Journal of Engineering Science and Technology
Vol. 20, No. 4 (2025) 1209 - 1225
© School of Engineering, Taylor's University

PREDICTIVE MODELLING OF OLTC SYSTEM FAULTS: COMPARATIVE EVALUATION AND VALIDATION OF MACHINE LEARNING APPROACHES

FIRAS BASIM. ISMAIL[1,4,5,*], SITI ROSILAH ARSAD[2,7],
HUSSEIN AL-FAIZ[3], YOUNG ZAIDEY YANG GHAZALI[6]

[1]Smart Power Generation Unit, Institute of Power Engineering,
Universiti Tenaga Nasional (UNITEN), Kajang, 43000, Malaysia.
[2]Department of Electrical & Electronics Engineering, College of Engineering,
Institute of Sustainable Energy, Universiti Tenaga Nasional, 43000, Kajang, Malaysia
[3]College of Graduate Studies (COGS), Universiti Tenaga Nasional
(UNITEN), 43000 Kajang, Selangor, Malaysia
[4]Faculty of Engineering, Sohar University, PO Box 44, Sohar PCI 311, Oman.
[5]Al-Bayan university, Engneering College, Baghdad, Iraq
[6]Distribution DivisionTenaga Nasional BerhadWisma TNB,
Jalan Timur, PetalingJaya 46200, Selangor, Malaysia
[7]Department of Engineering, School of Engineering and Technology,
Sunway University, Bandar Sunway, Petaling Jaya, 47500, Malaysia
*Corresponding Author: Firas@uniten.edu.my

Abstract

This study presents a machine learning (ML)-based framework for on-load tap changer (OLTC) fault detection. By integrating the Duval Triangle Method (DTM) with five ML classifiers, the study addresses the limitations of traditional dissolved gas analysis (DGA), such as subjectivity and dataset imbalance. Among the five ML models evaluated, Fine-KNN exhibited superior accuracy and robustness in classifying transformer faults, achieving a validation accuracy of 75%. The key novelty of this research lies in the comparative evaluation of multiple ML models for OLTC fault prediction and the use of data augmentation techniques to enhance model performance. The results suggest that ML provides a promising solution for early fault detection in OLTC transformers, reducing the likelihood of unexpected failures. Future work should explore the integration of real-time monitoring systems, hybrid ML models, and digital twin technology to further improve predictive accuracy and deployment feasibility.

Keywords: Duval triangle method, Incipient fault detection, Machine learning, On-load tap changer (OLTC), Transformer monitoring (T).

1. Introduction

Power transformers play a crucial role in power transmission and distribution networks, with OLTCs ensuring voltage regulation. However, OLTCs are among the most vulnerable components, responsible for approximately 30% of transformer failures due to mechanical wear and electrical faults [1]. Therefore, effective fault detection and condition monitoring of OLTCs are essential for maintaining transformer reliability.

The OLTC is both the most expensive and failure-prone component, requiring mechanical movement that increases the risk of operational failures. OLTC failures generally result from a combination of mechanical and electrical issues, where mechanical problems often trigger subsequent electrical faults. To monitor OLTC conditions, various offline and online diagnostic techniques are employed. Offline methods include dissolved gas analysis (DGA), static winding resistance measurement (SWRM), and dynamic resistance measurement (DRM), while online approaches include online DGA, vibro-acoustic analysis, and infrared (IR) thermography [2]. This highlights the importance of timely condition assessment for this component. OLTC failures typically result from a combination of mechanical and electrical issues, with mechanical problems often leading to electrical faults.

Several techniques are employed for monitoring the condition of OLTCs, which can be broadly classified into offline and online methods. Offline techniques include dissolved gas analysis (DGA), static winding resistance measurement (SWRM), and dynamic resistance measurement (DRM). Conversely, online approaches comprise online DGA, vibro-acoustic analysis, and infrared (IR) thermography [3]. This study addresses the following key research question: Can machine learning models improve the predictive accuracy of OLTC fault detection compared to traditional DGA methods? The primary aim of this research is to develop and evaluate machine learning models for early fault detection in OLTC systems, leveraging historical DGA data.

The study seeks to determine the most effective model for fault classification, thereby enhancing the reliability of OLTC monitoring. The findings of this study have direct practical implications for power utilities. The proposed machine learning models can be deployed for real-time OLTC fault detection, allowing utility companies to shift from reactive maintenance to predictive maintenance strategies.

By integrating the developed models with Internet of Things (IoT) platforms, real-time monitoring of OLTCs can be achieved, providing early warnings of potential faults and reducing unexpected transformer failures. Utility providers such as Tenaga Nasional Berhad (TNB) could benefit from these AI-driven solutions by implementing predictive maintenance frameworks, leading to improved grid reliability, extended transformer lifespan, and optimized maintenance costs.

DGA is widely used for monitoring OLTCs in transformers, with gas chromatography (GC) as the primary gas detection method. The advantage of DGA is its ability to assess both new and repaired transformer units, schedule maintenance effectively, and monitor transformers susceptible to overload conditions. This technique detects specific gases in OLTC oil to diagnose faults, as each type of fault produces characteristic gas signatures. Transformer faults are identified when specific gas concentrations exceed predefined thresholds.

To diagnose OLTC-related faults, DGA measures gases such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2). These gases originate from electrical and thermal faults within transformer oil. Various diagnostic techniques are used to analyse these gases, including the Doernenburg Ratio Method (DRM), Rogers Ratio Method (RRM), IEC Ratio Method (IRM), Duval Triangle Method (DTM), and Duval Pentagon Method (DPM) [4].

