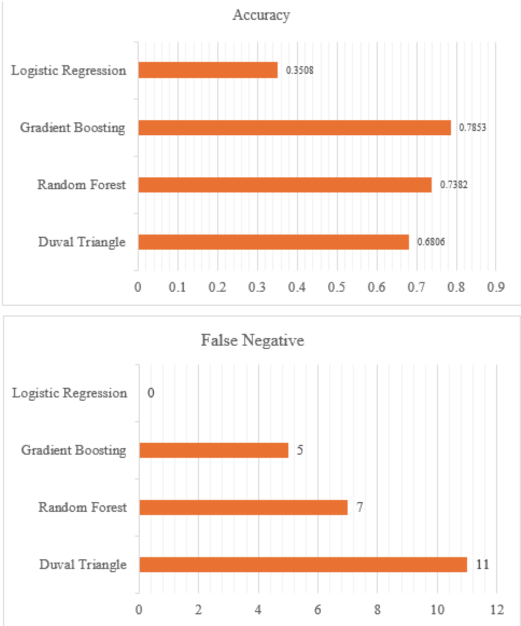


Fig. 5 *Model performance at base case (train:test 70:30)*

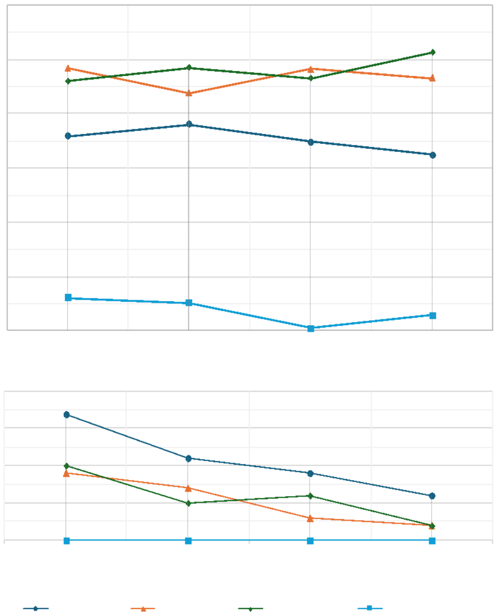


Fig. 6 *Model performance with varied train-test ratio*

3.3 Oversampling

The second preprocessing method executed was oversampling to treat imbalanced data which might lead to poor performance in predicting the minority classes. Oversampling increased the number of underrepresented cases in the datasets. “Synthetic Minority Oversampling Technique” (SMOTE) was applied in this work. It was a statistical method used to increase the number of cases in a dataset in a balanced manner. It increased the number of rare cases and returns a dataset that contains the original samples, as well as a number of synthetic minority samples equivalent to the number of observations from the imbalanced class.

Figure 7 compares the performance of the 3 machine learning models with and without oversampling. The accuracy gain was negative for all models, indicating decrease in accuracy with oversampling. However, for Gradient Boosting, there was a reduction in false negative from 2 to 1. Gradient Boosting trees grew in a sequential fashion, with each new growth repairing the mistakes of the older one. Gradient Boosted trees performed well in imbalanced circumstances because of their ability to concentrate on misclassified occurrences through sequential learning. Random Forests were made up of many decision trees that had been trained using arbitrary subsets of features and data. Random Forests improved generalization by reducing overfitting and strengthening robustness against imbalanced datasets by mixing numerous trees. As such, oversampling did not have significant impact on Gradient Boosting and Random Forest. Logistic regression assumed a linear relationship between the features and the log-odds of the outcome. When the oversampled data introduced non-linearities or noise, the model might struggle to fit the data accurately, resulting in worse performance. Duval Triangle’s performance was not impacted in this experiment as oversampling of only applied to the machine learning models. The best accuracy was achieved by the Gradient Boosting model, without oversampling, with an accuracy of 81.25% and 2 false negatives.

3.4 Outlier Treatment

An outlier is defined as an observation that lies an abnormal distance from the mass volume of the data. These outliers might create bias in the diagnosis. Pre-processing the data with outlier treatment involved applying the “ClipPeaks” function by specifying an upper boundary. Values greater than the boundary values were replaced with the specified threshold value. This method ensured that extreme values are moderated without completely removing the outliers, thereby retaining the essential variability in the data. The threshold values for each gas type were determined with reference to [12], applying a factor of 20 to the upper range of typical gas concentration values observed in power transformers. The upper boundaries for each input feature can be referred in Figure 9 (after outlier treatment).

