ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct417

Machine Learning Diagnosis of Power Transformer Faults From Dissolved Gas Analysis: Application to a 120 kV Mining Substation

Moma Munga Ferdinand[1]; Lumanji Mbunga Luc[2]; Mavungu Mayela Alain[3]; Tshikudi Tshikudi Dennis[4]; Minda Mwilambwe Georges[5]

[1;2;3;4;5]Institut Supérieur de Techniques Appliquées (ISTA), Kinshasa, Democratic Republic of the Congo

Publication Date: 2026/10/23

Abstract: This article applies the concepts of a companion paper to the design of a monitoring and diagnosis system for the power transformers of a 120 kV substation supplying a copper–cobalt operation. Because continuous dissolved gas records were not available on site, the models were developed on a public simulated data set of 4,000 samples described by four gases (H_2, CO, C_2H_2, C_2H_4) and four states: normal, partial discharge, thermal fault and arc discharge. Logistic regression, k-nearest neighbours, a support vector machine, a random forest and a stacked ensemble were compared on 1,000 held-out samples. Accuracy recomputed from the confusion matrices was 83.8%, 87.9%, 83.7%, 91.4% and 91.7% respectively. The random forest (95% CI [89.5%, 93.0%]) was significantly better than k-nearest neighbours (z = 2.57, p = .010); the gain from stacking (0.3 point) was not significant (z = 0.24, p = .81). In every model the dominant errors were arc discharge classified as partial discharge and normal operation classified as a fault (14% to 23% false alarms), while 0.5% to 1.9% of faults were classified as normal. The insulation life model is formulated but was not validated numerically. Results on simulated data establish feasibility of the processing chain; they do not establish field performance.

Keywords: *Dissolved Gas Analysis, Power Transformer, Random Forest, Stacking, Fault Diagnosis, Predictive Maintenance, Mining Substation.*

How to Cite: Moma Munga Ferdinand; Lumanji Mbunga Luc; Mavungu Mayela Alain; Tshikudi Tshikudi Dennis; Minda Mwilambwe Georges (2026) Machine Learning Diagnosis of Power Transformer Faults From Dissolved Gas Analysis: Application to a 120 kV Mining Substation. *International Journal of Innovative Science and Research Technology*, 11(10), 1368-1373. https://doi.org/10.38124/ijisrt/26oct417

I. INTRODUCTION

The mining and metallurgical plants of the Lualaba province depend on a small number of high-voltage substations, and the loss of one power transformer stops production for the time needed to repair or replace it. At the site studied, transformer condition is assessed by dissolved gas analysis (DGA) of oil samples taken once every 12 months and interpreted with the Duval method. An annual sample cannot follow the growth of a fault between two campaigns, and interpretation by ratio methods varies with the analyst and leaves ambiguous cases (Nanfak et al., 2024).

Machine learning on DGA data is now well documented. Reviews report that support vector machines, tree ensembles and neural networks outperform ratio methods on labelled databases, and stress the role of class imbalance and parameter tuning (Acikgoz et al., 2025; Zhang et al., 2022). Reported accuracy varies widely with the data: 87.7% for bagged trees on 138 South African samples (Ngwenyama & Gitau, 2024), about 96% for a random forest with minority over-sampling (Prasojo et al., 2023), 93.4% for a boosted ensemble with feature engineering (Xu et al., 2024), and about 99% for k-nearest neighbours on a database whose labels were themselves produced by ratio methods (Wang et al., 2021). A recent evaluation is a useful warning: a random forest reaching 83.8% on 589 field-verified records fell to about 67% on two independent data sets (Adu-Gyamfi, 2026).

This article reports a feasibility study carried out for the KADI substation. It describes the installation and the monitoring architecture, the data and models, and a statistical re-analysis of the classification results; it then states what these results do and do not show. Equations are numbered within this article; derivations are in the companion article.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct417

II. MATERIALS AND METHOD

➢ *Installation*

The KADI high-voltage substation is supplied at 120 kV and houses four three-winding step-down transformers, 120 kV/15 kV/6.6 kV, 50 Hz, star–star–delta, rated 30 MVA (two units) and 40 MVA (two units). The proposed system (Figure 1) adds an on-line multi-gas sensor on the oil circuit. Its readings are acquired by a programmable logic controller (Siemens S7-1500, programmed in TIA Portal) which manages alarms and the operator panel, then stored in a SQL Server database. A Python service reads the database, applies the classifier and publishes the predicted state and class probabilities through a web interface (Gradio); gas trends are displayed in Power BI.

