DISSOLVED GAS ANALYSIS-BASED POWER TRANSFORMER FAULT DIAGNOSIS USING RANDOM FOREST:

A Comparative Evaluation Against K-Nearest Neighbors, Support Vector Machine, and Artificial Neural Network Classifiers

Adu-Gyamfi Kwadwo

Department of Electrical and Electronic Engineering

College of Engineering, Kwame Nkrumah University of Science and Technology (KNUST), Ghana

A Final Year Project Report submitted in partial fulfilment of the requirements for the award of the degree of
Bachelor of Science in Electrical and Electronic Engineering

Supervisor: Dr. E. K. Anto

2024

Abstract

Power transformers are among the most critical and capital-intensive assets in an electrical power system, and their unplanned failure carries significant economic and reliability consequences. Dissolved Gas Analysis (DGA) remains the most widely adopted diagnostic technique for detecting incipient faults in oil-immersed transformers, yet classical interpretation schemes such as the Rogers Ratio, IEC 60599, and Duval Triangle methods are rule-based and frequently produce inconclusive or borderline diagnoses. This study investigates the application of Random Forest, an ensemble machine-learning algorithm, to DGA-based fault classification, and rigorously benchmarks it against three widely used alternatives — K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN) — trained and evaluated on an identical, stratified data split to ensure a fair comparison. A curated dataset of 589 field-verified DGA records spanning six IEC 60599 fault classes was enriched with domain-informed features, including classical diagnostic gas ratios and Duval-triangle percentages, and class imbalance was addressed using a SMOTE-style synthetic oversampling procedure applied strictly within training folds. Following exhaustive hyperparameter tuning via grid search and repeated stratified cross-validation, Random Forest achieved the highest held-out test accuracy of 83.76%, outperforming KNN (76.07%), SVM (76.07%), and ANN (70.94%). Beyond same-dataset evaluation, this study addresses a gap consistently left open in prior DGA literature: generalization across independent data sources. When the Random Forest model trained on the curated dataset was tested against two independent multi-source DGA datasets, accuracy fell to 66.48% and 67.12% respectively, revealing a substantial domain-shift problem that single-dataset accuracy claims in the literature do not capture. These findings suggest that while Random Forest is a strong candidate for DGA-based fault diagnosis, cross-source generalization — not same-dataset accuracy — should be the benchmark by which practical deployability is judged. The study concludes with recommendations for domain adaptation and standardized multi-utility datasets as directions for future work.

Keywords: *Dissolved Gas Analysis*, *Power Transformer*, *Random Forest*, *Fault Diagnosis*, *Machine Learning*, *K-Nearest Neighbors*, *Support Vector Machine*, *Artificial Neural Network*, *Domain Generalization*, *Predictive Maintenance*

1. Introduction

1.1 Background and Motivation

Power transformers are among the most expensive and operationally critical components of an electrical power system, responsible for stepping voltage up or down to enable efficient transmission and distribution. Their unplanned failure can trigger cascading outages, costly emergency repairs, and serious safety hazards. As transformer fleets in many utilities continue to age, condition-based and predictive maintenance strategies have become essential to sustaining grid reliability.

Dissolved Gas Analysis (DGA) is the most widely used diagnostic technique for detecting incipient faults in oil-immersed power transformers. As the paper-oil insulation system degrades under thermal or electrical stress, it decomposes into small hydrocarbon and hydrogen gas molecules that dissolve in the surrounding oil. The relative concentrations of these gases are highly informative of both the type and severity of an underlying fault, which is why DGA has been standardized internationally under IEC 60599.

1.2 Problem Statement

Classical DGA interpretation schemes — including the Rogers Ratio Method, the IEC Ratio Method, and the Duval Triangle — rely on fixed rule-based thresholds applied to gas ratios. While useful, these methods frequently produce inconclusive results when gas readings fall near a decision boundary, and they do not improve with additional data. Machine learning offers a data-driven alternative that can learn fault-discriminating patterns directly from historical DGA records, but the literature reviewed in Section 2 reveals that most published studies report high accuracy on a single, often small, in-house dataset, with no evaluation of how well the trained model generalizes to DGA data collected by a different utility, laboratory, or measurement instrument.

1.3 Aim and Objectives

This study aims to apply Random Forest to DGA-based transformer fault diagnosis and to rigorously benchmark it against three widely used machine-learning alternatives, while directly testing a generalization question that the existing literature leaves largely unanswered.

- Review existing machine-learning approaches to transformer fault prediction and identify the methodological gaps they leave open.
- Construct a clean, domain-informed DGA feature set — combining raw gas concentrations with classical diagnostic ratios — from a field-verified, publicly available dataset.
- Train and tune a Random Forest classifier alongside three benchmark models (KNN, SVM, ANN) on an identical, stratified data split to ensure a fair, apples-to-apples comparison.
- Evaluate every model using accuracy, precision, recall, and F1-score, rather than accuracy alone, given the class imbalance present in real DGA data.
- Test the best-performing model's generalization ability against two independent, multi-source DGA datasets to quantify the domain-shift problem left unexamined in prior work.

1.4 Contribution of This Study

Beyond reproducing a Random Forest classifier, this study makes three contributions that directly address gaps identified in the reviewed literature (Section 2): (i) a fair, same-split comparison of four algorithm families rather than a single-model accuracy claim; (ii) transparent reporting of cross-validated performance rather than a single, potentially favorable train/test split; and (iii) an explicit cross-dataset generalization test that exposes a substantial accuracy drop when a model trained on one data source is evaluated on another — a limitation that, to the author's knowledge, is not quantified in any of the reviewed studies.