Figure 8 compares the performance of the 3 machine learning models with and without outlier treatment. Outlier treatment uniformly improved accuracy across all models. Random Forest showed a notable improvement in accuracy and reduction in false negatives with outlier treatment. Outliers could distort the decision boundaries of the individual trees in the Random Forest, leading to less accurate predictions. By removing these outliers, the model could focus on the more representative data points, resulting in a more stable and accurate ensemble prediction. Gradient Boosting's accuracy gain was minimal, but the elimination of false negatives suggested that outlier treatment helped the model better identify the minority class. Gradient Boosting built models sequentially, where each new model aimed to correct the errors (residuals) of the previous model. Where outliers were present, the residuals could become skewed as the model tries to fit these extreme values. Outlier treatment allowed the residuals to become more representative of the actual patterns in the data. The model could better focus on learning the true pattern including those of the minority class. There was substantial increase in accuracy for Logistic Regression, but the increase in false negatives indicated a trade-off. Logistic Regression was particularly sensitive to outliers because it relied on a linear decision boundary. Outliers could disproportionately influence the model's coefficients, leading to poor performance. By removing outliers, the model's coefficients were not skewed by extreme values, resulting in a more accurate model overall. While the overall accuracy improved, the increase in false negatives suggested that the model might be overfitting to the majority class. This happened because the model became less flexible in handling variations in the minority class after outlier removal.

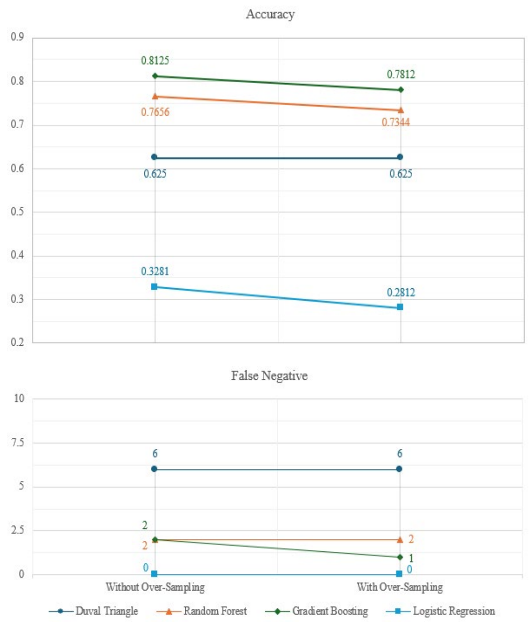


Fig. 7 *Model performance with and without oversampling*

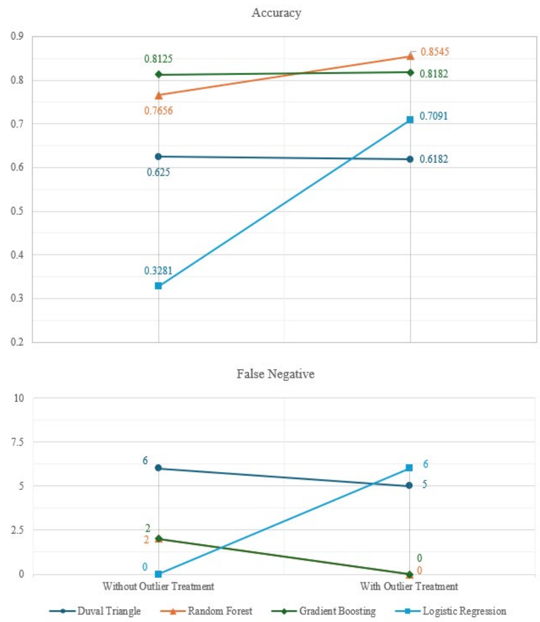


Fig. 8 *Model performance with and without outlier treatment*

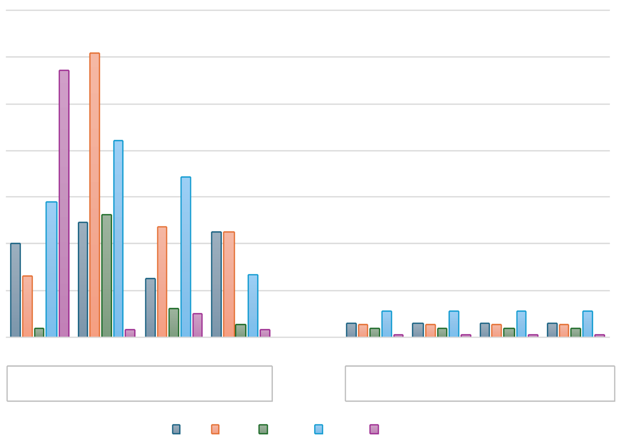


Fig. 9 *Comparison of the values of input features before and after clipping peaks*

In this experiment, the best accuracy was achieved by the Random Forest model, with outlier treatment, with an accuracy of 85.45% and no false negatives. It is crucial to note that, as the predetermined upper boundary values were applied in the value clipping process, some datasets with different fault classifications (i.e. D2, T1/T2, and T3) now had similar values across all 5 input features. This inconsistency could confuse the model, as identical input feature values corresponded to different output fault classes. Such examples are shown in Figure 9.

