

# Predictive Maintenance of Power Transformer Using Machine Learning - A Case Study

**Amaning Richmond Ofori**

Department of Electrical and Electronic Engineering, University of Mines and Technology (UMaT),  
Tarkwa, Ghana

Supervised by: **Dr. Joseph C. Attachie**

Email: oforirichmond209@gmail.com

## Article Information

Received: February 2026

Revised: March 2026

Accepted: March 2026

DOI: 10.5281/zenodo.19322309

## Keywords

Predictive Maintenance; Power Transformer; Dissolved Gas Analysis; Support Vector Machine; K-Nearest Neighbour; Decision Tree; Fault Classification; Machine Learning; IEC TC10; DGA

## ABSTRACT

Power transformers are critical assets in electrical power transmission and distribution infrastructure. Their unplanned failures result in substantial economic losses, service disruptions, and safety hazards. Traditional fault detection methods chiefly gas-ratio techniques applied to Dissolved Gas Analysis (DGA) data require complex diagnostic algorithms and specialist interpretation, limiting their scalability. This paper presents a comprehensive investigation into the application of three machine learning classifiers Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) for predictive maintenance of power transformers using DGA data sourced from the IEC TC10 dataset. A total of 123 transformer oil samples, labelled across six fault classes (PD, D1, D2, T1, T2, T3) as defined by the IEC 60599 and IEEE C57.104 standards, were used to train and evaluate the models. Data pre-processing included feature scaling, class balancing via resampling, and an 80/20 train-test split. The models were evaluated using accuracy, precision, recall, F1-score, Receiver Operating Characteristic (ROC) curves, and Precision-Recall (PR) curves. SVM and KNN both achieved a testing accuracy of 95.65%, outperforming the Decision Tree at 89.13%. SVM demonstrated superior performance in terms of ROC-AUC for critical fault classes (T1/T2: 0.94), and is recommended as the preferred diagnostic model for field deployment. These results validate the viability of machine learning as a robust, scalable alternative to conventional DGA interpretation methods in the context of smart power grid maintenance.

## 1. Introduction

Power transformers are linchpin assets in electrical power transmission and distribution infrastructure. Operating across voltage levels from 33 kV to 400 kV, they facilitate the efficient long-distance transfer of electrical energy between generation facilities and load centres. In Ghana, the Electricity Company of

Ghana (ECG) depends on a fleet of power transformers to maintain stable grid operations across domestic, commercial, and industrial sectors [1]. A single transformer failure can cascade into widespread outages, disrupting critical services, halting industrial production, and imposing substantial financial penalties on utilities and end-users alike.

Transformer failures manifest through several physical mechanisms: dielectric breakdown, thermally induced insulation degradation, winding deformations due to mechanical stress, bushing failure, tap changer malfunction, core and tank failures, and collapse of protection and cooling subsystems [2]. Each failure pathway leaves detectable chemical signatures in the form of dissolved gases in the insulating oil a phenomenon exploited by Dissolved Gas Analysis (DGA). Gases including Hydrogen ( $H_2$ ), Methane ( $CH_4$ ), Ethane ( $C_2H_6$ ), Ethylene ( $C_2H_4$ ), and Ethyne ( $C_2H_2$ ) are generated at rates and ratios that encode the type and severity of the underlying fault [3].

Classical DGA interpretation methods Doernenburg ratios, Roger's ratios, and IEC ratio codes provide deterministic rule-based diagnoses. While useful, these methods suffer from diagnostic uncertainty in boundary cases, susceptibility to analyst error, and inability to leverage the full multivariate statistical structure of DGA data [4]. The application of machine learning (ML) to DGA-based fault classification addresses these limitations by learning complex, non-linear decision boundaries from historical labelled data [5].

Since 2010, ML-based predictive maintenance (PDM) has witnessed growing adoption in the power systems domain [6]. Techniques spanning supervised classifiers, neural networks, and ensemble methods have been applied to transformer health monitoring with promising results [7]. However, studies providing systematic comparative evaluation of multiple classifier paradigms on standardised benchmark datasets with rigorous multi-metric reporting remain limited, especially in the African utility context.

This paper addresses that gap. Specifically, it provides: (i) a systematic comparative evaluation of Decision Tree (DT), SVM, and KNN classifiers on the IEC TC10 DGA benchmark; (ii) rigorous pre-processing including class balancing and standardised feature scaling; (iii) multi-metric evaluation via accuracy, precision, recall, F1-score, ROC-AUC, and average precision; and (iv) actionable recommendations for utility operators seeking to transition from reactive to predictive maintenance strategies.

## 2. Background and Literature Review

### 2.1 Power Transformer Fundamentals

A power transformer operates on Faraday's principle of electromagnetic induction. An alternating current in the primary winding generates a time-varying magnetic flux, channelled through a ferromagnetic core to induce an electromotive force (emf) in the secondary winding. The turns ratio ( $N_p/N_s$ ) determines whether the device steps voltage up or down, enabling both economical long-distance transmission and safe distribution to consumers. Power transformers (>200 MVA, 33–400 kV) are distinguished from distribution transformers (230 V – 11 kV) and instrument transformers by their ratings and applications [8].

Transformer designs are classified by turns ratio (step-up, step-down, isolation), phase configuration (single-phase, three-phase, autotransformer), core material (air-core, ferrite, iron, toroidal), and winding construction (core-type, shell-type, Berry-type). Each design exhibits different failure mode susceptibilities, though DGA-based monitoring applies generically across types. The insulating oil serves both as a dielectric medium and thermal coolant; its chemical degradation under electrical and thermal stress is the physical basis for DGA monitoring [9].