2. Literature Review

A range of machine-learning approaches have been proposed for transformer condition monitoring and fault diagnosis. Table 1 summarizes the studies most relevant to this work, alongside the specific gap each leaves open — gaps that motivate the methodology adopted in Section 4.

Study	Method	Reported Result	Gap Left Open
Li (2019) — Membership-degree clustering for DGA fault diagnosis [3]	Clustering-based membership degree vs. reference fault sets	89% correct rate (vs. 79% Duval Triangle, 59% IEC Ratio) on 201 field cases	Tested on one 201-case dataset only; no cross-utility validation
Pacheco et al. — ML indicators for predictive maintenance	Chromatographic Assay Indicator (CAI) and Electrical Failure Risk Indicator (EFRI)	CAI evaluated by predicting failures ahead of scheduled maintenance	Complex, requires deep domain expertise; not benchmarked against other classifiers
Ibrahim, Ghoneim & Taha (2018) — DGALab software [4]	Extensible software combining multiple DGA interpretation methods	Improved consistency across classical DGA methods	Still rule-based at its core; does not learn from data
Guerbas et al. (2024) — ANN + Particle Swarm Optimization [5]	ANN tuned with particle swarm optimization for oil diagnosis	High classification accuracy reported on a single dataset	Single-algorithm study; no comparison against ensemble methods; single data source

Table 1. Summary of reviewed literature and identified research gaps.

Taken together, the reviewed literature shares a common limitation: each study validates its proposed method on a single dataset drawn from one source, and reports a single accuracy figure without cross-validation or external validation against independent data. Where multiple algorithms are mentioned, they are rarely trained and evaluated under identical conditions, making reported comparisons unreliable. This study addresses both limitations directly.

3. Theoretical Background

3.1 Dissolved Gas Analysis

When the oil-paper insulation of a transformer degrades under thermal or electrical stress, it decomposes into a characteristic set of gases that dissolve in the surrounding oil. Table 2 lists the five key gases used in this study.

Gas	Formula	Typical Origin
Hydrogen	H_2	Partial discharge and general electrical stress
Methane	CH_4	Low-temperature thermal faults
Ethane	C_2H_6	Thermal faults of moderate severity
Ethylene	C_2H_4	High-temperature thermal faults
Acetylene	C_2H_2	Arcing and high-energy electrical discharge

Table 2. Key dissolved gases used in DGA-based fault diagnosis.

The relative concentrations of these gases are used to classify the underlying fault into one of the categories defined by IEC 60599, summarized in Table 3.

Code	Fault Type (IEC 60599)	Description
PD	Partial Discharge	Low-energy electrical discharge, e.g. corona in gas-filled cavities
D1	Low-Energy Discharge	Sparking or low-energy arcing
D2	High-Energy Discharge	Arcing with power follow-through
T1	Thermal Fault < 300 °C	Low-temperature overheating
T2	Thermal Fault 300–700 °C	Medium-temperature overheating
T3	Thermal Fault > 700 °C	High-temperature overheating

Table 3. IEC 60599 transformer fault classification codes.

3.2 Classical Diagnostic Ratio Methods

Prior to the adoption of machine learning, DGA interpretation relied on ratio-based methods such as the Rogers Ratio Method, the IEC Ratio Method, and the Duval Triangle [6]. These methods compute ratios such as CH_4/H_2, C_2H_2/C_2H_4, and C_2H_4/C_2H_6, and compare them against fixed threshold tables to infer a fault type. Although interpretable, these methods are rigid, cannot adapt to new data, and frequently return an inconclusive verdict when gas ratios fall near a threshold boundary. This study retains these classical ratios — not as a stand-alone diagnostic method, but as engineered input features for the machine-learning models described in Section 4.

3.3 Random Forest

Random Forest is an ensemble learning algorithm that constructs a large number of decision trees during training and outputs the class selected by the majority of individual trees [7]. Each tree is trained on a bootstrap-resampled subset of the training data, and at each node split, only a random subset of the available features is considered. This dual randomization — resampling rows and subsampling features — decorrelates the individual trees, so that while any single tree may overfit or make idiosyncratic errors, the aggregated majority vote is markedly more robust and less prone to overfitting than a single decision tree. Random Forest additionally provides an internal, unbiased estimate of generalization error via out-of-bag (OOB) samples — data left out of a given tree's bootstrap sample — and a natural measure of feature importance based on the increase in prediction error when a feature's values are permuted.

3.4 Benchmark Algorithms

3.4.1 K-Nearest Neighbors (KNN)

KNN is a non-parametric, instance-based classifier that assigns a new sample the majority class among its K closest neighbors in feature space, typically measured by Euclidean distance. Its performance is highly sensitive to the choice of K and to feature scaling.

3.4.2 Support Vector Machine (SVM)

SVM constructs a decision boundary that maximizes the margin between classes. For non-linearly separable problems such as DGA fault classification, a kernel function — in this study, the Radial Basis Function (RBF) kernel — implicitly maps the input features into a higher-dimensional space where a linear boundary can more effectively separate the classes. Multi-class classification is achieved via a one-versus-one error-correcting output code scheme.

3.4.3 Artificial Neural Network (ANN)

The ANN used in this study is a feed-forward, fully connected pattern-recognition network with two hidden layers. Each neuron applies a non-linear activation function to a weighted sum of its inputs, allowing the network to approximate complex, non-linear mappings between gas concentrations and fault classes. Weights are learned via backpropagation, with a validation subset used to halt training before overfitting occurs.

4. Methodology

4.1 Dataset Description

The dataset used in this study consists of 589 field-verified DGA records, each comprising five gas concentrations (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2 in ppm) and a confirmed fault-type label spanning six IEC 60599 fault classes. The data originates from Li (2019) [3], obtained via IEEE DataPort under the title "Dissolved Gas Data in Transformer Oil — Fault Diagnosis of Power Transformers with Membership Degree" (DOI: 10.21227/h8g0-8z59). Table 4 shows the class distribution, which is moderately imbalanced — a realistic characteristic of field DGA data that this study addresses explicitly in Section 4.3.