To address this issue, all values equal to or exceeding the upper boundary across all 5 input features were removed. This resulted in the elimination of 92 rows, reducing the dataset from 639 to 547 rows. While outlier treatment might slightly improve accuracy of the best model (by approximately 0.04), it also led to significant data loss. Consequently, the model might struggle to make accurate predictions if the test data fell outside the established boundaries. As such, Gradient Boosting model without outlier treatment was still considered as the best model. It was also observed that there was slight change in the accuracy of Duval Triangle due to removal of rows.

3.5 False Negatives

As false negatives are costly errors in fault detection, this section takes a deep dive into the only 2 false negatives produced by the best machine learning model, i.e. the Gradient Boosting model, with train-test ratio of 90:10, without oversampling and outlier treatment. Table 5 presents the 2 datasets where the model failed to detect a fault. 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 [12], except for C2H2. This was because C2H2 is an indication of sparking discharges and arcs, which was 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 1 diagnosed these transformers as "No Fault," whereas, in reality, T1/T2 faults were observed when the transformers were de-energized and inspected. Although not ideal, the consequence of the false negatives in this study was not considered as catastrophic since T1/T2 were considered low risk faults where no immediate shutdown was required. The recommended action was more frequent sampling and monitoring.

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- 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- Different diagnostic tools and techniques have varying levels of sensitivity. Duval Triangle 1 might not have been sensitive enough to detect the T1/T2 faults, especially at early stage. Triangle 1 is known to be more sensitive to CH4 and faults T3; (3) External Factors- External factors should not be discounted as potential influencers of diagnostic accuracy. 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 collectively introduce variability into the diagnostic process, potentially resulting in erroneous diagnoses.

Table 5 *Datasets of false negatives predictions*

Gas Concentration (μl/l)					Fault Prediction		
H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	Machine Learning Models	Duval Triangle	Ground Truth
86	8	2.5	2.5	0	GB	NF	T1/T2
34	30	34	2.5	0	GB	NF	T1/T2
100	80	55	170	2	Threshold for Fault Detection		

3.6 Field Trials

A field trial entails deploying the model and validating it with actual field data, which includes known faults. This exercise aimed to assess the practicality and robustness of the model and to provide an opportunity to verify the ease of model deployment, retraining and improvement. The best Gradient Boosting model, with train-test ratio of 90:10, without oversample and outlier treatment was selected for field trials.

The field trials were conducted using datasets from 4 failed transformers in different operating units in the company. Table 6 presents the DGA datasets of the 4 transformers. From the forensic investigation, 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-

Int. Journal of Integrated Engineering Vol. 18 No. 1 (2026) p. 349-367

temperature thermal fault. This coincides with a T3 fault. 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 sulfide 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 D1 fault. 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 D2 fault. TR-4 experienced an internal flashover due to the reaction between corrosive sulfur 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.

Table 7 shows the prediction results of the field trials. The results of Duval Triangle prediction were included as well for comparison purposes. The models successfully predicted the faults correctly for all 4 transformers, which were consistent with the forensic findings (ground truth). There was no false negative produced by the Gradient Boosting model. This demonstrates the model's potential for integration into real-world transformer health monitoring systems, providing accurate and timely diagnostics.

Table 6 *Datasets for field trials*

Tag	Gas Concentration (µl/l)				
	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
TR-1	2241	9841	3724	10443	0
TR-2	85	62	8	64	86
TR-3	689	89	104	8	248
TR-4	156	508	129	424	0

Table 7 *Results of field trials*

Model Type	TR-1	TR-2	TR-3	TR-4
Ground Truth	T3	D2	D1	T2
Duval Triangle	T3	D2	D1	T2
Gradient Boosting	T3	D2	D1	T1/T2

3.7 Comparison with Recent AI-Based DGA Diagnosis Methods

A study is conducted to compare the proposed work against recent AI-based DGA diagnosis methods, considering diagnostic accuracy, implementation platform, no-code or low-code approach, inclusion of no-fault classification, false negative analysis and field validation. These aspects directly relate to the practical challenges faced by field engineers, as discussed in Section 1.2. Table 8 presents comparison of this work with 20 published research works from 2021-2026.