These gases are generated due to electrical and thermal faults in transformer oil. Fault types are determined by analysing the gas concentrations using established techniques such as the Doernenburg Ratio Method (DRM), Rogers Ratio Method (RRM), IEC Ratio Method (IRM), Duval Triangle Method (DTM), and Duval Pentagon Method (DPM) [5].

While conventional DGA methods such as the Doernenburg Ratio Method, Rogers Ratio Method, and IEC Ratio Method have been widely used for OLTC fault diagnosis, they rely heavily on predefined gas ratio thresholds. These thresholds may not always generalize well across different transformer operating conditions, leading to potential misclassifications.

Additionally, manual interpretation of DGA results introduces subjectivity, making real-time monitoring impractical. Recent studies have attempted to overcome these limitations using AI-based techniques, but many fail to address key challenges such as dataset imbalance, overfitting, and the need for real-time deployment.

Among the conventional methods, the Duval Triangle Method (DTM) stands out due to its accuracy of up to 96%, leaving only a 4% probability of incorrect diagnosis. DTM uses a graphical approach that plots three key gases CH_4, C_2H_4, and C_2H_2 on a triangular diagram in parts per million (ppm). This method is recognized for its high accuracy, efficiency, and repeatability in fault detection. Studies [6, 7] have demonstrated the superiority of DTM in fault diagnosis due to its ability to focus on three critical gases that indicate fault severity:

- CH4: Indicates low-energy or low-temperature faults.
- C2H4: Detects high-temperature faults.
- C2H2: Identifies very high-energy or arcing faults.

DTM's reliability has led to its adoption as a benchmark in studies exploring new DGA interpretation methods, underscoring its role as a dependable standard for DGA analysis. Although the Duval Pentagon Method (DPM) is a newer approach with potential, it requires further validation. For now, DTM remains the most trusted and well-established method for diagnosing faults in transformer insulation [8].

The IEC Ratio Method (IRM) [9] and IEEE C57.104 standard [10] provide critical guidelines for interpreting dissolved gas concentrations, though their reliance on static thresholds limits adaptability to dynamic operating conditions. Recent advancements in DGA interpretation, such as those outlined in the CIGRE Technical Brochure D1/A2.47 [11], emphasize the need for adaptive techniques to improve diagnostic accuracy. Table 1 shows the DGA interpretation for fault identification methods.

Table 1. DGA interpretation for fault identification methods [9, 11].

DGA Interpretation methods	Description	Fault Identification and Normal Aging	Advantages	Disadvantages
Doernenburg Ratio Method (DRM)	Utilizes the ratios of four gases (CH_4/H_2, C_2H_2/C_2H_4, C_2H_2/CH_4, and C_2H_6/C_2H_2) to diagnose faults.	Identifies thermal decomposition, partial discharge (PD), and arcing faults.	-Helps classify fault types graphically. - Accurate in detecting overheating faults.	-Provides only an initial and basic assessment. - Higher likelihood of misdiagnosis or false negatives.
Rogers Ratio Method (RRM)	Employs the ratios of three gases (CH_4/H_2, C_2H_2/C_2H_4, and C_2H_4/C_2H_6) for fault identification.	Detects normal aging, PD, arcing, low-temperature faults, and thermal faults below or above 700°C.	-Comprehensive coding with no blind spots. - Satisfactory accuracy for compound fault analysis., and the accuracy is satisfactory.	-Cannot pinpoint the specific location of the fault.
IEC Ratio Method (IRM)	Examines the ratios of three gases (CH_4/H_2, C_2H_2/C_2H_4, and C_2H_4/C_2H_6) for fault classification.	Covers PD, low- and high-energy discharges, and thermal faults within three temperature ranges (<300°C, 300–700°C, >700°C).	-Fault sequences are logically arranged from initial to severe stages.	Some faults remain undiagnosed due to incomplete coding. - Requires expert knowledge for precise fault classification.
Duval Triangle Method (DTM)	Uses three gases (CH_4, C_2H_4, and C_2H_2) aligned with increasing fault temperature or energy content.	Detects PD, low- and high-energy discharges, and thermal faults in three temperature ranges (<300°C, 300–700°C, >700°C).	-Offers high accuracy and repeatability. - Diagnoses a wide variety of faults efficiently.	-Lacks the ability to describe normal transformer aging. - May confuse thermal and electrical faults.
Duval Pentagon Method (DPM)	Evaluates faults by analysing five gases (H_2, CH_4, C_2H_4, C_2H_2, and C_2H_6) based on energy or temperature levels.	Detects normal aging, PD, low- and high-energy discharges, and thermal faults across various temperature ranges (<300°C, 300–700°C, >700°C).	-Predicts normal conditions and stray gas from insulation degradation. - Differentiates PD from low-energy thermal faults.	-Relatively new and requires further validation through additional studies.

Cheng and Yu [12] surveyed intelligent DGA approaches for large oil-immersed transformers, highlighting the potential of machine learning to address limitations in conventional ratio methods. Traditionally, gas chromatography (GC) has been used in DGA methods for diagnosing faults such as overheating, electrical arcing, and insulation damage. While effective, these GC-based diagnostic methods combined with DGA interpretation techniques have certain limitations. For example, the characteristic gas method often lacks precision and efficiency, making it challenging to pinpoint fault locations accurately. Faults with similar gas characteristics can result in misjudgements, and the Duval Triangle Method (DTM) may fail to detect partial discharge (PD) faults.