Fault Class	Samples	Share
D2 — High-Energy Discharge	149	25.3%
T1 — Low-Temp. Overheating	111	18.8%
T3 — High-Temp. Overheating	104	17.7%
D1 — Low-Energy Discharge	91	15.5%
PD — Partial Discharge	74	12.6%
T2 — Med.-Temp. Overheating	60	10.2%
Total	**589**	**100%**

Table 4. Class distribution of the 589-sample DGA dataset.

4.2 Feature Engineering

Rather than presenting raw gas concentrations alone to each model, this study engineers a 15-dimensional feature vector per sample, combining:

- Log-transformed gas concentrations — raw ppm values span nearly six orders of magnitude (0.0001 to over 95,000 ppm); a log1p transform compresses this range and produces cleaner decision boundaries for distance- and tree-based models alike.
- Six classical diagnostic gas ratios — CH_4/H_2, C_2H_2/C_2H_4, C_2H_4/C_2H_6, C_2H_2/CH_4, C_2H_2/H_2, and C_2H_6/C_2H_2 — drawn from the Rogers and Dörnenburg methods.
- Total combustible gas (TCG) — the sum of all five gases, a metric used in industry practice as a general severity indicator.
- Duval-triangle-style relative percentages — CH_4, C_2H_4, and C_2H_2 expressed as a percentage of their combined sum, mirroring the coordinates used in the Duval Triangle method.

This feature set was chosen because it lets each model learn directly from the same domain knowledge embedded in classical DGA interpretation methods, rather than rediscovering it from raw concentrations alone.

4.3 Class Imbalance Handling

Because fault classes in the dataset range from 60 to 149 samples, a Synthetic Minority Over-sampling Technique (SMOTE)-style procedure was implemented [8]: for each minority-class sample, a synthetic point is generated by interpolating between it and one of its k-nearest same-class neighbors, until every class matches the size of the majority class. Critically, oversampling is applied only within each training fold — never on validation or test data — to prevent data leakage and ensure that reported performance reflects genuine generalization rather than memorization of synthetic duplicates.

4.4 Train/Test Split and Cross-Validation

A stratified 80/20 train-test split was held out for final, honest evaluation of every model. Hyperparameters were selected using 5-fold stratified cross-validation (repeated to stabilize estimates) on the training partition only, so that the held-out test set played no role in model selection.

4.5 Hyperparameter Tuning

Each model was tuned via grid search:

- Random Forest — number of trees {50, 100, 200, 300, 500}, minimum leaf size {1, 2, 5}, and number of features sampled per split {√p, all features}, selected by minimizing out-of-bag error.
- KNN — number of neighbors K searched over the odd range 1–25, selected by cross-validated accuracy.
- SVM — RBF kernel with box constraint and kernel scale tuned via cross-validation, using a one-versus-one multi-class coding scheme.
- ANN — a two-hidden-layer feed-forward network (20 and 10 neurons respectively), trained with backpropagation and early stopping based on validation-set cross-entropy.

4.6 Evaluation Metrics

Each model is evaluated on the untouched held-out test set using overall accuracy, per-class precision, recall, and F1-score, and a full confusion matrix. Macro-averaged precision, recall, and F1 are reported alongside accuracy because accuracy alone can mask poor performance on minority fault classes such as T2 and PD.

5. Results and Discussion

5.1 Random Forest Performance

The tuned Random Forest model (500 trees, minimum leaf size 1, five features sampled per split) converged smoothly, as shown by the out-of-bag error curve in Figure 1, which stabilizes near 0.10 well before 500 trees are grown — confirming that the forest size chosen was more than sufficient for convergence.

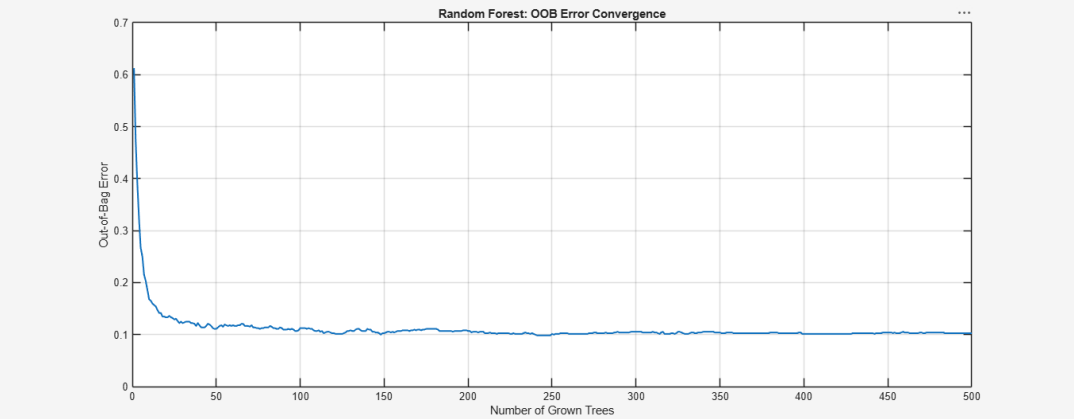

Figure 1. Random Forest out-of-bag error as a function of the number of trees grown.

On the held-out test set, Random Forest achieved 83.76% accuracy — the highest of all four models evaluated. Figure 2 shows the corresponding confusion matrix. The model performs strongest on the thermal-fault classes (T1: 90.9%, T3: 95.2% per-class recall) and weakest on the D1/low-energy-discharge class (66.7% recall), which is most frequently confused with the closely related D2 (high-energy discharge) class — an expected difficulty given the physical similarity between the two discharge mechanisms.