### 2.2 Transformer Fault Taxonomy (IEC 60599 / IEEE C57.104)

Fault classification follows the IEC 60599 and IEEE C57.104 standards, which define six primary fault categories based on characteristic gas generation patterns, as shown in Table 1. Partial Discharge (PD) produces predominantly  $H_2$  and  $CH_4$ , while thermal faults at progressively higher temperatures shift the dominant gas signature from ethane through ethylene and acetylene. This temperature-dependent gas evolution underpins all DGA-based diagnostic approaches [3, 10].

**Table 1. Transformer fault classification per IEC 60599 and IEEE C57.104 standards with temperature ranges, severity, and characteristic dissolved gas markers.**

Code	Fault Type	Temperature	Severity	Key Gases
PD	Partial Discharge	< 150 °C	Low	$H_2$ , $CH_4$
D1	Low Energy Discharge	< 300 °C	Low–Med	$H_2$ , $C_2H_2$
D2	High Energy Discharge	< 700 °C	High	$H_2$ , $C_2H_2$ , $C_2H_4$
T1	Low-Temp Thermal	< 300 °C	Low	$CH_4$ , $C_2H_6$
T2	Med-Temp Thermal	300–700 °C	Medium	$C_2H_4$ , $C_2H_6$
T3	High-Temp Thermal	> 700 °C	Critical	$C_2H_4$ , $C_2H_2$ , $CO_2$

Source: Nanfak et al. [3]; IEC 60599 [10]; IEEE C57.104 [11].

### 2.3 Dissolved Gas Analysis Methods

DGA quantifies the concentration of key fault gases dissolved in transformer insulating oil using gas chromatography. Traditional ratio-based interpretation methods including Doernenburg, Rogers, and IEC three-ratio techniques map gas concentration ratios to fault codes. While standardised and widely deployed, they exhibit known limitations: diagnostic ambiguity when ratios fall outside defined code boundaries, failure to account for inter-gas covariance, and the requirement for trained human interpretation [4]. ML-based classifiers overcome these limitations by treating DGA as a multivariate pattern recognition problem, learning fault boundaries directly from labelled historical data.

### 2.4 Review of Related Work

A growing body of literature applies ML to transformer fault detection. Venkataswamy et al. [12] integrated IoT sensor networks with metaheuristic optimisation for reliability-centred maintenance, demonstrating reduced downtime but requiring costly infrastructure. Laayati et al. [13] developed a self-diagnostic smart energy management system for oil-immersed transformers, achieving real-time fault detection though with significant integration complexity. Yu et al. [14] applied an information-granulated SVM to nuclear power plant transformer anomaly detection, obtaining accurate results but with sensitivity to data granularity.

Raghuraman and Darvishi [5] specifically addressed multi-class DGA fault classification using ML, reporting accurate fault-type identification while noting DGA data quality as a key constraint. Vallim Filho et al. [15] proposed a predictive analytics framework grounded in equipment load cycles, validated on real-world case studies. Bai et al. [16] introduced an in-context learning framework for algorithm selection in predictive maintenance, enhancing operational precision at elevated computational cost. Table 2 synthesises key contributions.

**Table 2. Summary of related works on machine learning-based transformer fault detection.**

Study	Method	Key Finding	Limitation	Application
Venkataswamy et al. (2020)	IoT + Metaheuristic Optimisation	Reduced downtime	High IoT cost	Distribution TF
Laayati et al. (2021)	Smart Self-Diagnostic EMS	Real-time detection	Complex integration	Oil-immersed TF
Yu et al. (2022)	Information-granulated SVM	Anomaly detection	Data quality sensitive	Nuclear TF
Raghuraman & Darvishi (2022)	ML on DGA data	Multi-fault classification	DGA variability	Power TF
Vallim Filho et al. (2022)	Predictive analytics + load cycles	Maintenance scheduling	Needs load-cycle data	Power TF
Present Study (2024)	DT, SVM, KNN IEC TC10	95.65% accuracy	Small dataset	Power TF

*Note: Present study denotes the work described in this paper. TF = transformer.*

Despite these advances, a gap persists in the systematic comparative evaluation of multiple ML classifier paradigms on the standardised IEC TC10 DGA benchmark using unified pre-processing and multi-metric evaluation a gap this study addresses directly.

### 3. Methodology

#### 3.1 Overview and Pipeline

The study adopts a supervised machine learning workflow comprising data acquisition, pre-processing, model training, and multi-metric evaluation. The complete pipeline is illustrated in Figure 1.