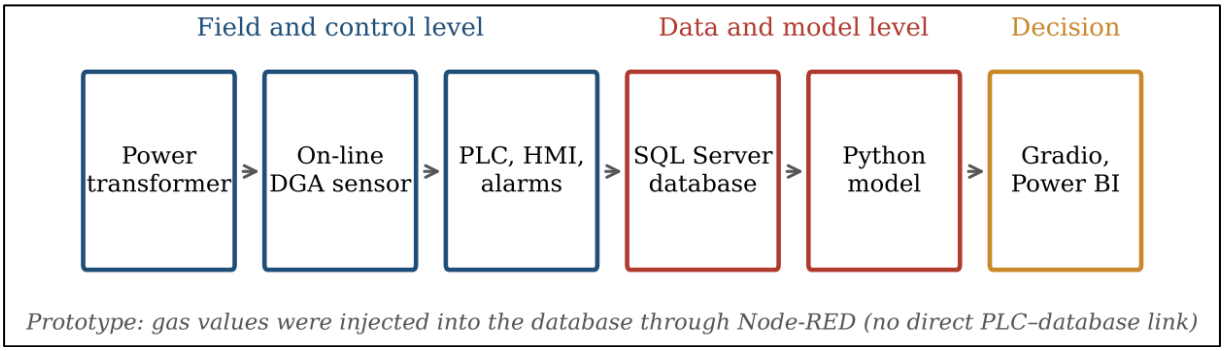


Fig 1 Architecture of the Monitoring and Diagnosis System

- *Note.* Sensor gases: H_2, CO, C_2H_2, C_2H_4. Transformers: 120 kV/15 kV/6.6 kV, 30 and 40 MVA.

➢ *Physical Basis of the Input Variables*

The four gases were selected by the simplified key-gas method. Hydrogen indicates low-energy partial discharge; ethylene indicates thermal faults, becoming dominant above about 500 °C; acetylene requires the 800–1,200 °C of an arc; carbon monoxide comes from cellulose and indicates ageing of the paper (International Electrotechnical Commission [IEC], 2022). The input vector and target are:

$$\mathbf{x} = [\mathrm{H_2},\ \mathrm{CO},\ \mathrm{C_2H_2},\ \mathrm{C_2H_4}]^{\top}, y \in \{\text{Normal, Partial discharge, Thermal, Arc}\} \quad (1)$$

For insulation life, the study adopts the loading-guide relation between hot-spot temperature θ_h and the ageing acceleration factor, with a normal life of 180,000 h at 110 °C:

$$F_{AA} = \exp\left(\frac{15{,}000}{383} - \frac{15{,}000}{\theta_h + 273}\right), \qquad L(\theta_h) = \frac{180{,}000}{F_{AA}}\ \text{h} \quad (2)$$

➢ *Data*

Continuous gas records from the substation did not exist, so a public data set from the Kaggle platform was used. It contains 4,000 simulated records representing readings taken every 12 hours on transformers of nuclear power plants, each with the four gas concentrations, a transformer identifier and a state label. There are no missing values. The four classes are balanced at about 1,000 records each. An interquartile-range screen flagged 352 CO values and one H_2 value as outliers and none for C_2H_4 or C_2H_2; they were kept, since high gas values are the signal of interest.

➢ *Models and Evaluation*

Four classifiers were trained with scikit-learn in Python (Google Colaboratory): multinomial logistic regression (LR), k-nearest neighbours (k-NN), a support vector machine with one-versus-rest decomposition (SVM) and a random forest (RF). The random forest was then tuned by grid search with cross-validation, and a stacked ensemble was built with SVM, k-NN and the tuned forest as base learners and logistic regression as meta-learner. Learning curves were drawn by cross-validation on increasing training sizes.