Ghoneim and Taha [13]. proposed a novel DGA interpretation technique using probabilistic classifiers, achieving higher accuracy than traditional methods. Recent work by Nanfak et al. [14] introduced a data-driven approach for labelled fault analysis, demonstrating improved reliability in gas interpretation.

Early neural network applications in transformer diagnosis, as explored by Guardado et al. [15], laid the groundwork for modern AI-driven methods. Miranda et al. [16] further advanced fault diagnosis using auto associative neural networks, while Wei et al. [17] integrated support vector machines (SVMs) with feature prioritization for DGA analysis. Mansour [18] developed a graphical DGA technique based on five combustible gases, offering a robust alternative to traditional triangle methods.

With the rapid advancements in computer technology and AI theory, this study introduces machine learning (ML) as an innovative strategy for early prediction and diagnosis of OLTC faults in transformers. Employing ML to analyse DGA data for identifying OLTC faults offers several advantages. One such advantage is the automation of the fault identification process, allowing the model to analyse DGA data without manual intervention, thereby saving time and resources. This approach also reduces the potential for human error in DGA data analysis, contributing to more consistent and reliable fault detection [19]. ML systems can be specifically designed for continuous monitoring of DGA data, providing real-time insights into the condition of OLTC. This continuous monitoring allows for timely responses to emerging faults.

Additionally, when properly trained on relevant data, ML algorithms have the potential to provide more accurate predictions compared to traditional methods [20]. This accuracy can enhance the reliability of OLTC fault diagnosis. To implement this approach, three ML models namely, ANN, Decision Tree (DT), and K-Nearest Neighbours (KNN) are utilized to train the DGA data. The selection of these models is based on their widespread applicability to monitoring machine parameter conditions, quick learning capabilities, and their ability to predict outcomes by considering various inputs [21]. Model accuracy will be assessed as part of performance evaluation and compared with results obtained from the DGA analysis conducted in the chemical laboratory of the Malaysian electric utility, Tenaga Nasional Berhad (TNB).

Fan et al. [22] highlighted the importance of data preprocessing for reliable knowledge discovery, a critical consideration for DGA datasets. Although various machine learning models, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT), have been applied to transformer fault diagnosis, their performance is often constrained by small datasets and a lack of real-world validation. Many studies train models on simulated or limited experimental data, which may not reflect actual transformer operating conditions.

Moreover, deep learning techniques such as convolutional and transformer-based networks have shown promising results in fault classification but require large-scale datasets, which are often unavailable in OLTC monitoring applications. This study addresses these gaps by using a real-world dataset from

Tenaga Nasional Berhad (TNB) and implementing the Synthetic Minority Oversampling Technique (SMOTE) to mitigate class imbalance, ensuring a more robust and generalizable fault prediction model.

To address dataset imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) [23] was applied during preprocessing. Chawla et al. [23] demonstrated SMOTE's effectiveness in handling class imbalance for fault classification tasks. Materials informatics principles [24] guided feature engineering, ensuring robust representation of gas concentration patterns. Alqudsi and El-Hag [25] demonstrated the application of machine learning for predicting transformer health indices, highlighting the potential of AI-driven approaches in predictive maintenance. Their work underscores the importance of integrating historical DGA data with modern algorithms to enhance diagnostic accuracy.

Similarly, Mateus et al. [26] proposed a hybrid framework combining fuzzy logic and neural networks for transformer fault detection, achieving robust performance under noisy operational conditions. Their methodology aligns with the need for adaptive models in real-world power systems.

Recent advancements in machine learning have significantly improved OLTC fault diagnosis. For instance, integrating acoustic emission signals with machine learning techniques has enabled precise, non-intrusive diagnostics of OLTCs [27]. Comprehensive reviews have highlighted the role of artificial intelligence in enhancing transformer fault diagnosis accuracy and reliability [28]. Additionally, the application of Transformer-based models has shown promise in intelligent fault diagnosis tasks, offering superior modelling capabilities for complex mechanical systems [29].

Novel frameworks like FaultFormer demonstrate the potential of self-supervised pretraining in improving fault classification performance, especially in scenarios with limited data [30, 31]. However, many of these approaches lack real-world validation on datasets from operational power transformers. This study addresses this gap by utilizing actual OLTC data from Tenaga Nasional Berhad (TNB), ensuring a practical evaluation of machine learning-based fault diagnosis.

This work makes three key contributions: (1) It presents a comparative analysis of five machine learning models—Fine-KNN, Weighted-KNN, Ensemble Bagged Trees, Ensemble Subspace KNN, and Wide Neural Network—to identify the most effective approach for OLTC fault detection. (2) It integrates the Duval Triangle Method (DTM) for feature selection, optimizing the classification process. (3) It employs data augmentation using the Synthetic Minority Oversampling Technique (SMOTE) to balance the dataset, improving the robustness of the models. The study's findings provide valuable insights into the application of AI-driven solutions for power transformer fault monitoring, paving the way for more reliable predictive maintenance strategies.

While several studies have explored OLTC fault diagnosis using conventional DGA interpretation methods and early AI-based approaches, they often suffer from limitations such as reliance on expert-defined thresholds, lack of real-world validation, and imbalanced datasets. This study addresses these gaps by systematically comparing multiple machine learning models trained on a real-world dataset obtained from Tenaga Nasional Berhad (TNB).