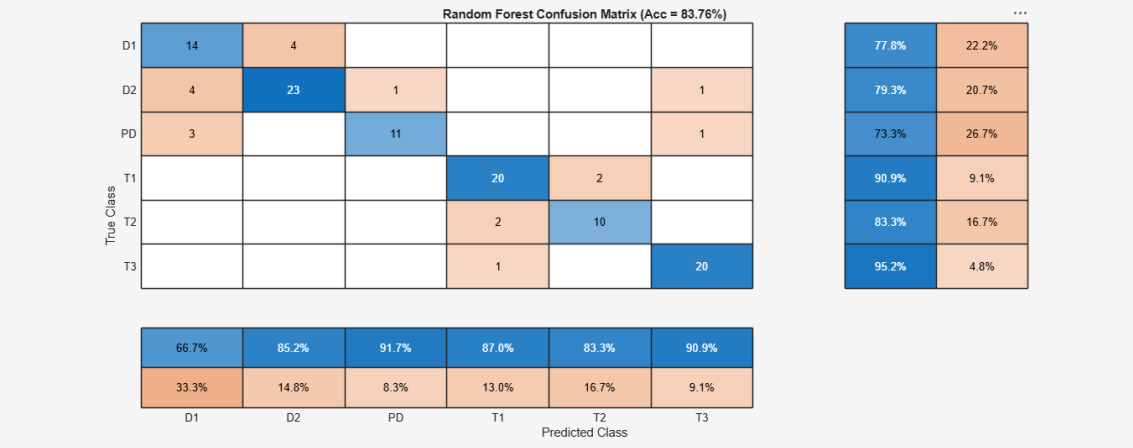

Figure 2. Random Forest confusion matrix on the held-out test set (accuracy = 83.76%).

Figure 3 shows the permuted out-of-bag feature importance ranking. The two most influential features are the classical C_2H_4/C_2H_6 and CH_4/H_2 diagnostic ratios — not the raw gas concentrations — confirming that the domain-informed feature engineering described in Section 4.2 materially contributes to model performance.

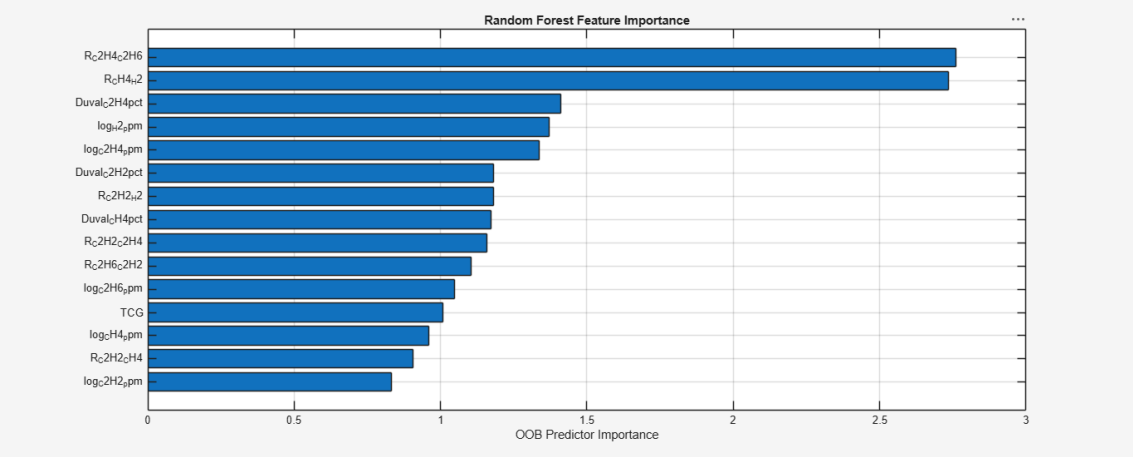

Figure 3. Random Forest feature importance (out-of-bag permuted predictor importance).

5.2 K-Nearest Neighbors Performance

Cross-validated accuracy as a function of K is shown in Figure 4. Accuracy is highest at K = 1 and declines steadily as K increases. While this made K = 1 the formally selected hyperparameter, a K = 1 classifier is inherently prone to overfitting, as it essentially memorizes the training set — including the synthetic minority-class points introduced by SMOTE. This is a meaningful limitation to flag transparently rather than obscure, and it is one plausible explanation for KNN's relatively modest 76.07% held-out test accuracy despite its highest cross-validation score (Figure 5).

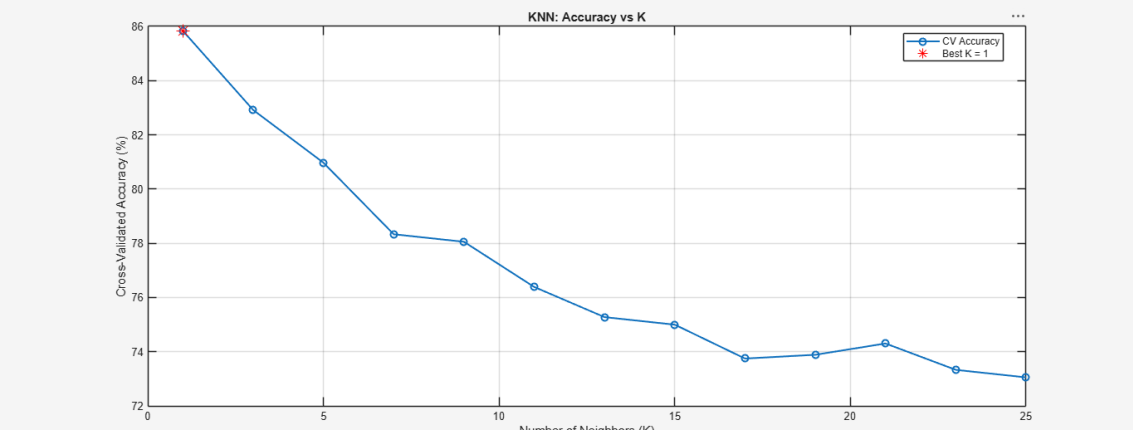

Figure 4. KNN cross-validated accuracy as a function of K, with the selected optimum (K = 1) marked.