Most studies employed advanced AI techniques and often reported high diagnostic accuracy exceeding 90%. These approaches were predominantly implemented using code-intensive platforms such as MATLAB or Python, requiring specialized machine learning expertise. As a result, field engineers remain dependent on data scientists for model development, tuning, and retraining, limiting adaptability when new fault data emerge. Majority studies included a no-fault class to distinguish healthy transformers, but false negatives which represented missed fault detections with direct implications for asset integrity and operational risk, were rarely analyzed or reported. Field validation was also not commonly implemented.

In this work, a Gradient Boosting model developed using a no-code approach in Microsoft Azure Machine Learning was presented as a practical alternative. Although the achieved accuracy is lower than some code-based approaches, the model was trained using verified ground truth datasets and included explicit evaluation of false negatives. The model was validated through field trials on failed transformers with known forensic outcomes. This demonstrated a deliberate trade-off between prediction accuracy and practical usability, positioning the proposed approach as a field-deployable diagnostic tool that directly addressed the operation needs of field engineers.

Table 8 *Comparison with recent AI-Based DGA diagnosis methods*

No.	Ref.	AI Model				Diagnosis		Field Trial	
		AI Method	Implementation Platform	No Code/ Low Code	Diagnostic Accuracy (%)	No-Fault Class Included	False Negative Analyzed	Field Validated	Field Data Size
1	[16]	MCSVM	MATLAB	No	88.9	No	Not considered	No	-
2	[26]	IF & KNN	Python	No	88.0	Yes	Not considered	No	-
3	[27]	BO-GB	Python	No	99.9	Yes	Not considered	Yes	5
4	[28]	MNN	MATLAB	No	96.0	No	Not considered	No	-
5	[29]	ANN	Not disclosed	No	99.0	Yes	Not considered	No	-
6	[30]	SVM	Not disclosed	No	98.4	Yes	Not considered	No	-
7	[31]	DT	SPINA	No	99.4	No	Not considered	No	-
8	[32]	RS & CART	Not disclosed	No	92.3	Yes	Not considered	No	-
9	[15]	MMNB	Not disclosed	No	85.7	Yes	Not considered	No	-
10	[33]	RBPNN & SVM	Not disclosed	No	92.0	Yes	Not considered	No	-
11	[34]	SKFSF	Not disclosed	No	91.2	No	Not considered	No	-
12	[35]	RFE	Not disclosed	No	98.8	Yes	Yes	Yes	305
13	[36]	TMF	MATLAB	No	84.9	Yes	Not considered	No	-
14	[37]	ANFIS	MATLAB	No	82.1	Yes	Not considered	No	-
15	[38]	FL	Not disclosed	No	65.3	No	Not considered	Yes	1
16	[39]	BTC	MATLAB	No	99.5	Yes	Not considered	No	-
17	[40]	ETC	Not disclosed	No	98.0	Yes	Not considered	No	-
18	[41]	RF	Not disclosed	No	97.0	Yes	Not considered	No	-
19	[42]	ANN	MATLAB	No	85.0	Yes	Not considered	No	-
20	[43]	RF	Not disclosed	No	100	No	Not considered	No	-
This work		GB	Azure ML	Yes	81.25	Yes	Yes	Yes	4

Note:
MCSVM: Multi-Class Support Vector Machine;
IF: Isolation Forest;
KNN: K-Nearest Neighbor;
BO-GB: Bayesian Optimization-Based Gradient Boosting;
MNN: Middle Neural Network;
ANN: Artificial Neural Network;
DT: Decision Tree;
RS: Random Search;

CART: Classification Regression Tree;
MMNB: Multi-Methods Naïve Bayes;
RBPNN: Residual BP Neural Network;
SKFSF: Sequential Kalman Filter Sensor Fusion;
RFE: Recursive Feature Elimination;
TMF: Triangular Membership Functions;
ANFIS: Adaptive Neuro-Fuzzy Inference System;
FL: Fuzzy Logic;
BTC: Bagged Tree Classifier;
ETC: Extra Tree Classifier;
RF: Random Forest;
GB: Gradient Boosting.

3.8 Model Interpretability and Feature Contribution Analysis

While prediction accuracy is an important indicator of model performance, interpretability is equally important for gaining the confidence of field engineers and supporting practical deployment. To better understand how the best-performing Gradient Boosting model arrived at its predictions, model behavior was examined at a global level using the Permutation Feature Importance component Microsoft Azure Machine Learning. It computed a set of feature importance scores for the dataset.