Unlike previous works, this study also employs data augmentation through the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance, ensuring more reliable fault detection. The integration of the Duval Triangle Method (DTM) with machine learning further enhances interpretability, providing a novel approach to OLTC fault diagnosis. To the best of our knowledge, this is one of the first studies to combine these techniques for improving OLTC fault prediction.

2. Research Methodology

This section outlines the methodology used in developing and evaluating the ML models for OLTC fault diagnosis. The methodology encompasses data collection, preprocessing, model construction, and performance evaluation.

2.1. Data collection

This study analyses OLTC fault data obtained from the chemical laboratory of TNB, Malaysia. The dataset comprises DGA information from 2996 samples of transformers situated in a selected region. The DGA data, presented in an MS Excel spreadsheet, includes percentages of combustion gases such as H2, O2, N2, CH4, CO, CO2, C2H4, C2H6 and C2H2. The initial categorization of the 2996 samples of transformers into Normal and Faulty conditions is determined using DTM interpretation methods in accordance to the guideline in IEEE C57.104-2019 [10], as outlined in Table 2.

Table 2. Total number of transformers with the initial predicted condition based on the DTM interpretation method.

Initial predicted condition	Number of Transformer
Normal condition	1997
Abnormal arcing(D1)	15
Abnormal arcing(D1) or Thermal fault (X1)	233
Fault T3 or T2 in progress or severe arcing (D2)	502
Severe thermal fault (T2)	64
Severe thermal fault (T3)	185

Initially, 1997 transformers are identified as being in a normal condition, while 999 transformers exhibit Faulty conditions. Among the Faulty transformers, 502 are classified under fault types T3, T2, or severe arcing (D2), with 233 showing Abnormal arcing (D1) or Thermal fault (X1). Furthermore, 185 transformers exhibit Severe thermal fault (T3), 64 have Severe thermal fault (T2), and only 15 are categorized as having Abnormal arcing (D1) faults. This dataset will serve as input for training the ML model used in identifying OLTC faults in transformers.

2.2. Pre-processing of data

Data pre-processing is an essential step in machine learning, ensuring datasets are clean, relevant, and well-structured for analysis [22]. This process plays a pivotal role in improving data quality, making it suitable for effective model training and performance evaluation, ultimately leading to the development of accurate and reliable machine learning models. Key steps in data pre-processing include data augmentation (artificially increasing dataset size), data cleaning (eliminating outliers), data scaling (normalizing data within a 0-1 range), data reduction (minimizing data volume while retaining important information), and data splitting (dividing data into training, testing, and validation sets).

This study employs data augmentation to increase the dataset size from 2996 to 48,000 using the Synthetic Minority Oversampling Technique (SMOTE) [23]. Additionally, this technique addresses the imbalanced dataset by distributing it, aiming to prevent biased models through the generation of synthetic samples based on the minority class. In this study, the minority samples are augmented to 8000 datasets for each condition as shown in Table 3. SMOTE was applied to handle class imbalance by synthetically generating minority class instances. While this technique is effective in addressing dataset limitations, future work could focus on incorporating real-world datasets from multiple transformer systems to improve model generalization.

Table 3. Augmented data based on SMOTE.

Types of condition	Total Number
Normal condition	8000
Abnormal arcing(D1)	8000
Abnormal arcing(D1) or Thermal fault (X1)	8000
Fault T3 or T2 in progress or severe arcing (D2)	8000
Severe thermal fault (T2)	8000
Severe thermal fault (T3)	8000

The original dataset is merged with the augmented dataset in an Excel file, which is then imported and pre-processed in MATLAB. During the data cleaning process, outliers are removed by imputing problematic data and making necessary modifications, replacements, or deletions. The significance of dataset in ML is crucial, as the presence of contamination can hinder system functioning or yield inaccurate results.

Subsequently, data scaling is performed to maintain predictive modelling integrity, normalizing input variables between 0 and 1, especially when they exhibit varying scales. Data reduction is achieved through feature selection, where relevant characteristics are identified. In this study, manual feature selection using the Duval triangle technique is applied, reducing the dataset to three crucial features out of ten, specifically CH4, C2H4, and C2H2. Before ML model creation, the dataset is divided into two subsets for training and testing, employing the k-fold cross-validation method. Finally, data splitting is applied, allocating 70% for data training and 30% for data testing/validation. The flowchart depicting the data pre-processing steps is illustrated in Fig. 1.

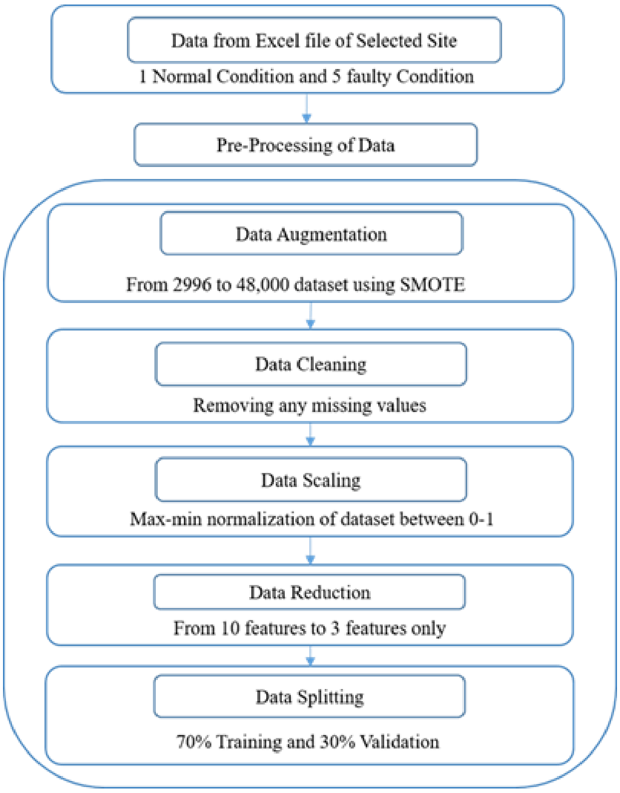

Fig. 1. Flowchart of the pre-processing phase.