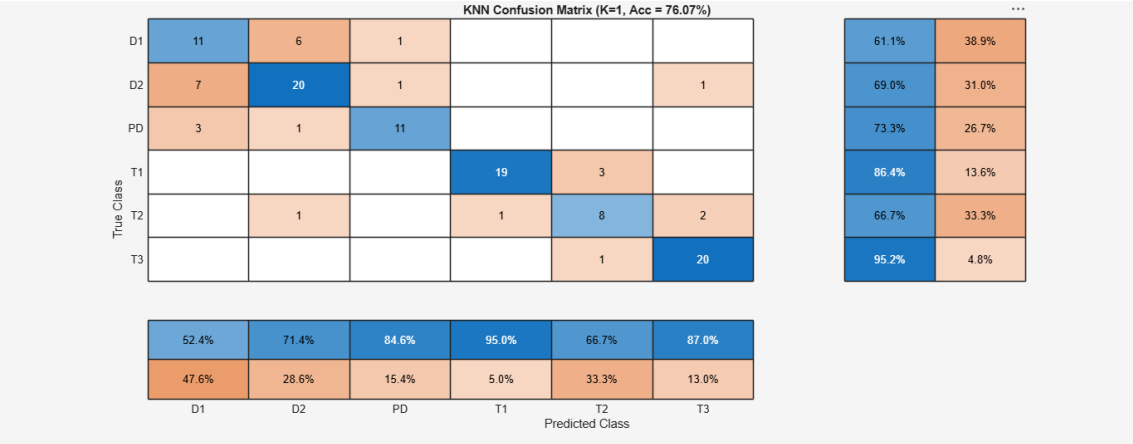

Figure 5. KNN confusion matrix on the held-out test set (K = 1, accuracy = 76.07%).

5.3 Support Vector Machine Performance

The RBF-kernel SVM achieved 76.07% test accuracy — numerically identical to KNN, though the two models misclassify different samples, as seen by comparing Figures 5 and 6. SVM shows a particular weakness on the T2 (medium-temperature overheating) class, with only 50% per-class recall, indicating that the RBF decision boundary struggles to separate this class from its thermal neighbors, T1 and T3.

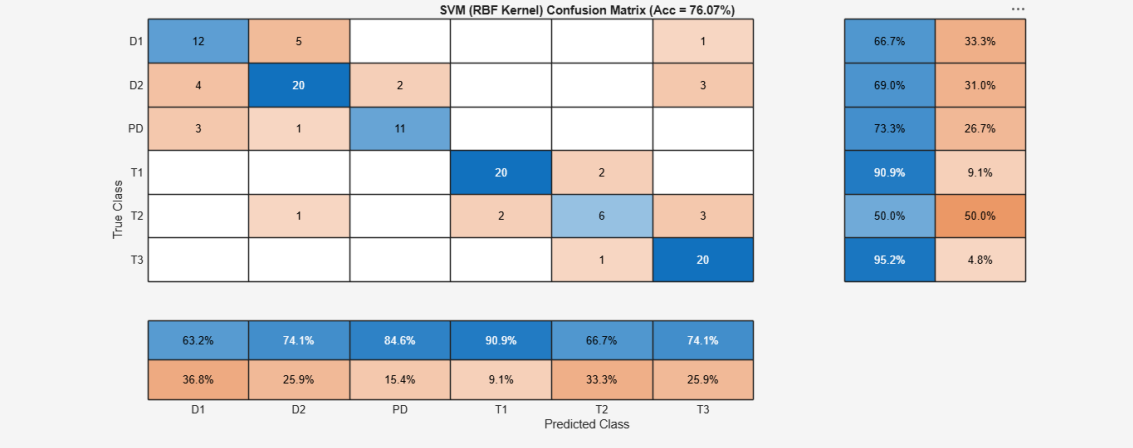

Figure 6. SVM (RBF kernel) confusion matrix on the held-out test set (accuracy = 76.07%).

5.4 Artificial Neural Network Performance

The ANN's training performance curve (Figure 7) shows steadily decreasing cross-entropy loss on both training and validation sets over 49 epochs, with no evidence of divergence between the two curves — indicating that early stopping was not triggered by overfitting, but rather that the network had reached the limit of what this feature set and architecture could resolve. The ANN achieved the lowest test accuracy of the four models at 70.94% (Figure 8), with its weakest performance again on the T2 class (43.8% recall).