The performance figures below were recomputed from the published confusion matrices. Accuracy p_o on N test samples is given with its Wilson 95% interval, and two accuracies p_1, p_2 are compared by the two-proportion statistic

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{N} + \frac{p_2(1-p_2)}{N}}} \quad (3)$$

This test treats the two samples as independent. The models were scored on the same test set, for which a paired test (McNemar) would be more powerful, but it requires per-sample predictions that are not available. Three operational rates are also reported: the false-alarm rate (normal samples predicted as any fault), the missed-fault rate (faulty samples predicted as normal) and the share of arc samples predicted as partial discharge.

III. RESULTS

➢ *Overall Performance*

Figure 2 shows the five confusion matrices and Table 1 the derived measures. The random forest and the stacked ensemble form a leading group at 91.4% and 91.7%; k-NN follows at 87.9%; logistic regression and the SVM are level at 83.8% and 83.7%.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct417

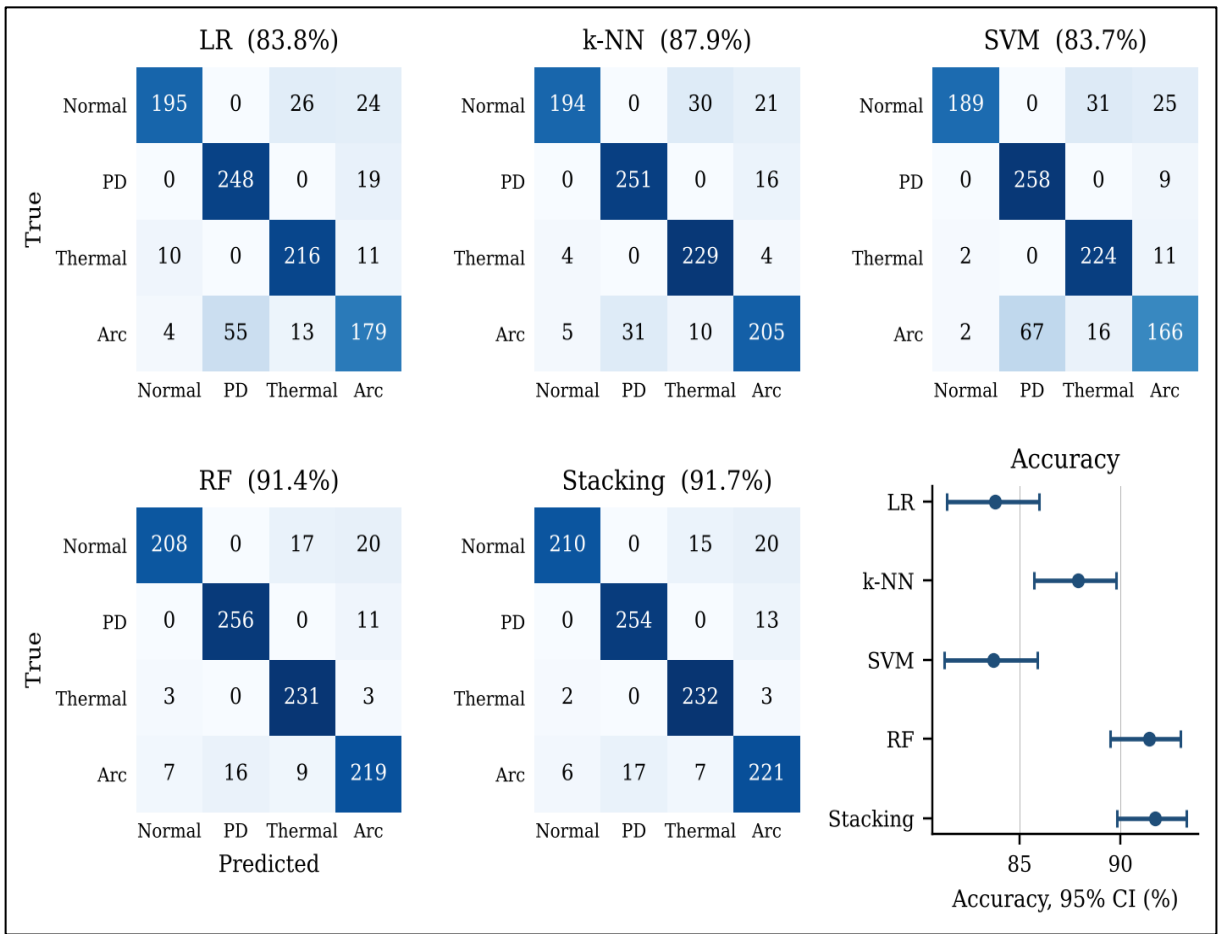


Fig 2 Confusion Matrices of the Five Models and Accuracy with 95% Confidence Intervals (N = 1,000)

- *Note.* Rows: true class; columns: predicted class. PD = partial discharge. Colour is the row-normalised rate.

Table 1 Overall Performance of the Five Models on the Test Set