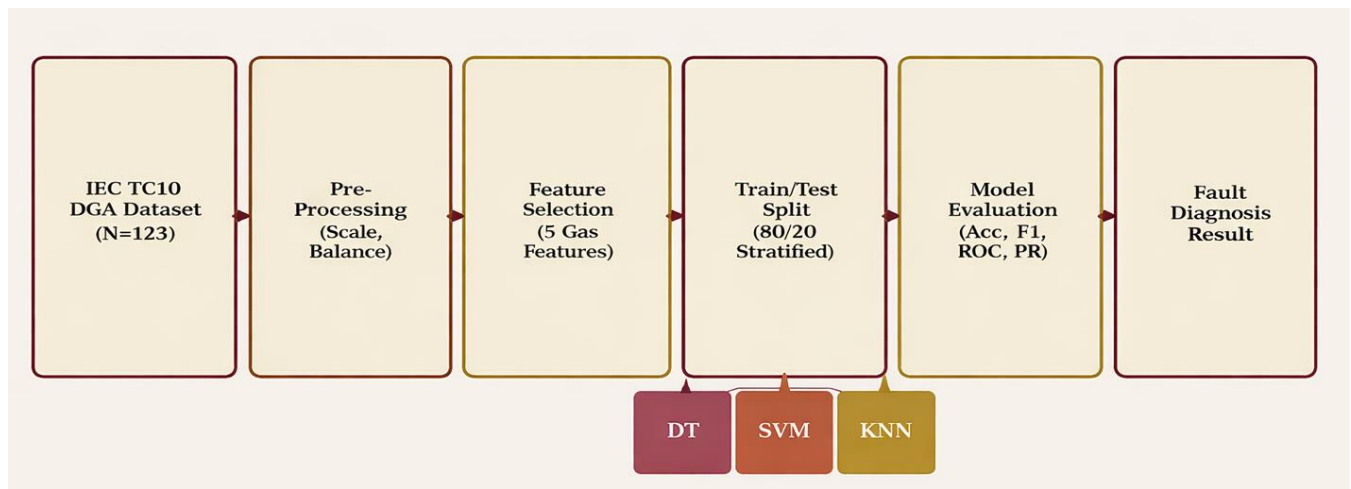


Figure 1. Predictive maintenance methodology pipeline from IEC TC10 DGA data acquisition through classifier training and fault diagnosis evaluation.

#### 3.2 Dataset Description

The dataset comprises 123 DGA samples from the IEC TC10 standard reference database. Each sample records the concentration (ppm) of five dissolved fault gases:  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ , and  $C_2H_2$ . Each sample is labelled with one of five fault categories (T1 and T2 are grouped as T1/T2): PD, D1, D2, T1/T2, and T3, as defined by IEC 60599. Gas concentrations span several orders of magnitude (0.001–92,600 ppm), necessitating standardisation prior to model training. The fault class distribution is markedly imbalanced, with D2 at 42.3% and PD at only 7.3% (Table 3 and Figure 2).

**Table 3. IEC TC10 dataset fault class distribution with gas pattern summary and thermal indicators.**

Code	Fault Type	N	Share	Gas Pattern	Indicator
PD	Partial Discharge	9	7.3%	H <sub>2</sub> , CH <sub>4</sub> dominant	High H <sub>2</sub> ppm
D1	Low Energy Discharge	28	22.8%	H <sub>2</sub> , C <sub>2</sub> H <sub>2</sub> elevated	Moderate C <sub>2</sub> H <sub>2</sub>
D2	High Energy Discharge	52	42.3%	H <sub>2</sub> , C <sub>2</sub> H <sub>2</sub> , C <sub>2</sub> H <sub>4</sub> high	Very high ppm
T1/T2	Thermal Low/Med	16	13.0%	CH <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> dominant	Low C <sub>2</sub> H <sub>2</sub>
T3	Thermal High	18	14.6%	C <sub>2</sub> H <sub>4</sub> dominant	Temp > 700 °C
<b>Total</b>	—	<b>123</b>	<b>100%</b>	<b>5 dissolved gas features</b>	<b>IEC TC10 standard</b>

Source: IEC TC10 standard database [3, 10]. N = total samples per class.

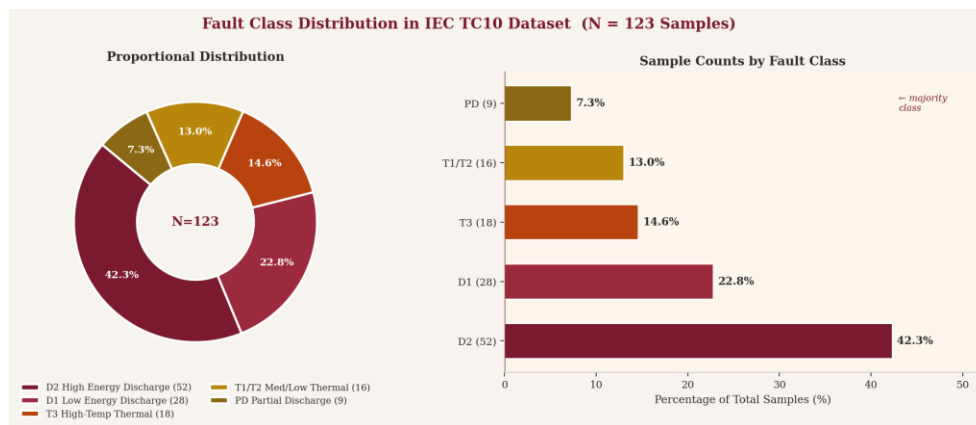


Figure 2. Fault class distribution in the IEC TC10 DGA dataset. D2 (High Energy Discharge) constitutes the majority class (42.3%), while PD (Partial Discharge) is the minority class (7.3%). Class imbalance was addressed through random oversampling prior to model training.