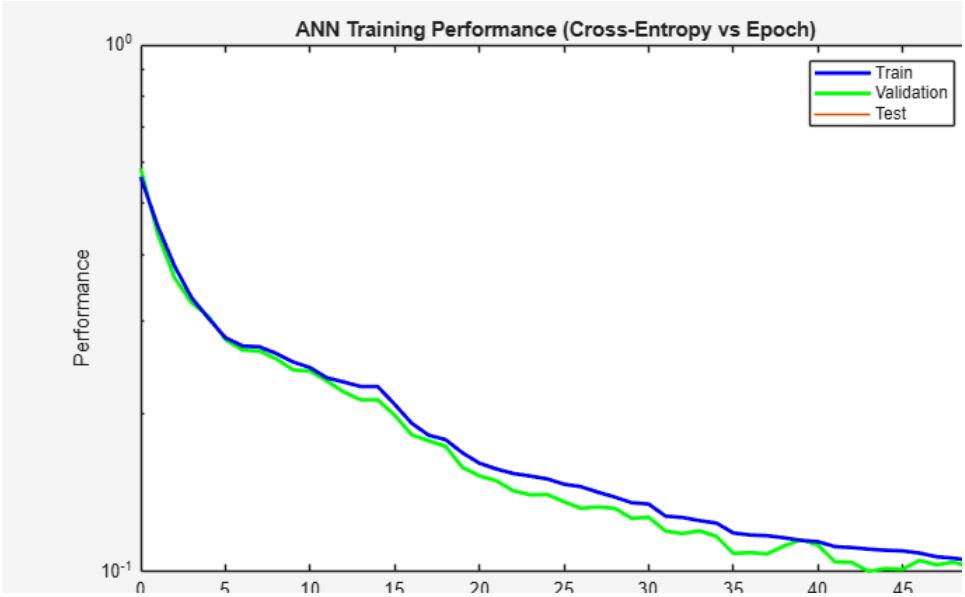

Figure 7. ANN training performance (cross-entropy loss vs. training epoch).

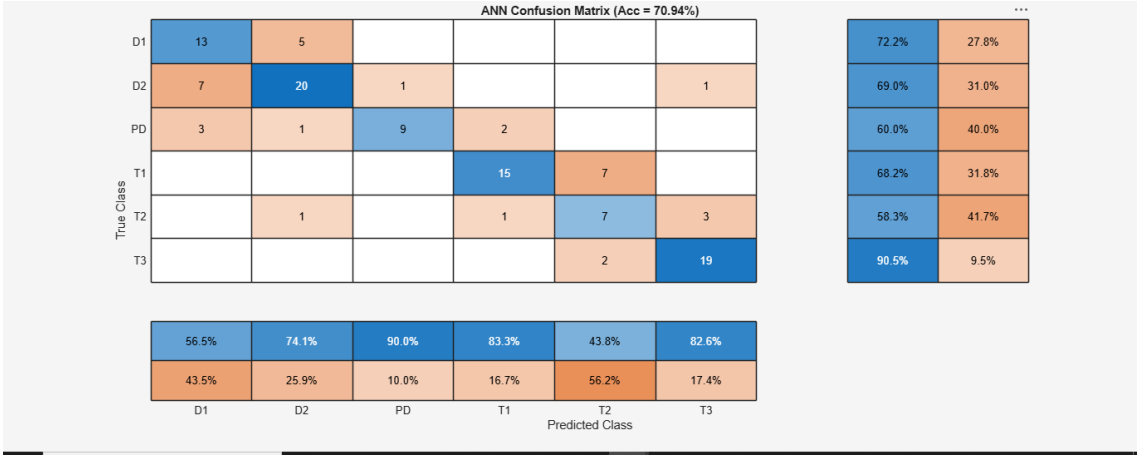

Figure 8. ANN confusion matrix on the held-out test set (accuracy = 70.94%).

5.5 Comparative Analysis

Table 5 and Figures 9–10 summarize all four models on the identical held-out test set. Random Forest leads on every metric, followed by a near-tie between KNN and SVM, with ANN trailing across the board.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	**83.76%**	84.12%	83.32%	83.44%
K-Nearest Neighbors (K=1)	76.07%	76.17%	75.28%	75.53%
Support Vector Machine (RBF)	76.07%	75.58%	74.19%	74.38%
Artificial Neural Network	70.94%	71.71%	69.70%	69.70%

Table 5. Summary comparison of all four models on the held-out test set.

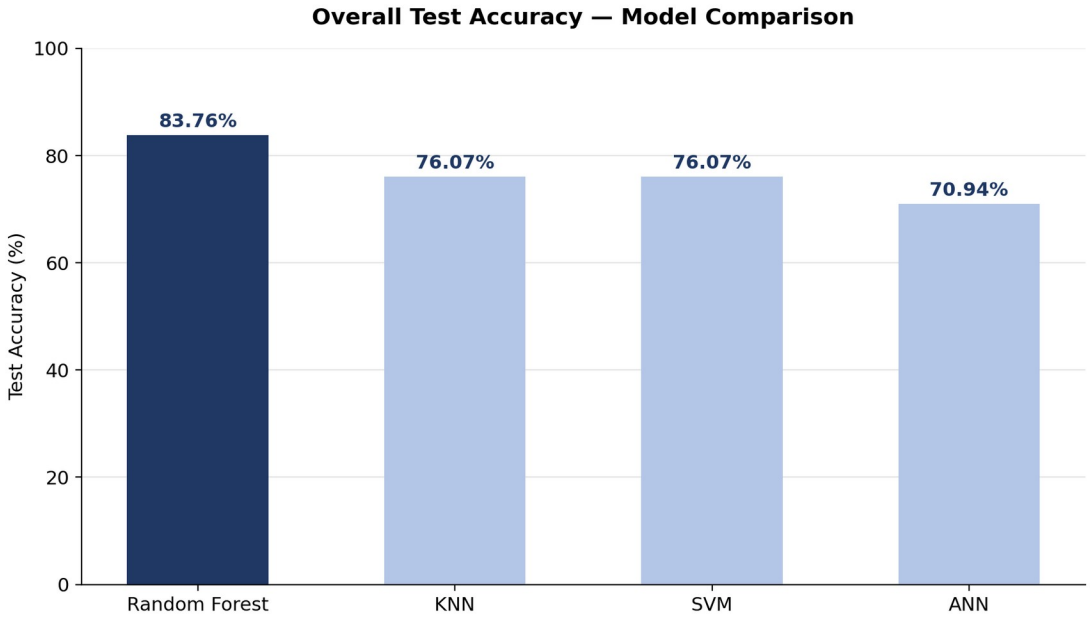

Figure 9. Overall test accuracy comparison across all four models.