Model	Accuracy (%)	95% CI (%)	Macro F1	Cohen's κ
Logistic regression	83.8	[81.4, 86.0]	0.837	0.784
k-nearest neighbours	87.9	[85.7, 89.8]	0.878	0.839
Support vector machine	83.7	[81.3, 85.9]	0.834	0.782
Random forest	91.4	[89.5, 93.0]	0.913	0.885
Stacked ensemble	91.7	[89.8, 93.3]	0.917	0.889

- *Note.* Computed from the counts in Figure 2. CI = Wilson confidence interval.

Applying Equation 3, the random forest is better than k-NN (z = 2.57, p = .010) and k-NN better than logistic regression (z = 2.63, p = .008). The difference between stacking and the random forest, 3 samples in 1,000, is far inside the sampling uncertainty (z = 0.24, p = .81). The reported learning curves are consistent with this ranking: the forest fits the training folds almost perfectly, k-NN reaches 0.87–0.89 in training and 0.84–0.85 in validation, and the SVM curves converge near 0.83–0.84.

➢ *Per-Class Behaviour*

Figure 3 and Table 2 show that performance is uneven across classes. Partial discharge and thermal faults are recognised with recall of 93% to 98% by every model. Arc discharge is the weakest class: its recall ranges from 66.1% (SVM) to 88.0% (stacking). Normal operation has high precision (93% to 98%) but recall of only 77% to 86%.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct417

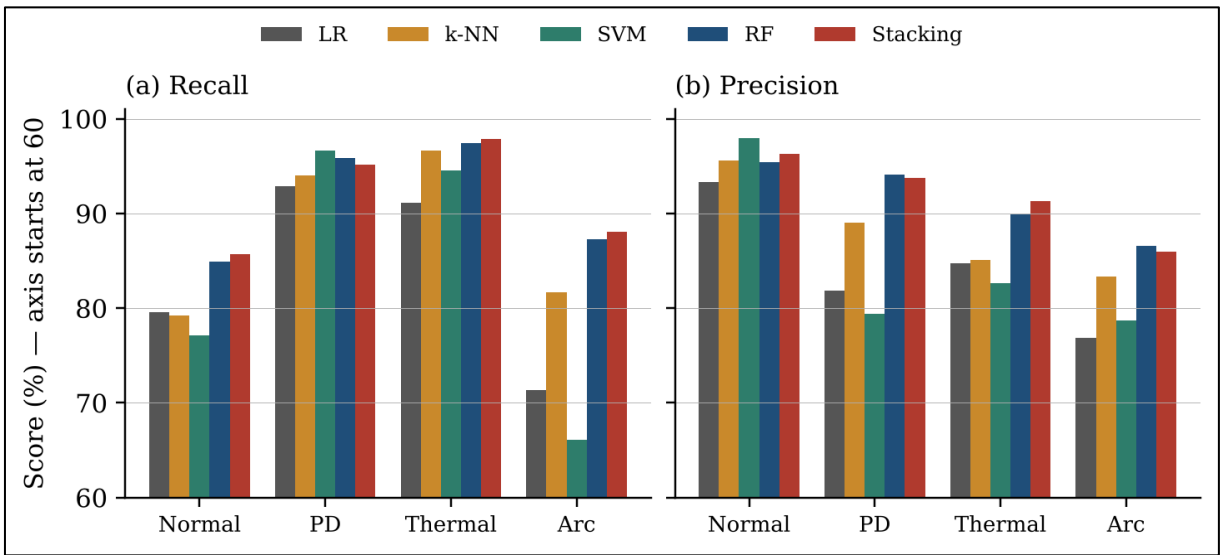


Fig 3 Per-Class Recall and Precision of the Five Models

- *Note.* The vertical axis starts at 60% to make differences visible.