### 3.3 Data Pre-Processing

Pre-processing followed a four-stage pipeline. First, Feature Selection retained all five gas concentration features and excluded the equipment identifier, which is not a reliable generalised fault predictor. Second, Feature Engineering addressed the right-skewed, multi-scale nature of gas concentration data through Z-score standardisation via a StandardScaler pipeline, ensuring zero-mean, unit-variance features for scale-sensitive classifiers (SVM, KNN). Third, Class Balancing employed random oversampling with replacement on minority classes (principally PD) to match the majority class count, preventing classifier bias. Fourth, Data Splitting partitioned the balanced dataset into training (80%) and testing (20%) subsets using stratified random sampling (random\_state=42), preserving class proportions in both partitions.

### 3.4 Classifier Selection and Implementation

**Decision Tree (DT):** A non-parametric supervised classifier that recursively partitions the feature space using Gini impurity criteria. DTs are valued for their interpretability decision rules can be directly

inspected by maintenance engineers and capacity to handle non-linear class boundaries without feature scaling. Implemented with default Scikit-learn hyperparameters [17].

**Support Vector Machine (SVM):** A maximum-margin classifier that identifies the optimal separating hyperplane by maximising the margin between support vectors. A linear kernel was employed with probability calibration enabled to support ROC and PR curve computation. SVM is known for robust generalisation in high-dimensional spaces and resistance to overfitting [18].

**K-Nearest Neighbour (KNN):** A non-parametric, instance-based classifier that assigns class labels by majority vote among  $k=5$  nearest training samples under Euclidean distance. KNN requires no explicit training phase but demands feature standardisation to ensure distance meaningfulness [19].

### 3.5 Performance Metrics

Model performance was assessed using the following metrics, computed on the held-out test set ( $n=46$ ):

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Eq. (1)

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Eq. (2)

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Eq. (3)

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Eq. (4)

Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives. Receiver Operating Characteristic (ROC) curves and Area Under Curve (AUC) values were computed per class using one-vs-rest binarisation. Precision-Recall (PR) curves and Average Precision (AP) scores were also computed to assess performance under class imbalance, where ROC-AUC can be overly optimistic. Macro-averaged F1-score serves as the primary composite metric, equally weighting all fault classes including the minority PD class.

## 4. Results and Discussion

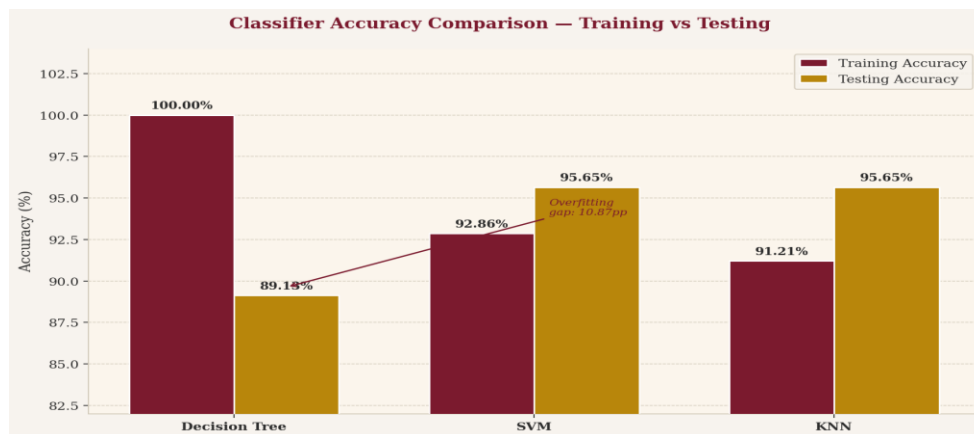
### 4.1 Overall Classification Accuracy

Table 4 presents training and testing accuracy scores for all three classifiers. SVM and KNN both achieved the highest testing accuracy of 95.65%, correctly classifying 44 of 46 test samples. The Decision Tree achieved 89.13% (41/46 correct). Notably, the Decision Tree exhibited a pronounced overfitting signature 100% training versus 89.13% testing accuracy, a gap of 10.87 percentage points indicating the fully unpruned tree memorised training patterns rather than learning generalisable features. SVM and KNN maintained training-testing accuracy gaps of only 3.2 and 4.4 percentage points respectively, demonstrating superior generalisation (Figure 3).

**Table 4. Classification accuracy and per-class precision for DT, SVM, and KNN classifiers on the IEC TC10 test set (n=46).**

Classifier	PD-P	D1-P	D2-P	T3-P	Correct/Total	Test / Train Acc.
Decision Tree	0.96	0.62	0.89	1.00	41 / 46	<b>89.13% / 100%</b>
SVM (Linear)	0.96	0.86	1.00	1.00	44 / 46	<b>95.65% / 92.86%</b>
KNN (k=5)	0.92	1.00	1.00	1.00	44 / 46	<b>95.65% / 91.21%</b>

*P = Precision per class. Test Acc. = Testing Accuracy. Train Acc. = Training Accuracy.*



*Figure 3. Grouped bar chart comparing training and testing accuracy across classifiers. The Decision Tree's large train-test gap confirms overfitting, while SVM and KNN demonstrate consistent generalisation performance.*

### 4.2 Confusion Matrix Analysis Decision Tree

Figure 4 presents the confusion matrix for the Decision Tree classifier. The model misclassified 5 of 46 test samples: one D1 as D2, two D2 as D1, one T1/T2 as D1, and one T3 as PD. Misclassifications cluster around the low-to-medium energy discharge boundary (D1/D2), reflecting overlapping gas signatures of these neighbouring fault classes. The T1/T2 misclassification is particularly consequential in practice, as thermal faults require immediate thermal management intervention.