2.3. Data preprocessing

The development of machine learning (ML) models involves four main steps: data collection, data representation, algorithm selection, and model optimization [24]. The initial stages of data collection and representation have already been covered as part of the preprocessing process.

The next phase focuses on selecting a suitable algorithm that aligns with the dataset's characteristics. Commonly used algorithms include decision trees (DT), artificial neural networks (ANN), and k-nearest neighbours (KNN). A brief introduction to these methods will be provided in the subsequent sections.

In this research, the chosen algorithms are constructed through the MATLAB Classification Learner Application. The dataset is trained using classification algorithms, and the top five algorithms with the highest accuracy are identified through the MATLAB Classification Learner Application.

After training the model, performance evaluation is executed using metrics specifically designed for classification tasks. Upon achieving consistent performance, verification is sought through the integration of real fault data from various on-site OLTCs. Subsequently, a validation test is conducted to assess the efficacy of the selected trained model. The test involves 60 units of unseen data, evenly distributed as shown in Table 4.

**Table 4. Data breakdown for validation test:
60 samples evenly allocated among OLTC fault types.**

Types of condition	**Total number**
Normal condition	10
Abnormal arcing(D1)	10
Abnormal arcing(D1) or Thermal fault (X1)	10
Fault T3 or T2 in progress or severe arcing (D2)	10
Severe thermal fault (T2)	10
Severe thermal fault (T3)	10

The validation test performance is assessed using metrics such as the confusion matrix and accuracy. These results provide an overall evaluation of the model's effectiveness. Figure 2 presents a summary of the steps involved in the development of the machine learning algorithm.

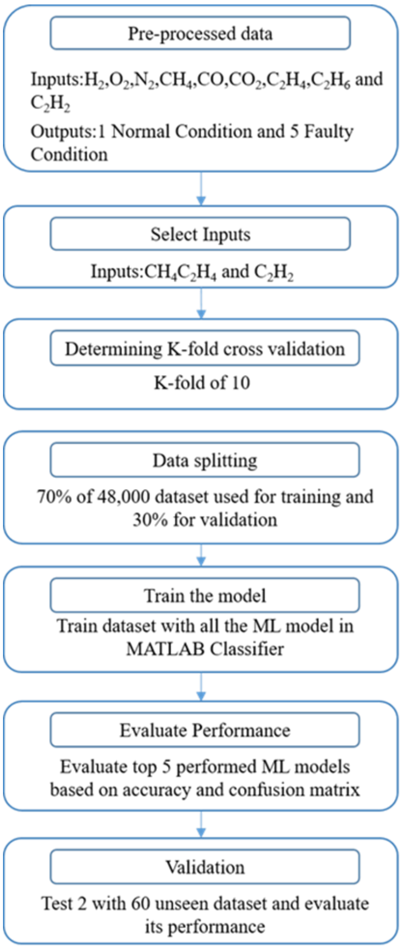

Fig. 2. Building the ML model.

2.4. Model Development

Evaluating the performance of a machine learning (ML) model is crucial to confirm its accuracy in predicting new, unseen data and to ensure it meets the desired objectives. Common metrics and techniques for this evaluation include the confusion matrix, accuracy, recall, precision, and F1 score [25].

In this study, the confusion matrix is used to provide a detailed analysis of the model's performance by highlighting the number of instances that were correctly or incorrectly classified within each category. This analysis helps pinpoint specific areas where the model may be prone to errors, offering valuable insights into its strengths and weaknesses.

The confusion matrix is especially effective in identifying tendencies toward false positives or false negatives, which is critical for diagnosing faults, particularly OLTC faults in transformers [26]. Accuracy serves as a high-level metric to provide a quick overview of the model's performance. To gain a deeper insight, especially when dealing with imbalanced datasets, additional metrics such as precision, recall, and F1 score are utilized. Together, these metrics provide a comprehensive evaluation of the model's effectiveness and reliability.

3. Evaluation and Results

Following the preprocessing stage, the machine learning models were trained using 70% of the dataset, comprising 48,000 data points, while the remaining 30% was allocated for testing. To ensure robust model performance, 10-fold cross-validation was applied during training. The models were assessed using evaluation metrics such as accuracy, recall, precision, and F1-score.

The MATLAB Classification Learner app was employed to identify the top-performing machine learning models. The five best-performing models, as detailed in Table 5, include Fine K-Nearest Neighbour (KNN), Weighted KNN, Ensemble (Subspace KNN), Ensemble (Bagged Trees), and Neural Network (Wide Neural Network).

Table 5. Performance results of the top 5 ML models based on the MATLAB classification learner app.

Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
KNN (Fine-KNN)	99.9	99.88	99.9	99.89
KNN (Weighted- KNN)	99.8	99.78	99.78	99.78
Ensemble (Bagged Tree)	99.5	99.55	99.55	99.55
Ensemble (Subspace KNN)	98.8	98.78	98.77	99.77
ANN (Wide Neural Network)	99.3	99.35	99.33	99.34

The results in Table 5 indicate that the Fine-KNN model achieved the highest accuracy of 99.9%, with excellent precision (99.9%), recall (99.88%), and F1-score (99.89%). Similarly, the Weighted-KNN model performed well, achieving 99.8% accuracy, with precision, recall, and F1-score values of 99.78%.