Table 2 Per-Class Measures of the Two Best Models

Class	Support	P (RF)	R (RF)	F1 (RF)	P (stack)	R (stack)	F1 (stack)
Normal	245	0.954	0.849	0.898	0.963	0.857	0.907
Partial discharge	267	0.941	0.959	0.950	0.937	0.951	0.944
Thermal	237	0.899	0.975	0.935	0.913	0.979	0.945
Arc	251	0.866	0.873	0.869	0.860	0.880	0.870

- *Note.* P = precision; R = recall; RF = random forest; stack = stacked ensemble.

➢ *Structure of the Errors*

The matrices share a pattern that is more informative than the totals. In all five models, no normal sample is predicted as partial discharge, no partial-discharge sample is predicted as normal or thermal, and no thermal sample is predicted as partial discharge. Partial-discharge errors go only to the arc class. Errors therefore concentrate in two places, quantified in Figure 4. The first is arc predicted as partial discharge: 26.7% of arc samples for the SVM and 21.9% for logistic regression, against 6.4% for the random forest. The second is false alarms: 14.3% to 22.9% of normal samples are predicted as thermal or arc faults. Missed faults are much rarer, from 0.5% (SVM, 4 of 755) to 1.9% (logistic regression, 14 of 755); the random forest misses 10 and stacking 8.

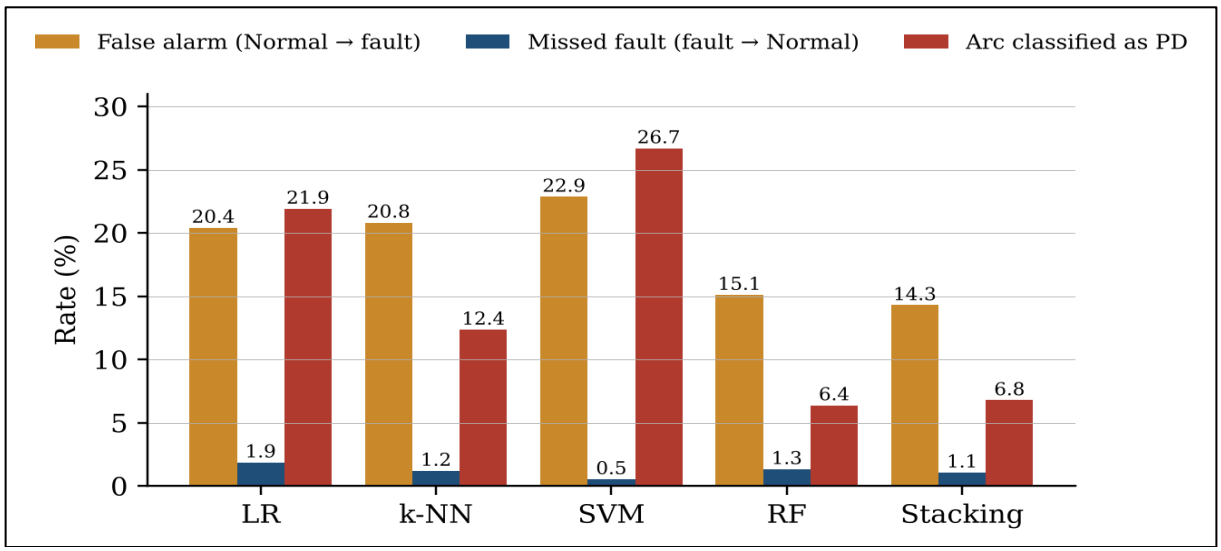


Fig 4 Operational Error Rates of the Five Models

- *Note.* False alarm: share of 245 normal samples predicted as a fault. Missed fault: share of 755 faulty samples predicted as normal. Arc as PD: share of 251 arc samples predicted as partial discharge.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct417

➢ *Insulation Life*

Table 3 evaluates Equation 2 for four hot-spot temperatures. It is a scenario calculation: no hot-spot or load record of the KADI units was available, and the source study reports no numerical estimate of remaining life.