Figure 10 shows the global feature importance ranking of dissolved gases from the Gradient Boosting model in Microsoft Azure Machine Learning. C_2H_2 and C_2H_4 were the most influential input features across the different fault classes, followed by CH_4, C_2H_6 and H_2. This ranking was consistent with established transformer diagnostic practice in [11] and [12], where C_2H_2 is commonly associated with high energy electrical discharges (D1/D2), while elevated levels of C_2H_4 are indicative of high-temperature thermal faults (T2/T3). In comparison, CH_4 and C_2H_6 showed moderate influence and were particularly relevant for distinguishing low-temperature thermal faults (T1/T2) from normal operating conditions. Lastly, H_2, being a broad indicator across multiple fault types, carried lower global feature importance score in the model.

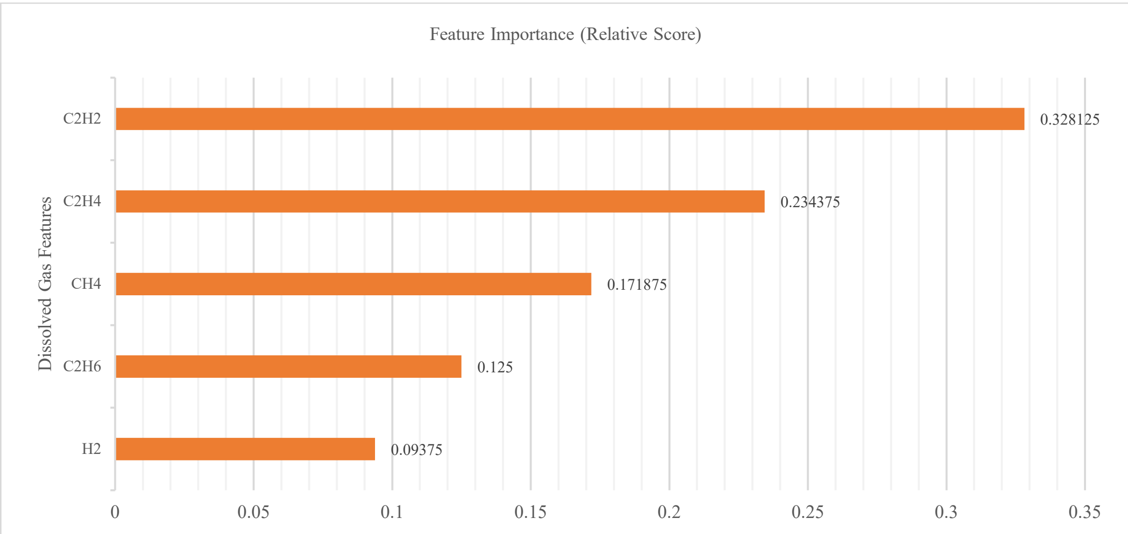


Fig. 10 *Global feature importance ranking for the Gradient Boosting model*

Overall, the relative importance of the dissolved gases aligned well with conventional dissolved gas analysis principles, suggesting that the model has learnt diagnostically meaningful patterns rather than relying on incidental correlations, supporting confidence in its use for field diagnosis.

3.9 Limitations

One of the primary challenges faced in this study was the availability and quality of data. Machine learning models require substantial volumes of high-quality data for effective training and validation. Ground truth data, being the most accurate and reliable, is ideal for this purpose. However, acquiring such data can be difficult and expensive. The ground truth datasets used in this study were sourced from various publications and expert networks, but the amount and diversity of the data were sometimes lacking. Due to data unavailability, it was also not feasible to develop a machine learning model which could also distinguish arching fault in paper (which signified high consequence fault) and oil. The issue of imbalanced data was a significant challenge in this study. Certain fault classes were underrepresented in the dataset, which could bias the model towards the majority classes.

Oversampling technique like SMOTE was applied to address this issue, but it led to a decrease in model accuracy. Balancing the dataset without introducing noise or overfitting remained a delicate task. Model interpretability was addressed at a global level through feature importance analysis and alignment with established dissolved gas–fault relationships. It answered which gases are generally most important for fault diagnosis. However, instance-level explainability, which answers why this transformer is classified as this fault, was beyond the scope of this study. While the field trials provided valuable insights into the model's performance in real-world conditions, they were limited by the number of transformers tested. The results were promising, but a larger sample size would be necessary to fully validate the model's robustness and reliability. Additionally, the field trials were conducted within a specific operational context, which might not fully represent the diversity of conditions encountered in different industries and environments.

4. Conclusion

This research successfully addressed the critical need for a more accurate, efficient, and accessible method of transformer fault diagnosis based on dissolved gas composition. The Gradient Boosting model consistently demonstrated superior performance, achieving the highest accuracy and lowest false negatives across experiments using various preprocessing techniques, including variation of train-test ratio, oversampling and outlier treatment. It outperformed both Random Forest and Logistic Regression models, as well as the conventional Duval Triangle method. Field trials validated the model's robustness in real-world conditions, with successful fault predictions for all tested transformers. These results highlight the model's practical applicability and reliability in operational environments. The deployment framework in Microsoft Azure Machine Learning further supported continuous improvement through retraining pipelines, enabling the model to adapt to evolving fault patterns and maintain diagnostic accuracy over time. Model interpretability was addressed through feature importance analysis at a global level and alignment with established dissolved gas-fault relationships, providing transparency into the model's fault classification. This could help to build trust with domain users and improve field acceptance.