The Ensemble Bagged Tree model, a variation of the Decision Tree (DT) algorithm, demonstrated a slightly lower accuracy of 99.5% but still maintained high precision and recall scores. The Ensemble Subspace KNN model performed with 98.8% accuracy, while the Wide Neural Network (ANN) attained an accuracy of 99.3%, showcasing strong classification capabilities. Overall, the Fine-KNN model emerged as the best performer, demonstrating superior predictive ability for OLTC fault classification.

In the next step, the evaluation process will include the construction of confusion matrices for each model. The confusion matrix provides a visual representation of the performance of a model, detailing accurate predictions and types of misclassifications. This step is crucial for gaining deeper insights into how well each model distinguishes between different classes and will contribute to a comprehensive assessment of their effectiveness in practical applications.

3.1. Fine-KNN model

The confusion matrix for the Fine-KNN model is presented in Table 6. The model demonstrated outstanding classification accuracy, particularly for Classes 3, 5, and 6, which were correctly predicted 100% of the time. Misclassification rates were minimal: 0.2% for Normal Condition (Class 1), 0.3% for Abnormal Arcing (D1) (Class 2), and 0.2% for Fault T3 or T2 in Progress or Severe Arcing D2 (Class 4). These results highlight the Fine-KNN model's capability in detecting severe faults, including high-temperature thermal faults.

Table 6. Confusion matrix of Fine-KNN (%).

True class	1	2	3	4	5	6
1	99.8	0.1		0.1		
2	0.3	99.7				
3			100.0		0.0	
4	0.1			99.8	0.0	0.0
5					100.0	
6						100.0

Predicted class

3.2. Weighted-KNN model

Table 7 presents the confusion matrix for the Weighted-KNN model. This model exhibited strong classification performance but showed slightly higher misclassification rates compared to Fine-KNN. Notably, 0.5% of Abnormal Arcing D1 (Class 2) instances were misclassified, and 0.4% of Fault T3 or T2 in Progress cases (Class 4) were incorrectly identified.

Table 7. Confusion matrix of Weighted-KNN (%).

True class	1	2	3	4	5	6
1	99.8	0.8	0.3	0.5	0.0	0.2
2	1.0	98.9	0.1			
3	0.3	0.1	99.4	0.1	0.1	0.0
4	0.9	0.2	0.4	97.7	0.2	0.5
5	0.0	0.0	0.1	0.1	99.7	0.0
6	0.1		0.2	0.5	0.3	98.9

Predicted class

3.3. Ensemble bagged trees model

Table 8 shows the confusion matrix for the Ensemble Bagged Tree model. Misclassifications occurred in 0.9% for Normal Condition (1), 0.7% for Abnormal Arcing D1 (2), 0.9% for Fault T3 or T2 in Progress or Severe Arcing (D2) – Light Coking or Increased Contact Resistance (4), and 0.2% for Severe Thermal Fault T3 (6). The remaining instances were correctly classified.

The table reveals that utilizing the Ensemble Bagged Tree model for incipient fault prediction is less suitable, particularly in distinguishing between normal conditions and Abnormal Arcing (D1). Additionally, the model exhibits a 0.2% misclassification for severe thermal fault (T3), posing a potential risk due to oversight.

Table 8. Confusion matrix of Ensemble Bagged Trees (%).

True class	1	2	3		4	5	6
1	99.1	0.7	0.0		0.2		
2	0.7	99.3			0.0		
3			100.0		0.0	0.0	
4	0.6	0.0	0.1		99.1		0.1
5						100.0	
6	0.0				0.1	0.1	99.8

Predicted class

3.4. Ensemble Subspace KNN model

Table 9 outlines the confusion matrix for the Ensemble Subspace KNN model. Misclassification rates were as follows: 1.9% for Normal Condition (1), 1.1% for Abnormal Arcing D1 (2), 0.6% for a combination of Abnormal Arcing (D1) or Thermal Fault in Progress (X1) (3), 2.3% for Fault T3 or T2 in Progress or Severe Arcing D2 with Light Coking or Increased Contact Resistance (4), 0.3% for Severe Thermal Fault (T2) (5), and 1.1% for Severe Thermal Fault T3 (6).

The matrix indicates that employing the Ensemble Subspace KNN model for incipient fault prediction is not suitable, as it does not achieve 100% accuracy in predicting any incipient fault.

Table 9. Confusion matrix of Ensemble SUBSPACE KNN (%).

True class	1	2	3	4	5	6
1	99.7	0.2		0.1		
2	0.5	99.5				
3	0.0		99.9		0.1	
4	0.1		0.0	99.6	0.1	0.0
5					100.0	
6				0.0		100.0

Predicted class

3.5. Wide neural network model

Table 10 presents the confusion matrix for the Wide Neural Network model. Misclassifications occurred in 2.0% for Normal Condition (1), 1.5% for Abnormal Arcing D1 (2), and 0.4% for Fault T3 or T2 in Progress or Severe Arcing D2 - Light Coking or Increased Contact Resistance (4). The remaining instances were classified correctly. This matrix indicates that using the Wide Neural Network model for incipient fault prediction encounters challenges in distinguishing between normal conditions and Abnormal Arcing (D1). Nevertheless, the model achieves 100% accuracy in predicting severe faults, which is a notable accomplishment.