Table 3 Ageing Acceleration Factor and Expected Insulation Life at Constant Hot-Spot Temperature

Hot-spot temperature (°C)	F_{AA}	Expected life (h)	Expected life (years)
98	0.282	638,890	72.9
110	1.000	180,000	20.5
120	2.709	66,447	7.6
130	6.984	25,773	2.9

- *Note.* Equation 2 with a normal life of 180,000 h.

IV. DISCUSSION

➢ *Physical Reading of the Errors*

Both systematic confusions have a physical explanation. Partial discharge and arcing are electrical faults and both release hydrogen; what separates them is acetylene, which forms only at arc temperatures. A low-energy arc, or an arc sampled before acetylene has accumulated, lies close to partial discharge in a four-gas space. Linear boundaries cannot follow this separation, which is why logistic regression and the SVM, at the same accuracy, fail on the same cells, while the tree ensemble, which thresholds acetylene directly, reduces this confusion to about 6% of arc samples. The false alarms on normal operation involve the thermal and arc classes, never partial discharge: a healthy transformer always contains some ethylene and CO from normal ageing, so the normal class overlaps the low end of the thermal class. Separating them calls for rates of change as well as levels, and for the hydrocarbons absent from this data set (methane, ethane) on which ratio methods rely.

➢ *Model Choice*

The random forest is the reasonable choice for deployment: it is statistically indistinguishable from the stacked ensemble, simpler to maintain, and provides class probabilities. Stacking gains little because its base learners make largely the same errors, which is the condition under which ensembles of different models cannot help (Mienye & Sun, 2022). The SVM result deserves attention for another reason. It has the lowest accuracy but also the fewest missed faults (4, against 10 for the forest). If the cost of an undetected fault is much higher than that of an unnecessary oil sample, the operating point should be chosen on the predicted probabilities with that cost ratio, not at the default majority vote.

➢ *Comparison with the Literature*

The 91.4% obtained lies within the range reported for tree ensembles on DGA data (Ngwenyama & Gitau, 2024; Prasojo et al., 2023; Xu et al., 2024), and the ranking of model families agrees with the larger comparison of Rao et al. (2021). The comparison has limited meaning, however, because the present data are simulated and restricted to four gases, whereas the cited studies use laboratory or field records.

V. LIMITATIONS

Six limitations bound the conclusions. First, the data are simulated and unrelated to the KADI transformers; the drop of about 17 points observed when a forest was moved to independent field data (Adu-Gyamfi, 2026) indicates the size of the risk. Second, the source documents are inconsistent on the test set: the text states an 80/20 stratified split (800 test samples), whereas every confusion matrix sums to 1,000 samples with unequal supports (245, 267, 237, 251); the matrices are used here. Third, hyper-parameters are not reported, and the forest is described with four trees in its formulation, a size at which the variance reduction of bagging is small. Fourth, the logistic-regression learning curve is reported to converge near 0.55, which does not match its 83.8% test accuracy. Fifth, the remaining-life function is specified but neither linked to the CO measurement nor validated. Sixth, the prototype did not close the loop: direct transfer from the controller to the database failed with the available supervision licence, pressure sensors stood in for gas sensors, and gas values were injected into the database by Node-RED.

VI. IMPLEMENTATION AND FURTHER WORK

The estimated cost of the system is USD 65,940 (equipment USD 19,800, of which the multi-gas sensor is USD 12,800; development, deployment and training USD 38,500; first-year maintenance USD 7,640). Before commissioning, the priorities are: (a) to build a local labelled base from the annual laboratory DGA results and from on-line readings; (b) to extend the inputs to the full gas set and to gas ratios, which also improve robustness to measurement noise (Taha et al., 2021); (c) to validate on independent data and to report paired tests; (d) to link the controller and the database through OPC UA; and (e) to couple the life model to measured top-oil temperature and load, complemented by chemical markers of paper degradation (Behjat et al., 2021), keeping in mind that ageing also proceeds below the reference temperature (Leu et al., 2023).