By integrating machine learning techniques with a no-code methodology, the study empowered field engineers to independently create, train, and deploy models without requiring programming expertise or reliance on data scientists. This provided greater autonomy, accelerated innovation, and enhanced maintenance strategies. The ability to differentiate between healthy and faulty transformers while precisely identifying fault types and severity contributed to optimized operational efficiency and system reliability. Additionally, the proposed framework demonstrated strong scalability potential. Its modular design allowed for seamless extension to other electrical assets such as switchgears, generators, and cables. With tailored input features and output fault classes, engineers could rapidly develop high-performing diagnostic models across diverse asset types.

Future work could focus on expanding access to larger and diverse datasets with verified ground truth to improve model generalization. With sufficient data, the model can be expanded to distinguish more specific faults, such as arching faults in paper and oil. Advanced resampling techniques, such as Adaptive Synthetic Sampling (ADASYN) could be explored to better address class imbalance and improve minority fault classification performance. The model interpretability could be extended by incorporating instance-level explainability techniques, such as Shapley Additive Explanations (SHAP), to quantify each input feature's contribution to a prediction. This would enable engineers to understand the specific role of each dissolved gas for a given transformer, supporting more informed decision making in critical maintenance and risk management scenarios.

Acknowledgement

This work was supported in part by Projects, Engineering and Technology Division, PETRONAS, Malaysia.

Conflict of Interest

There is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Chee Ying Chan, Nursyarizal Mohd Nor, Nor Zaihar Yahaya;* ***data collection:*** *Chee Ying Chan;* ***analysis and interpretation of results:*** *Chee Ying Chan;* ***draft manuscript preparation:*** *Chee Ying Chan. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] L. M. Fonseca, J. P. Domingues, and A. M. Dima, "Mapping the sustainable development goals relationships," *Sustainability*, vol. 12, no. 8, p. 3359, Apr. 2020, doi: 10.3390/SU12083359.

[2] R. Mohamed, S. Abd Halim, S. Sahrani, and M. A. Talib, "Accelerated Heating Paper Pressboard and Transformer Oil for Transformer Aging Assessment," *Jurnal Kejuruteraan*, vol. 36, no. 5, pp. 1875–1883, Sep. 2024, doi: 10.17576/jkukm-2024-36(5)-08.

[3] PETRONAS, "PETRONAS Technical Standards - Pre-commissioning, Commissioning and Maintenance of Electrical Installations and Equipment, PTS 13.02.01," Oct. 2017.

[4] M. Duval, E. Alzieu, M. Banovic, and P. Boman, *Advances in DGA interpretation, JWG D1/A2.47 Technical Brochure*. Cigre, 2019.

[5] R. M. Rampersad, S. Bahadoorsingh, and C. Sharma, *Multifactorial Frameworks Modelling Linkages of Power Transformer Failure Modes*. 2018.

[6] H. Zhang, C. Zhai, G. Xiao, and T. C. Pan, "Identifying critical risks of cascading failures in power systems," *IET Generation, Transmission and Distribution*, vol. 13, no. 12, pp. 2438–2445, Jun. 2019, doi: 10.1049/iet-gtd.2018.5667.

[7] S. Pearce, H. Abb, and P. Grids, "Sustainable transformers-Solutions to increase reliability and ease maintenance," *TRANSFORMERS MAGAZINE, Volume 7, Issue 4*, vol. 7, pp. 76–79, 2020. [Online]. Available: www.transformers-magazine.com

[8] K. H. L. T. A. S. I. I. and A. A. T. Abbasi, "Condition Based Maintenance of Oil and Gas Equipment: A Review," in *3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2020*, 2020, pp. 1–9.

[9] E. Abbasi and O. P. Malik, "Failure Rate Estimation of Power Transformers Using Inspection Data," in *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016. doi: 10.1109/PMAPS.2016.7764150.

[10] Y. Zhang, Y. Tang, Y. Liu, and Z. Liang, "Fault diagnosis of transformer using artificial intelligence: A review," *Front. Energy Res.*, vol. 10, Sep. 2022, doi: 10.3389/fenrg.2022.1006474.

[11] IEEE Power & Energy Society, *IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers, C57.104-2019*. IEEE, 2019.