Table 10. Confusion matrix of wide neural network.

True class	1	2	3	4	5	6
1	98.0	1.7		0.4		
2	1.5	98.5				
3			100.0			
4	0.4			99.6	0.0	0.0
5					100.0	
6					0.0	100.0

Predicted class

3.6. Validation test of fine-KNN

A validation test was conducted to assess the performance of the trained Fine-KNN model. Sixty samples of previously unseen data, originally collected from the dataset, were utilized for the validation test, with performance evaluated through the confusion matrix. The data distribution is detailed in Table 4 in Section 2.3.3. Table 11 presents the Confusion Matrix for the tested Fine-KNN. The model performs well in predicting Normal Operation (1) and Abnormal Arcing (D1) or Thermal fault in progress (X1) (3).

However, challenges arise in determining Abnormal Arcing D1 (2), with 80% of instances misclassified as Normal Operation (1). For Fault T3 or T2 in Progress or Severe Arcing D2 (X3) - Light Coking or Increased Contact Resistance (4), 10% of the dataset is incorrectly classified as Normal Operation (1). For Severe Thermal Fault (T2) (5), around 20% are misclassified as Abnormal Arcing (D1) or Thermal Fault in Progress (X1) (3), and 10% are misclassified as Fault T3 or T2 in Progress or Severe Arcing D2 (X3) - Light Coking or Increased Contact Resistance (4). Lastly, for Severe Thermal Fault T3 (6), the model misclassifies 30% as Fault T3 or T2 in Progress or Severe Arcing D2 (X3) - Light Coking or Increased Contact Resistance (4).

Lastly, for Severe Thermal Fault T3 (6), the model misclassifies 30% as Fault T3 or T2 in Progress or Severe Arcing D2 (X3) - Light Coking or Increased Contact Resistance (4). According to the information derived from the confusion matrix, the model is predicted to have an accuracy rate of 75%, a commendable result for the fault identification of the OLTC transformer. The model achieved a validation accuracy of 75%, which reflects the challenges in fault classification due to dataset constraints. Additional hyperparameter tuning and feature engineering could further enhance performance. However, this study primarily aimed to evaluate machine learning feasibility for OLTC fault diagnosis rather than achieving peak accuracy.

Table 11. Validation testing confusion matrix for fine-KNN (%).

True class	1	2	3	4	5	6
1	100.0					
2	80.0	20.0				
3			100.0			
4	10.0			90.0		
5			20.0	10.0	70.0	
6				30.0	0.0	70.0

Predicted class

4. Conclusion and Recommendations

This study presents a significant advancement in OLTC fault detection by integrating machine learning models with the Duval Triangle Method (DTM), addressing the limitations of traditional dissolved gas analysis (DGA)-based techniques. Unlike previous works that primarily relied on static DGA interpretation, this research employs a comparative evaluation of five machine learning models—Fine-KNN, Weighted-KNN, Ensemble Bagged Trees, Ensemble Subspace KNN, and Wide Neural Network—to enhance predictive accuracy.

The study's novelty lies in the systematic validation of these models using a real-world dataset from Tenaga Nasional Berhad (TNB), demonstrating the Fine-KNN model's superior performance with a validation accuracy of 75%. Additionally, this work introduces data augmentation through the Synthetic Minority Oversampling Technique (SMOTE) to address dataset imbalances, ensuring more reliable fault classification.

The findings highlight the potential of machine learning as a transformative tool for proactive OLTC condition monitoring, reducing the risk of unexpected failures and improving maintenance efficiency. Future research should focus on real-time deployment of ML-based OLTC monitoring systems, integration with Internet of Things (IoT) platforms for continuous data acquisition, and the application of hybrid AI techniques, such as ensemble deep learning models, to enhance fault classification accuracy. Additionally, the incorporation of digital twin technology could provide a dynamic and real-time representation of the OLTC system, further improving fault diagnosis and predictive maintenance strategies.

While the proposed machine learning models have shown promising results for OLTC fault detection, several areas can be explored for future improvements. First, integrating real-time data collection from IoT-based sensors can enhance the responsiveness of the predictive maintenance framework. Second, deep learning techniques, such as convolutional neural networks (CNNs) and transformer-based architectures, could be investigated to improve fault classification accuracy. Third, implementing hybrid AI models that combine traditional ML algorithms with physics-based models may enhance fault diagnosis reliability. Finally, real-world deployment and testing in operational power grids would provide valuable insights into model adaptability and scalability. These future directions will contribute to a more robust, real-time OLTC monitoring system, ensuring higher transformer reliability and reduced maintenance costs.

Acknowledgments

The authors would like to express their sincere gratitude for the publication support provided by Tenaga Nasional Berhad (TNB) through the TNB Seeding Fund (Project No. U-TD-RD-21-07), and by Universiti Tenaga Nasional under the BOLD Research Grant 2024 (BOLD2024), Grant No. J510051055.

References

1. Feizifar, B.; Muller, Z.; Fandi, G.; and Usta, Ö. (2019). A collective condition monitoring algorithm for on-load tap-changers. *Proceedings of the IEEE International Conference on Environment and Electrical Engineering and*

IEEE Industrial and Commercial Power Systems Europe (*EEEIC/I and CPS Europe*), Genova, Italy, 1-6.