VII. CONCLUSION

Five classifiers were compared for four-gas diagnosis of transformer state. On 1,000 simulated test samples, the random forest reached 91.4% accuracy (95% CI [89.5%, 93.0%]) and the stacked ensemble 91.7%, a difference that is not statistically significant; both were significantly better

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct417

than k-nearest neighbours, logistic regression and the support vector machine. Errors were concentrated in two physically explicable confusions, arc with partial discharge and normal with thermal or arc, giving false-alarm rates of 14% to 23% against missed-fault rates below 2%. The monitoring chain from controller to web interface was demonstrated with injected data. The study establishes feasibility; field performance on the KADI transformers, and the remaining-life estimate, remain to be validated with local measurements.

REFERENCES

[1]. Acikgoz, N., Sazli, M. H., & Kosalay, I. (2025). A comprehensive review of transformer fault diagnosis studies based on dissolved gas analysis: Classical methods, historical development of the devices used, artificial intelligence based methods, accuracy of classifications of predictions. *Elektronika ir Elektrotechnika, 31*(3), 4–21. https://doi.org/10.5755/j02.eie.39824

[2]. Adu-Gyamfi, K. (2026). *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*. engrXiv. https://doi.org/10.31224/8150

[3]. Behjat, V., Emadifar, R., Pourhossein, M., Rao, U. M., Fofana, I., & Najjar, R. (2021). Improved monitoring and diagnosis of transformer solid insulation using pertinent chemical indicators. *Energies, 14*(13), Article 3977. https://doi.org/10.3390/en14133977

[4]. International Electrotechnical Commission. (2022). *Mineral oil-filled electrical equipment in service – Guidance on the interpretation of dissolved and free gases analysis* (IEC 60599:2022, 4th ed.).

[5]. Leu, B., Seriţan, G., & Enache, B. (2023). A new hybrid model to evaluate the loss of life of power transformers. *Revue Roumaine des Sciences Techniques, Série Électrotechnique et Énergétique, 68*(3), 277–282. https://doi.org/10.59277/RRST-EE.2023.3.5

[6]. Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access, 10*, 99129–99149. https://doi.org/10.1109/ACCESS.2022.3207287

[7]. Nanfak, A., Eke, S., Fofana, I., Meghnefi, F., Ngaleu, G. M., & Kom, C. H. (2024). Traditional fault diagnosis methods for mineral oil-immersed power transformer based on dissolved gas analysis: Past, present and future. *IET Nanodielectrics*. Advance online publication. https://doi.org/10.1049/nde2.12082

[8]. Ngwenyama, M. K., & Gitau, M. N. (2024). Discernment of transformer oil stray gassing anomalies using machine learning classification techniques. *Scientific Reports, 14*, Article 376. https://doi.org/10.1038/s41598-023-50833-7

[9]. Prasojo, R. A., Putra, M. A. A., Ekojono, Apriyani, M. E., Rahmanto, A. N., Ghoneim, S. S. M., Mahmoud, K., Lehtonen, M., & Darwish, M. M. F. (2023). Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique. *Electric Power Systems Research, 220*, Article 109361. https://doi.org/10.1016/j.epsr.2023.109361

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

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

[12]. Wang, L., Littler, T., & Liu, X. (2021). Artificial intelligence model for transformer fault diagnosis using a constructed database. In *The 16th IET International Conference on AC and DC Power Transmission (ACDC 2020)* (pp. 1474–1479). Institution of Engineering and Technology. https://doi.org/10.1049/icp.2020.0194

[13]. Xu, G., Zhang, M., Chen, W., & Wang, Z. (2024). Transformer fault diagnosis utilizing feature extraction and ensemble learning model. *Information, 15*(9), Article 561. https://doi.org/10.3390/info15090561

[14]. Zhang, Y., Tang, Y., Liu, Y., & Liang, Z. (2022). Fault diagnosis of transformer using artificial intelligence: A review. *Frontiers in Energy Research, 10*, Article 1006474. https://doi.org/10.3389/fenrg.2022.1006474