[12] IEC TC10, "Mineral Oil-Filled Electrical Equipment in Service – Guidance on the Interpretation of Dissolved and Free Gasses Analysis, IEC 60599," 2022.

[13] Hydroelectric Research and Technical Services Group, *Transformers: Basics, Maintenance, and Diagnostics*. Denver, Colorado: U.S. Department of the Interior - Bureau of Reclamation, 2005.

[14] S. K. Balivada and S. Karmakar, "Comparison of Dissolved Gas Interpretation Techniques in Mineral Oil Immersed Transformers," in *Proceedings of 2022 6th International Conference on Condition Assessment Techniques in Electrical Systems, CATCON 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 403–407. doi: 10.1109/CATCON56237.2022.10077623.

[15] R. A. Prasojo, H. F. W. Azizi, I. Kurniawan, Suwarno, H. Sutikno, and I. Ridzki, "Naïve Bayes Model for Combining Results from Multi-Methods Fault Identification of Power Transformer," in *Proceedings of 2023 4th International Conference on High Voltage Engineering and Power Systems, ICHVEPS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 199–203. doi: 10.1109/ICHVEPS58902.2023.10257469.

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

[17] K. Kaur, D. Bhalla, and J. Singh, "Fault Diagnosis For Oil Immersed Transformer Using Certainty Factor," *IEEE Transactions on Dielectrics and Electrical Insulation*, 2023, doi: 10.1109/TDEI.2023.3307513.

[18] Dr. M. Utmal, "Machine Learning Its Applications, Challenges & Tools: A Review," *International Journal of Computer Science and Mobile Computing*, vol. 10, no. 3, pp. 32–38, Mar. 2021, doi: 10.47760/ijcsmc.2021.v10i03.004.

[19] D. Zou *et al.*, "Power Transformer Fault Diagnosis Method Based on Machine Learning," in *Proceedings of the International Conference on Cyber-Physical Social Intelligence, ICCSI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 484–489. doi: 10.1109/ICCSI55536.2022.9970667.

[20] M. Duval and A. DePablo, "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, Mar. 2001, doi: 10.1109/57.917529.

[21] R. A. Hooshmand, M. Parastegari, and Z. Forghani, "Adaptive neuro-fuzzy inference system approach for simultaneous diagnosis of the type and location of faults in power transformers," *IEEE Electrical Insulation Magazine*, vol. 28, no. 5, pp. 32–42, 2012, doi: 10.1109/MEI.2012.6268440.

[22] N. Pattanadech, K. Sasomponsawatline, J. Siriworachanyadee, and W. Angsusatra, "The conformity of DGA interpretation techniques: Experience from transformer 132 units," in *Proceedings - IEEE International Conference on Dielectric Liquids*, IEEE Computer Society, Jun. 2019. doi: 10.1109/ICDL.2019.8796588.

[23] A.S. Monzaid, "UK GTF Main Transformer Buchholz Tripped (11th Jan 18, 15.22) Rep. INV/OEIMS/PGB/2018/0005/1," Paka, Terengganu, 2018.

[24] Data Science Team, "Azure ML Studio-Overview. Citizen Analytics (Level 3)," Kuala Lumpur, Malaysia, 2020.

[25] M. I. A. H. Hafiz *et al.*, "Application of Artificial Intelligence in Electronics and Semiconductor Industries," *Jurnal Kejuruteraan*, vol. 37, no. 7, pp. 3245–3253, Oct. 2025, doi: 10.17576/jkukm-2025-37(7)-11.

[26] R. Raghuraman and A. Darvishi, "Detecting Transformer Fault Types from Dissolved Gas Analysis Data Using Machine Learning Techniques," in *Proceedings of the 2022 IEEE Dallas Circuits and Systems Conference, DCAS 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/DCAS53974.2022.9845611.

[27] D. Paul, A. K. Goswami, R. L. Chetri, R. Roy, and P. Sen, "Bayesian Optimization-Based Gradient Boosting Method of Fault Detection in Oil-Immersed Transformer and Reactors," in *IEEE Transactions on Industry Applications*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1910–1919. doi: 10.1109/TIA.2021.3134140.

[28] N. Verma and A. K. Chandel, "Incipient Fault Diagnosis of Power Transformer Based on Duval Pentagon Using Medium Neural Network," Institute of Electrical and Electronics Engineers (IEEE), Jan. 2023, pp. 1–5. doi: 10.1109/stpes54845.2022.10006647.

[29] N. Poonnoy, C. Suwanasri, and T. Suwanasri, "Neural Network Approach to Dissolved Gas Analysis for Fault Analysis in Power Transformers," in *Proceedings of the 2022 International Electrical Engineering Congress, iEECON 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/iEECON53204.2022.9741595.