2. Dabaghi-Zarandi, F.; Behjat, V.; Gauvin, M.; Picher, P.; Ezzaidi, H.; and Fofana, I. (2023). Power transformers OLTC condition monitoring based on feature extraction from vibro-acoustic signals: Main peaks and Euclidean distance. *Sensors*, 23(16), 7020.
3. Ismail, F.B. et al. (2022). An offline and online approach to the OLTC condition monitoring: A review. *Energies*, 15(17), 6435.
4. Bustamante, S.; Manana, M.; Arroyo, A.; Laso, A.; and Martinez, R. (2020). Determination of transformer oil contamination from the OLTC gases in the power transformers of a distribution system operator. *Applied Sciences*, 10(24), 8897.
5. Wajid, A. et al. (2023). Comparative performance study of dissolved gas analysis (DGA) methods for identification of faults in power transformer. *International Journal of Energy Research*, 2023(1), 9960743.
6. Soni, R.; and Mehta, B. (2022). Graphical examination of dissolved gas analysis by ratio methods and Duval triangle method to investigate internal faults of power transformer. *Materials Today: Proceedings*, 62, 7098-7103.
7. Ali, M.S.; Omar, A.; Jaafar, A.S.A.; and Mohamed, S.H. (2023). Conventional methods of dissolved gas analysis using oil-immersed power transformer for fault diagnosis: A review. *Electric Power Systems Research*, 216, 109064.
8. Bustamante, S.; Manana, M.; Arroyo, A.; Castro, P.; Laso, A.; and Martinez, R. (2019). Dissolved gas analysis equipment for online monitoring of transformer oil: A review. *Sensors*, 19(19), 4057.
9. Cheng, L.; and Yu, T. (2018). Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: A survey. *Energies*, 11(4), 913.
10. IEEE. (2009). IEEE guide for the interpretation of gases generated in oil-immersed transformers. IEEE Std C57.104-2019 (Revision of IEEE Std C57.104-2008), 1-98.
11. CIGRE. (2019). D1/A2.47: Advances in DGA interpretation. *Technical Brochure*, 771.
12. Mao, W.; Wei, B.; Xu, X.; Chen, L.; Wu, T.; Peng, Z.; and Ren, C. (2022). Fault diagnosis for power transformers through semi-supervised transfer learning. *Sensors*, 22(12), 4470.
13. Ghoneim, S.S.; and Taha, I.B. (2016). A new approach of DGA interpretation technique for transformer fault diagnosis. *International Journal of Electrical Power and Energy Systems*, 81, 265-274.
14. Nanfak, A.; Eke, S.; Kom, C.H.; Mouangue, R.; and Fofana, I. (2021). Interpreting dissolved gases in transformer oil: a new method based on the analysis of labelled fault data. *IET Generation*, *Transmission and Distribution*, 15(21), 3032-3047.
15. Guardado, J.L.; Naredo, J.L.; Moreno, P.; and Fuerte, C.R. (2001). A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis. *IEEE Transactions on Power Delivery*, 16(4), 643-647.

16. Miranda, V.; Castro, A.R.G.; and Lima, S. (2012). Diagnosing faults in power transformers with autoassociative neural networks and mean shift. *IEEE Transactions on Power Delivery*, 27(3), 1350-1357.
17. Wei, C.; Tang, W.; and Wu, Q. (2014). Dissolved gas analysis method based on novel feature prioritisation and support vector machine. *IET Electric Power Applications*, 8(8), 320-328.
18. Mansour, D.E.A. (2015). Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases. *IEEE Transactions on Dielectrics and Electrical Insulation*, 22(5), 2507-2512.
19. Odongo, G.; Musabe, R.; and Hanyurwimfura, D. (2021). A multinomial DGA classifier for incipient fault detection in oil-impregnated power transformers. *Algorithms*, 14(4), 128.
20. Huang, X.; Yuan, Y.; and Li, J. (2023). A review of transformer fault diagnosis based on information system theory and machine learning. *Preprints*, 2023050036.
21. Mosavi, A.; Ozturk, P.; and Chau, K.W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536.
22. Fan, C.; Chen, M.; Wang, X.; Wang, J.; and Huang, B. (2021). A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9, 652801.
23. Chawla, N.; Bowyer, K.; Hall, L.; and Kegelmeyer, W. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
24. Rajan, K. (2005). Materials informatics. *Materials Today*, 8(10), 38-45.
25. Alqudsi, A.; and El-Hag, A. (2019). Application of machine learning in transformer health index prediction. *Energies*, 12(14), 2694.
26. Mateus, B.C.; Farinha, J.T.; and Mendes, M. (2024). Fault detection and prediction for power transformers using fuzzy logic and neural networks. *Energies*, 17(2), 296.
27. Cichoń, A. and Włodarz, M. (2023). OLTC fault detection based on acoustic emission and supported by machine learning. *Energies*, 17(1), 220.
28. Cao, H.; Zhou, C.; Meng, Y.; Shen, J.; Xie, X. (2024). Advancements in transformer fault diagnosis technology. *Frontiers in Energy Research*, 12, 1437614.
29. Wang, R.; Dong, E.; Cheng, Z.; Liu, Z.; and Jia, X. (2024). Transformer-based intelligent fault diagnosis methods of mechanical equipment: A survey. *Open Physics*, 22(1), 20240015.
30. Zhang, Y.; Tang, Y.; Liu, Y.; and Liang, Z. (2022). Fault diagnosis of transformer using artificial intelligence: A review. *Frontiers in Energy Research*, 10, 1006474.
31. Zhou, A.Y.; and Farimani, A.B. (2024). FaultFormer: pretraining Transformers for bearing fault classification. *IEEE Access*, 12, 70719-70728.