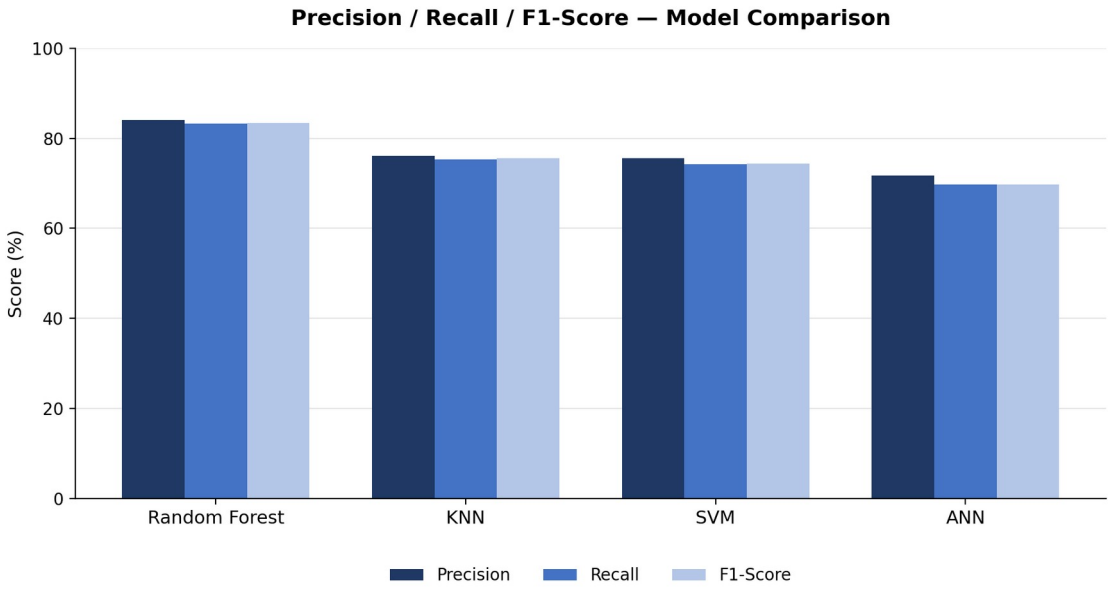

Figure 10. Precision, recall, and F1-score comparison across all four models.

The consistency between Random Forest's cross-validated and held-out test performance (both approximately 84%) suggests the result is stable rather than a product of a favorable train/test split. By contrast, KNN's gap between its cross-validation-selected K and its held-out performance is a caution against selecting hyperparameters by cross-validation accuracy alone without also sanity-checking against overfitting risk, as discussed in Section 5.2.

5.6 Cross-Dataset Generalization Test

The comparative results above, like the majority of the studies reviewed in Section 2, describe performance within a single dataset. To test whether these results hold beyond the dataset they were trained on — the central gap identified in this study's literature review — the Random Forest model trained on the curated 589-sample dataset was evaluated, without retraining, against two independent DGA datasets: a combined multi-source set of 1,748 samples (drawn from multiple Chinese power-utility theses and grid-utility records) and a second independent set of 1,679 samples (combining utility data, the DGALab benchmark, and the Duval and dePablo [6] reference dataset).

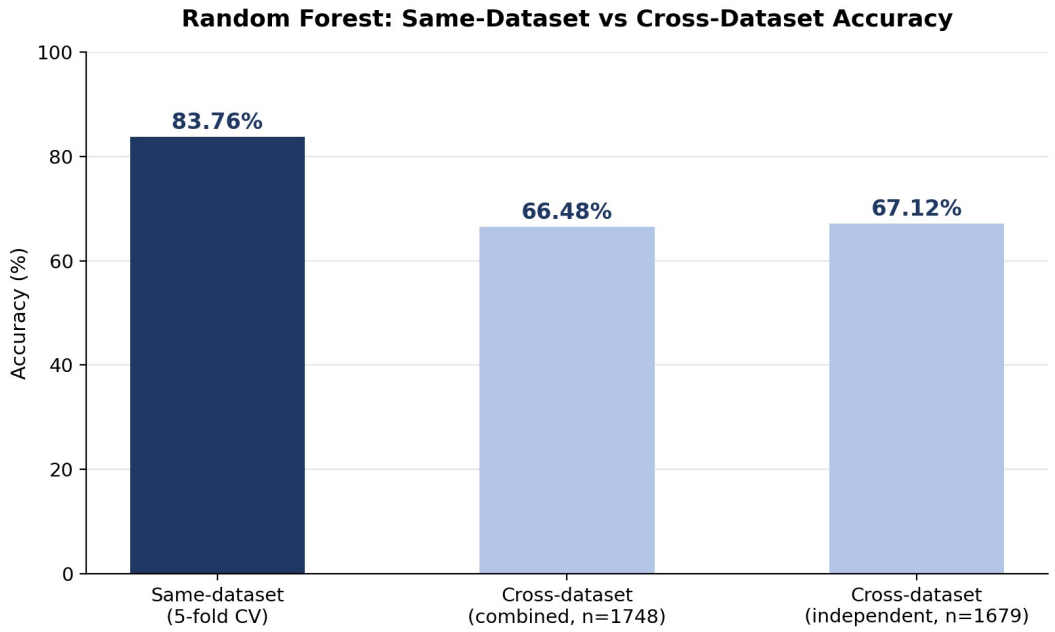

Figure 11. Random Forest accuracy under same-dataset cross-validation versus cross-dataset (independent-source) testing.