[30] J. Subalakshmi, J. J. Janet, J. J. S. Lakshmi, T. Sukumar, and B. Vigneshwaran, "Recognition of Power Transformer Defect Identification Based on Dissolved Gas Analysis using Support-Vector Machine Approach," in *6th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 104–108. doi: 10.1109/ICECA55336.2022.10009258.

[31] A. G. C. Menezes, M. M. Araujo, O. M. Almeida, F. R. Barbosa, and A. P. S. Braga, "Induction of Decision Trees to Diagnose Incipient Faults in Power Transformers," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 29, no. 1, pp. 279–286, Feb. 2022, doi: 10.1109/TDEI.2022.3148453.

[32] T. Wang, J. Yang, J. Shao, P. Wu, Y. Ma, and J. Liang, "Improvement of Power Transformer Faults Diagnosis Using Random Search with Classification and Regression Tree," in *2023 5th Asia Energy and Electrical Engineering Symposium, AEEES 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 653–658. doi: 10.1109/AEEES56888.2023.10114139.

[33] Y. Jin, H. Wu, J. Zheng, J. Zhang, and Z. Liu, "Power Transformer Fault Diagnosis Based on Improved BP Neural Network," *Electronics (Basel).*, vol. 12, no. 16, p. 3526, Aug. 2023, doi: 10.3390/electronics12163526.

[34] M. Demirci, H. Gözde, and M. C. Taplamacioglu, "Improvement of power transformer fault diagnosis by using sequential Kalman filter sensor fusion," *International Journal of Electrical Power and Energy Systems*, vol. 149, Jul. 2023, doi: 10.1016/j.ijepes.2023.109038.

[35] S. Das, A. Paramane, S. Chatterjee, and U. M. Rao, "Accurate Identification of Transformer Faults From Dissolved Gas Data Using Recursive Feature Elimination Method," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 30, no. 1, pp. 466–473, Feb. 2023, doi: 10.1109/TDEI.2022.3215936.

[36] T. Manoj, C. Ranga, S. S. M. Ghoneim, U. M. Rao, and S. A. M. Abdelwahab, "Alternate and Effective Dissolved Gas Interpretation to Understand the Transformer Incipient Faults," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 30, no. 3, pp. 1231–1239, Jun. 2023, doi: 10.1109/TDEI.2023.3237795.

[37] B. Jan, O. Rahman, S. Parveen, and S. A. Khan, "Multi-Stage Binary Classification Technique for Incipient Fault Diagnosis of Oil Immersed Power Transformers Based on ANFIS," in *2023 International Conference on Power, Instrumentation, Energy and Control, PIECON 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/PIECON56912.2023.10085900.

[38] Z. Gao, X. Wang, F. Yi, W. Xu, and M. Zhu, "Transformer DGA Diagnostic Method Based on Fuzzy Logic Function Method," in *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, IEEE, Dec. 2023, pp. 1639–1642. doi: 10.1109/ITAIC58329.2023.10408761.

[39] A. Dhiman, H. K. Azmeera, and C. C. Reddy, "Intelligent Diagnosis of Transformer Faults Using DGA," in *Proceedings of 2024 IEEE 7th International Conference on Condition Assessment Techniques in Electrical Systems, CATCON 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 166–170. doi: 10.1109/CATCON60527.2024.10831256.

[40] D. Das, S. Pani, S. Chakravorti, and A. Baral, "ML-Based Improved Incipient Fault Detection in Power Transformers using DGA Data," in *Proceedings of 2024 IEEE 7th International Conference on Condition Assessment Techniques in Electrical Systems, CATCON 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 77–81. doi: 10.1109/CATCON60527.2024.10831821.

[41] T. Luan, "Transformer Fault Diagnosis Based on Artificial Intelligence-Enhanced Dissolved Gas Analysis," in *2025 5th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)*, IEEE, Nov. 2025, pp. 158–163. doi: 10.1109/ICAIRC68035.2025.11385262.

[42] M. Tešanović, E. Čerkezović, A. Karišik, and M. Kasumagić, "Possibilities of ANN Application for Condition Monitoring of Power Transformers Using DGA," in *2025 24th International Symposium INFOTEH-JAHORINA, INFOTEH 2025 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/INFOTEH64129.2025.10959198.

[43] R. S. Tarade and P. Kundu, "A Performance Comparison of Fault Diagnosis Methods for Power Transformer using Machine Learning Techniques based on DGA," in *2026 International Conference on Electric Power and Renewable Energy (EPREC)*, IEEE, Jan. 2026, pp. 1–5. doi: 10.1109/EPREC66546.2026.11412063.