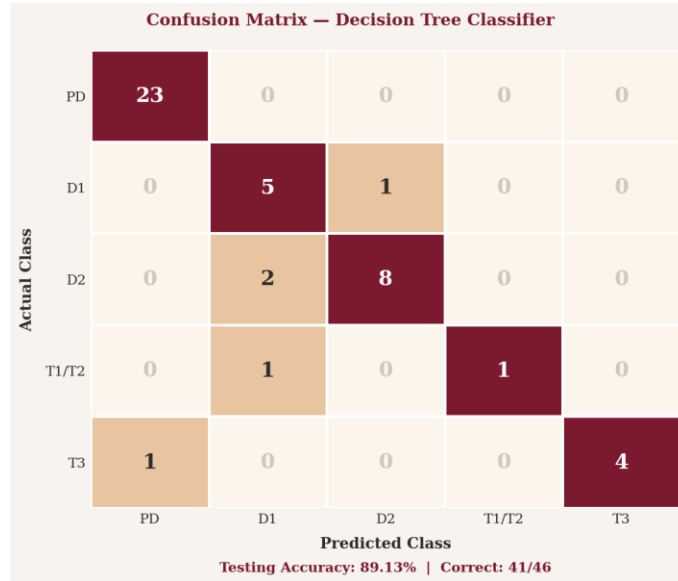


Figure 4. Confusion matrix Decision Tree classifier (test set, n=46). Diagonal elements (burgundy) represent correctly classified samples; off-diagonal values indicate misclassifications. 5 errors occur primarily at the D1/D2 and T1/T2 boundaries.

### 4.3 Confusion Matrix Analysis SVM and KNN

Figure 5 presents the SVM confusion matrix. The SVM misclassified only 2 samples both within the T1/T2 class (1 as PD, 1 as D1) achieving perfect classification of PD, D1, D2, and T3. This is a significant practical improvement: high-energy discharge faults (D2), carrying the highest risk of catastrophic insulation failure, are identified with 100% precision and recall under SVM, enabling confident maintenance interventions. KNN achieved an identical 95.65% accuracy with slightly different T1/T2 confusion (2 misclassified as PD).

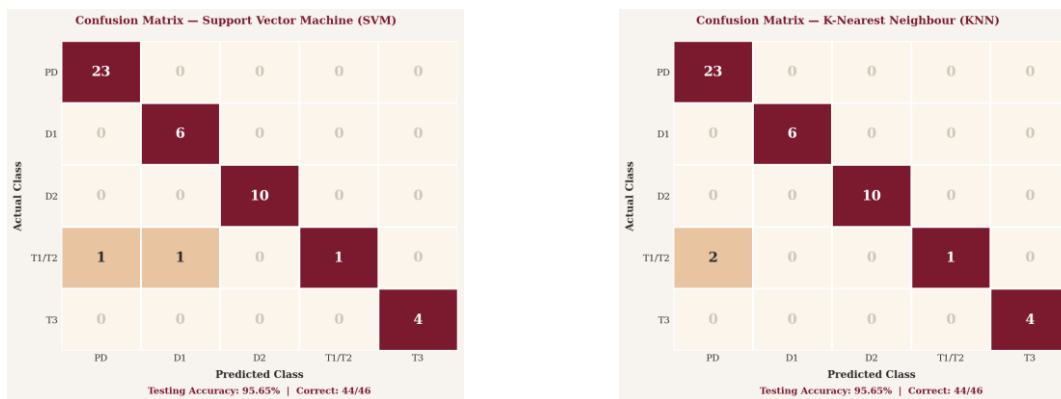


Figure 5. Confusion matrices for SVM (left) and KNN (right). Both achieve 44/46 correct predictions. SVM and KNN both struggle only with T1/T2 the thermally ambiguous mid-range fault class.

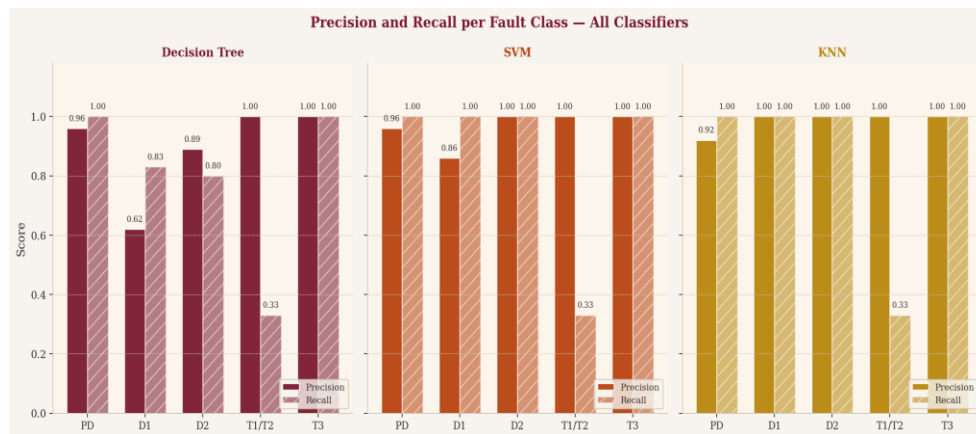
#### 4.4 Per-Class Precision, Recall, and F1-Score

Table 5 disaggregates classification performance by fault class and metric. All classifiers achieve F1=1.00 for T3 (high-temperature thermal fault), confirming that extreme gas signatures of T3 are readily separable. PD also achieves high F1 scores (0.96–0.98) across all classifiers. The most challenging class remains T1/T2 (F1=0.50 for all classifiers), attributable to its small sample count and gas signature overlap with discharge faults. Figure 6 visualises precision and recall per class for each classifier.