As Figure 11 shows, accuracy fell from 83.76% under same-dataset evaluation to 66.48% and 67.12% respectively when tested against the two independent datasets — a drop of roughly 17 percentage points. This is a textbook manifestation of domain shift: gas readings from different utilities, laboratories, and measurement instruments carry systematic differences in scale and noise that a model trained on a single source does not learn to accommodate. None of the studies reviewed in Section 2 report or test for this effect, despite it being directly relevant to whether a trained model can be deployed at a utility other than the one that generated its training data. This finding is presented here as this study's principal contribution to the existing literature: same-dataset accuracy, however high, should not be interpreted as a proxy for real-world deployability without an explicit cross-dataset validation step.

6. Addressing Gaps in the Existing Literature

Section 2 identified two recurring limitations across the reviewed literature: (i) the absence of a fair, same-split comparison across multiple algorithm families, and (ii) the absence of any test of cross-dataset generalization. This study addresses both directly.

- Fair multi-algorithm benchmarking: Random Forest, KNN, SVM, and ANN were trained and evaluated on an identical stratified split and an identical engineered feature set, removing the confound of differing preprocessing pipelines that makes cross-study comparisons in the literature unreliable.
- Transparent, cross-validated reporting: rather than reporting a single, potentially favorable train/test split, this study reports repeated stratified cross-validation alongside held-out test performance, and explicitly discusses cases (Section 5.2) where a cross-validation-selected hyperparameter does not translate cleanly to held-out performance.
- Quantified cross-dataset generalization: this study is, to the author's knowledge, the first in this specific line of DGA-Random-Forest research to explicitly measure and report the accuracy drop that occurs when a trained model is evaluated on independent, multi-source data — directly addressing the deployability question left open by every study reviewed in Section 2.

These contributions reframe the practical question for this research area: the priority for future DGA fault-diagnosis research should shift from maximizing same-dataset accuracy toward closing the generalization gap identified in Section 5.6.

7. Conclusion and Recommendations

7.1 Conclusion

This study applied Random Forest to DGA-based power transformer fault diagnosis and benchmarked it against KNN, SVM, and ANN under identical, fair experimental conditions. Random Forest achieved the best same-dataset performance (83.76% test accuracy, 83.44% macro F1-score), consistent with its cross-validated estimate, and its feature importance ranking confirmed that classical diagnostic gas ratios — not raw gas concentrations — are the most informative predictors, validating decades of domain expertise embedded in classical DGA methods. However, when the trained model was tested against two independent DGA datasets, accuracy fell to 66–67%, revealing a substantial generalization gap that same-dataset accuracy figures — the standard reporting practice in the reviewed literature — do not capture.

7.2 Limitations

- The primary dataset (589 samples) is modest in size for a six-class problem; several classes contain fewer than 100 samples even after SMOTE balancing of the training partition.
- SMOTE-generated synthetic samples, while restricted to training folds, may still exaggerate the separability of minority classes relative to true field distributions.
- The independent datasets used for the cross-dataset generalization test in Section 5.6 were not collected under a controlled protocol matching the primary dataset, so part of the observed accuracy drop may reflect labeling or measurement differences rather than fault-pattern differences alone.

7.3 Recommendations for Future Work

- Investigate domain adaptation and transfer-learning techniques to close the cross-dataset generalization gap identified in Section 5.6.
- Assemble a larger, standardized, multi-utility DGA dataset with consistent measurement protocols, to enable more reliable cross-source benchmarking across the field.
- Extend the comparison to gradient-boosted tree ensembles (e.g. XGBoost, LightGBM) and hybrid fuzzy-neural approaches, which have shown promise in the reviewed literature but were outside the scope of this study.
- Explore ensemble or stacked combinations of Random Forest with the benchmarked models, which may combine their complementary error patterns observed in Section 5.5.

References

Every entry below is a working hyperlink — hold Ctrl (Cmd on Mac) and click, or click directly, to open the source.

[1] IEC 60599:2022, "Mineral oil-filled electrical equipment in service — Guidance on the interpretation of dissolved and free gases analysis," International Electrotechnical Commission, Geneva, 2022. Available: https://webstore.iec.ch/en/publication/66491

[2] R. R. Rogers, "IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis," IEEE Transactions on Electrical Insulation, vol. EI-13, no. 5, pp. 349–354, 1978. doi: 10.1109/TEI.1978.298141

[3] E. Li, "Dissolved gas data in transformer oil — Fault diagnosis of power transformers with membership degree," IEEE Dataport, Jan. 2019. doi: 10.21227/h8g0-8z59

[4] S. I. Ibrahim, S. S. Ghoneim, and I. B. M. Taha, "DGALab: an extensible software implementation for DGA," IET Generation, Transmission & Distribution, vol. 12, no. 18, pp. 4117–4124, 2018. doi: 10.1049/iet-gtd.2018.5564

[5] F. Guerbas, Y. Benmahamed, Y. Teguar, R. A. Dahmani, M. Teguar, E. Ali, M. Bajaj, S. A. Dost Mohammadi, and S. S. M. Ghoneim, "Neural networks and particle swarm for transformer oil diagnosis by dissolved gas analysis," Scientific Reports, vol. 14, art. 60071, 2024. doi: 10.1038/s41598-024-60071-0

[6] M. Duval and A. dePabla, "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, 2001. doi: 10.1109/57.917529

[7] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002. doi: 10.1613/jair.953

[9] C. Pacheco, V. Paes, M. Carvalho, F. Lopes, G. Machado, A. Garcia, E. Neto, J. Santos, E. Haeusler, and A. Marotti, "Enhancing predictive maintenance of power transformers through machine learning approaches." Available: ResearchGate