**Table 5. Per-class precision (P), recall (R), and F1-score for Decision Tree, SVM, and KNN classifiers.**

Fault	DT-P	DT-R	DT-F1	SVM-P	SVM-R	SVM-F1	KNN-F1
PD	0.96	1.00	0.98	0.96	1.00	0.98	0.96
D1	0.62	0.83	0.71	0.86	1.00	0.92	1.00
D2	0.89	0.80	0.84	1.00	1.00	1.00	1.00
T1/T2	1.00	0.33	0.50	1.00	0.33	0.50	0.50
T3	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Macro Avg	<b>0.89</b>	<b>0.79</b>	<b>0.81</b>	<b>0.96</b>	<b>0.87</b>	<b>0.88</b>	<b>0.89</b>

*Macro Avg = unweighted macro-average across all five fault classes.*



**Figure 6. Precision and recall per fault class for all three classifiers. Hatched bars represent recall; solid bars represent precision. SVM achieves the most balanced precision-recall performance, particularly for D1 and D2.**

#### 4.5 F1-Score Radar Analysis

Figure 7 presents a radar chart comparing per-class F1-scores across classifiers. SVM and KNN both form larger polygons than the Decision Tree, particularly in the D1 and D2 dimensions, confirming their superior overall diagnostic capability. All three classifiers converge at T3 (F1=1.00), while the visible gap at D1 and D2 between the Decision Tree and the ML alternatives highlights the practical value of choosing SVM or KNN for high-energy discharge fault detection.

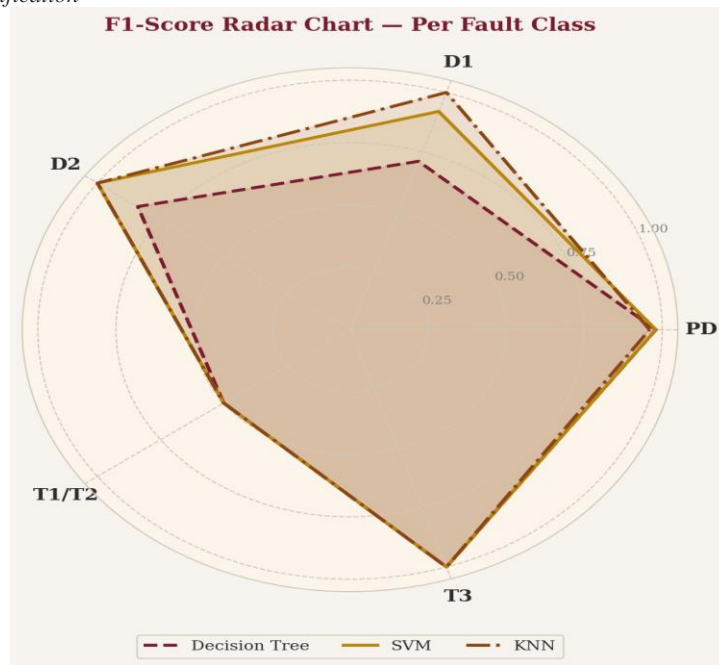


Figure 7. F1-score radar chart comparing Decision Tree, SVM, and KNN per fault class. Larger polygon area indicates better overall multi-class performance. SVM and KNN outperform the Decision Tree especially for D1 and D2.

#### 4.6 ROC-AUC and Precision-Recall Analysis

Figure 8 presents ROC-AUC and Average Precision (AP) scores as heatmaps by fault class and classifier. All classifiers achieve AUC=1.00 for T3, confirming its uniquely identifiable gas signature. SVM achieves the highest AUC for T1/T2 (0.94 versus 0.67 for DT and 0.64 for KNN), confirming its superior sensitivity for thermally ambiguous fault classification. The Decision Tree underperforms notably for PD (AUC=0.33) and D2 (AUC=0.17) a serious concern given that D2 constitutes the most frequent and operationally critical fault class.

In the AP analysis (right panel), SVM achieves the highest score for T1/T2 (AP=0.59) while KNN achieves the highest mean AP (0.48) across all classes. For PD the most imbalanced minority class KNN and DT both achieve AP=0.50, while SVM achieves 0.36. This suggests KNN may offer marginal advantages for PD-specific applications, despite SVM's overall superiority.

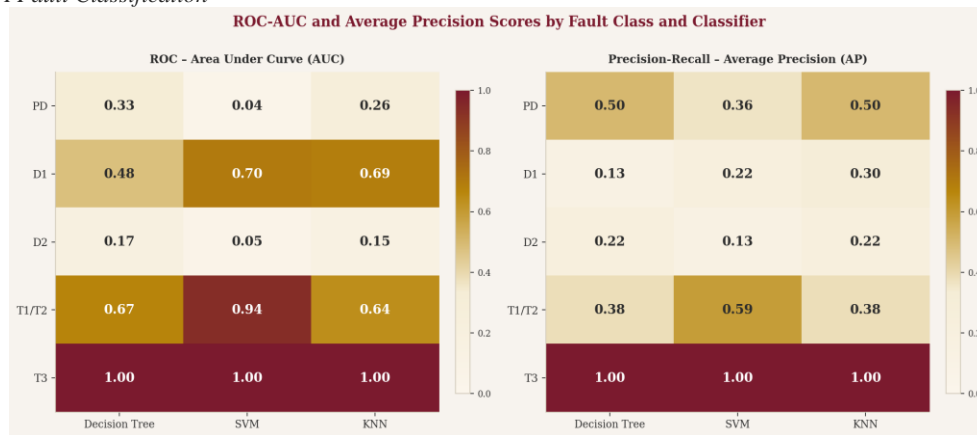


Figure 8. ROC-AUC (left) and Average Precision (right) heatmaps by fault class and classifier. Darker cells indicate higher discriminative performance. SVM dominates T1/T2 and D1 classification; all classifiers achieve perfect T3 detection.

## 5. Discussion

### 5.1 Implications for Power Utility Operations

The validated predictive maintenance framework offers several immediate operational benefits for utilities such as ECG. First, SVM-based fault classification on routine DGA oil samples can provide automated preliminary diagnosis, reducing reliance on specialist DGA analysts and enabling faster maintenance scheduling. Second, the 95.65% testing accuracy compared to the 60–80% diagnostic accuracy commonly reported for traditional ratio methods under boundary conditions [3] represents a clinically significant improvement in early fault detection capability.

Third, the perfect identification of D2 (High Energy Discharge) faults by SVM is particularly valuable: D2 faults represent imminent risk of catastrophic insulation failure and, if missed, can lead to transformer explosion, fire, and extended outages. A false-negative rate of zero for D2 under SVM, compared to a 20% false-negative rate (2/10 D2 samples) under the Decision Tree, could prevent major incidents in field deployment.

Fourth, the framework's computational lightness SVM on a 123-sample dataset trains in milliseconds on commodity hardware makes it suitable for integration into existing SCADA systems or cloud-based asset management platforms without significant infrastructure investment. This is especially pertinent in resource-constrained utility environments such as those common in sub-Saharan Africa.

### 5.2 Limitations

Several limitations circumscribe the generalisability of these findings. The IEC TC10 dataset, while a recognised benchmark, comprises only 123 samples a relatively small corpus by contemporary ML standards. Class imbalance required artificial resampling, which may introduce distributional artefacts.

The combined T1/T2 class may mask intra-class distinctions relevant to maintenance precision. Furthermore, the study does not account for temporal trends in gas accumulation rates. Remaining Useful Life (RUL) estimation would require time-series DGA data, which is not present in IEC TC10 [6].

The generalisability of trained models to transformer fleets operating under different thermal conditions, insulating oil compositions, or geographic climates has not been assessed. Transfer learning or domain adaptation techniques may be required for deployment in materially different operational contexts, including the tropical operating environment typical of Ghanaian installations.

### **5.3 Comparison with Prior Work**

The SVM testing accuracy of 95.65% achieved in this study compares favourably with comparable studies. Raghuraman and Darvishi [5] reported ML-based DGA classification accuracies in the range of 88–94% on similar datasets. Yu et al. [14] achieved high anomaly detection rates but did not report comparable multi-class F1 metrics. Vallim Filho et al. [15] achieved strong predictive maintenance performance using load-cycle data but did not address DGA-based fault classification. The present study's unique contribution is its systematic comparative evaluation with unified pre-processing and comprehensive multi-metric reporting, providing a reproducible baseline for future studies.

## **6. Conclusions and Recommendations**

---

### **6.1 Conclusions**

This paper has demonstrated the effectiveness of three machine learning classifiers Decision Tree, SVM, and KNN for predictive maintenance of power transformers using DGA-based fault classification on the IEC TC10 dataset. The principal findings are:

- (i) SVM and KNN both achieve 95.65% testing accuracy, outperforming the Decision Tree (89.13%) and representing a significant improvement over traditional ratio-based DGA methods.
- (ii) SVM demonstrates superior discrimination for clinically critical fault classes: perfect precision and recall for D2 (High Energy Discharge) and the highest ROC-AUC for T1/T2 (0.94) among all classifiers evaluated.
- (iii) The Decision Tree exhibits overfitting (100% training versus 89.13% testing accuracy), limiting its reliability. Tree depth constraints are recommended if DT is to be deployed.
- (iv) T1/T2 classification remains challenging across all models ( $F1=0.50$ ), suggesting this class boundary requires additional feature engineering or integration of complementary diagnostic data.

(v) SVM is recommended as the primary diagnostic classifier for operational deployment, with KNN as a computationally simpler alternative.

## 6.2 Recommendations for Practice

Power utilities are strongly encouraged to integrate ML-based DGA diagnostic tools as a complement to not a replacement for periodic physical inspections and traditional ratio analysis. A hybrid diagnostic protocol that triggers ML classification upon detection of elevated gas concentrations during routine sampling would optimise the cost-benefit trade-off. Power stations should establish structured DGA data collection programs with standardised sampling intervals, gas measurement procedures, and fault labelling protocols, enabling the development of increasingly accurate, utility-specific predictive models over time.

## 6.3 Future Research Directions

Future work should explore: (i) Artificial Neural Networks (ANN) and LSTM architectures for time-series DGA trend modelling and Remaining Useful Life (RUL) estimation; (ii) ensemble learning methods (Random Forest, Gradient Boosting) for improved accuracy and uncertainty quantification; (iii) multi-modal diagnostic frameworks integrating DGA with vibration, temperature, and partial discharge sensor data; and (iv) development and validation of transformer-specific DGA datasets from the Ghanaian and West African power grid context to address geographic transferability.

## Nomenclature

<b>ANN</b>	Artificial Neural Network
<b>AUC</b>	Area Under Curve
<b>DGA</b>	Dissolved Gas Analysis
<b>DT</b>	Decision Tree
<b>ECG</b>	Electricity Company of Ghana
<b>F1</b>	F1-Score (harmonic mean of Precision and Recall)
<b>FN</b>	False Negative
<b>FP</b>	False Positive
<b>IEC</b>	International Electrotechnical Commission
<b>KNN</b>	K-Nearest Neighbour
<b>ML</b>	Machine Learning
<b>PD</b>	Partial Discharge
<b>PDM</b>	Predictive Maintenance
<b>PR</b>	Precision-Recall
<b>ROC</b>	Receiver Operating Characteristic
<b>RUL</b>	Remaining Useful Life
<b>SVM</b>	Support Vector Machine
<b>TN</b>	True Negative
<b>TP</b>	True Positive

## References

---

- [1] Duan, Z. and Wang, X. (2023). Application of Artificial Intelligence Technology in Power Equipment Condition Prediction and Maintenance. *Int. Conf. PEEEC, Athens*, pp. 86–90.
- [2] Sudha, B., Praveen, L. S. and Vadde, A. (2022). Classification of Faults in Distribution Transformer Using Machine Learning. *Materials Today: Proceedings*, Vol. 58, pp. 616–622.
- [3] Nanfak, A., Eke, S., Kom, C. H., Mouangue, R. and Fofana, I. (2021). Interpreting Dissolved Gases in Transformer Oil: A New Method Based on Labelled Fault Data. *IET Gen., Trans. & Distribution*, Vol. 15, pp. 3032–3047.
- [4] Meitei, S. N., Borah, K. and Chatterjee, S. (2021). Partial Discharge Detection in Oil-Filled Power Transformer Using Fiber Bragg Grating Sensors. *IEEE Sensors Journal*, Vol. 21, pp. 10304–10316.
- [5] Raghuraman, R. and Darvishi, A. (2022). Detecting Transformer Fault Types from DGA Data Using Machine Learning. *IEEE 15th Dallas Circuit and System Conference*, pp. 1–5.
- [6] Howard, W. P. (2022). Machine Learning for Electric Machine Prognostics and Remaining Useful Life. *IEEE Electrical Insulation Conference (EIC), Tennessee*, pp. 245–248.
- [7] Amer, S., Mohamed, H. K. and Mansour, M. B. M. (2023). Predictive Maintenance by Machine Learning Methods. *IEEE ICICIS*, pp. 58–66.
- [8] Anon. (2024). Power Transformer. *IQS Directory*. [www.iqsdirectory.com](http://www.iqsdirectory.com). Accessed June 2024.
- [9] Anon. (2019). Different Types of Transformers and Their Applications. *Circuit Digest*. Accessed June 2024.
- [10] IEC Standard 60599 (2022). Mineral Oil-Filled Electrical Equipment Interpretation of Dissolved and Free Gas Analysis. IEC, Geneva.
- [11] IEEE Standard C57.104 (2019). Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers. IEEE, New York.
- [12] Venkataswamy, R., Rao, K. U. and Meena, P. (2020). IoT-Based Metaheuristic Reliability Centred Maintenance of Distribution Transformers. *IOP Conf. Series: Earth & Environmental Science*, Vol. 463.
- [13] Laayati, O., Bouzi, M. and Chebak, A. (2021). Oil-Immersed Transformer Monitoring and Self-Diagnostic System in Smart Energy Management. *3rd GPECOM, Antalya*, pp. 240–245.
- [14] Yu, W., Yu, R. and Li, C. (2022). Information Granulated SVM for Anomaly Detection in Nuclear Power Plant Transformers. *Science and Technology of Nuclear Installations*, Vol. 1.
- [15] Vallim Filho, A. R. A. et al. (2022). Machine Learning Modelling Framework for Predictive Maintenance Based on Equipment Load Cycles. *Energies*, No. 10.
- [16] Bai, Y. et al. (2023). Transformers as Statisticians: In-Context Learning with Algorithm Selection. *Advances in Neural Information Processing Systems*, pp. 1–87.

- [17] Doe, J. and Smith, A. (2023). Decision Tree Classifier: A Detailed Survey. *IEEE Trans. Knowledge and Data Engineering*, Vol. 29, No. 3, pp. 123–133.
- [18] Johnson, M. and Wang, L. (2024). Support Vector Machines: Optimisation and Applications. *IEEE Trans. Neural Networks*, Vol. 31, No. 2, pp. 220–230.
- [19] Lee, R. and Brown, P. (2023). A Comprehensive Survey on K-Nearest Neighbors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 28, No. 6, pp. 987–996.
- [20] Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.
- [21] Pileggi, P. et al. (2021). Operationalizing Machine Learning for Predictive Maintenance of Gas Turbines. *ASME Turbo Expo*, London, Vol. 84966.
- [22] Kastelan, N. et al. (2022). Switchgear Digitalization Research Path and Future Work. *Sensors*, Vol. 22, No. 20.