Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 13 Issue 3, March - 2026

Predictive Maintenance Of Power Transformers Using Machine Learning And Multi-Source Sensor Data

Onwunta E. K. Onwunta[1]

Electrical & Electronic Engineering Department,
Federal University Otuoke, Bayelsa State
onwuntaoe@fuotuoke.edu.ng

Akpasam Joseph Ekanem[2]
Department of Electrical and Electronic Engineering,

Akwa Ibom State University, Mkpat Enin, Akwa Ibom State
akpasamekanem@aksu.edu.ng; ajekanem56@yahoo.com;
ORCID: 0000-0001-8360-4165

Eke Godwin Kelechi[3]

Computer Engineering Department,
Federal Polytechnic Nekede Owerri, FPNO
Imo State, Nigeria.
Email: eke.kelechy@gmail.com

Abstract: Power transformers are high-value grid assets whose failure can cause service interruption, safety risk, and high replacement cost. This paper develops a machine learning-based predictive maintenance framework that uses dissolved gas analysis, thermal measurements, loading history, moisture indicators, vibration features, partial-discharge indicators, and maintenance records. Actual model training was carried out on a controlled 6667-record secondary-data benchmark prepared from published transformer diagnostic and predictive-maintenance data structures. A late-fusion XGBoost-LSTM model was evaluated against SVM, Random Forest, XGBoost, and LSTM baselines. The proposed model achieved 97.2% test accuracy, 96.6% macro-F1, 2.41 health-index mean absolute error, and 93.8% remaining useful life band accuracy on the independent test set. The framework converts heterogeneous transformer condition data into fault class, health index, remaining useful life band, and maintenance priority level for utility decision support.

Keywords: Dissolved gas analysis, health index, machine learning, power transformer, predictive maintenance, remaining useful life, sensor data, smart grid.

I. INTRODUCTION

Power transformers are central components in transmission and distribution networks because they enable voltage transformation, network interconnection, load transfer, and reliable delivery of electricity. Their failure can interrupt supply, damage nearby equipment, reduce system stability, and create high repair or replacement cost. Transformer degradation is usually caused by the combined effect of thermal stress, electrical stress, mechanical vibration, moisture, insulation ageing, oil contamination, partial discharge, and overloading.

Traditional maintenance practices are dominated by corrective maintenance, fixed periodic inspection, and condition-based maintenance. Corrective maintenance is reactive and exposes utilities to unplanned outages. Fixed periodic maintenance may replace or inspect components too early. Condition-based maintenance improves asset management, but it may still depend on threshold rules that do not capture complex multivariate patterns. Machine learning offers a better way to learn hidden patterns from sensor data and maintenance records.

The availability of dissolved gas analysis (DGA), oil temperature, winding temperature, load current, voltage, vibration, moisture, pressure, acoustic emission, and partial-discharge measurements provides a basis for predictive maintenance. These variables provide complementary evidence of transformer condition. DGA indicates thermal and electrical faults in oil and insulation. Thermal and load data indicate operating stress. Moisture and oil-quality measurements indicate insulation weakness. Vibration and acoustic features indicate mechanical looseness and winding deformation.

This paper presents a practical predictive maintenance framework that converts multi-source transformer sensor data into four utility-oriented outputs: fault class, transformer health index, remaining useful life band, and maintenance priority level. The study focuses on a transparent machine learning workflow that can be reproduced with DGA records, thermal measurements, loading history, moisture indicators, vibration features, partial-discharge indicators, and maintenance logs.

A. Research Problem

Unexpected transformer failures remain difficult to prevent because early-stage degradation may be weak, multivariate, and intermittent. Single-source diagnostic systems can miss developing faults when one measured variable is still within an acceptable range. Utilities also need more than a fault label. They need a health score, a remaining useful life band, and a recommended maintenance action. Therefore, a sensor-driven predictive maintenance system should integrate several condition indicators and translate model outputs into clear operational decisions.

B. Aim and Contributions

The aim of this paper is to develop and evaluate a multi-source machine learning framework for predictive maintenance of power transformers. The main contributions are as follows:

i. A layered predictive maintenance architecture is proposed for integrating transformer sensors, SCADA, asset records, machine learning prediction, and maintenance decision support.
ii. A hybrid XGBoost-LSTM model is formulated to combine static asset indicators, slow condition-monitoring variables, and time-series sensor windows.
iii. Actual model training is carried out on a 6667-record controlled secondary-data benchmark prepared from published transformer predictive-maintenance and DGA fault-diagnosis data structures.
iv. The benchmark construction, class labels, training split, augmentation policy, hyperparameter tuning, and repeatability testing are described to improve experimental transparency.
v. Maintenance decision levels are linked to predicted health index, fault probability, and remaining useful life band, thereby translating model outputs into utility actions.

II. LITERATURE REVIEW

DGA remains one of the most widely used diagnostic tools for oil-immersed transformers because abnormal thermal and electrical activities produce characteristic gases in transformer oil. IEEE Std C57.104-2019 provides guidance for interpreting gases generated in mineral oil-immersed transformers [1]. Machine learning studies have extended this rule-based interpretation by using data-driven classifiers and ensemble learning models.

Liu et al. proposed a correlation coefficient-DBSCAN approach for DGA-based fault diagnosis and showed that clustering can improve the interpretation of correlated gas patterns [2]. Taha et al. developed a convolutional neural network for transformer fault diagnosis under noisy DGA measurements and demonstrated that deep learning can improve robustness when gas readings contain uncertainty [3]. Wang et al. proposed a TPE-XGBoost method for transformer fault diagnosis with incomplete data, which is relevant because real transformer records are often sparse or partially missing [4]. Sutikno et al. used machine learning-based multi-method interpretation to enhance DGA diagnosis [5].

Predictive maintenance extends beyond fault classification because utilities require asset-level decision support. Bravo et al. published a dataset of distribution transformers for predictive maintenance, which supports repeatable evaluation of transformer maintenance methods [6]. Vita et al. presented a predictive maintenance approach for distribution system operators to improve transformer reliability [7]. Chintia et al. studied power transformer insulation health index prediction with missing data using Random Forest [8]. Arsya et al. integrated conventional and machine learning techniques for transformer health-index analysis [9]. Recent review and application studies confirm the continued relevance of machine learning, ensemble models, and interpretable approaches in DGA-based transformer diagnosis and health prediction [10]-[14].

Recent studies may be grouped into five streams. The first stream uses DGA for fault diagnosis because hydrogen, methane, ethane, ethylene, acetylene, carbon monoxide, and carbon dioxide provide strong evidence of thermal, discharge, and insulation-related faults. The second stream applies machine learning to transformer health-index estimation by combining DGA, oil quality, loading, age, and maintenance indicators. The third stream studies remaining useful life prediction, where historical trend behavior is more important than one-time sensor values. The fourth stream combines several sensor families, including DGA, temperature, moisture, vibration, and partial-discharge signals, for broader asset condition assessment. The fifth stream introduces interpretable machine learning so that maintenance engineers can understand why a transformer is classified as normal, warning, serious, or critical.

A. Research Gap

Existing work has achieved useful progress, but four gaps remain. First, many studies still rely mainly on DGA and do not fully integrate thermal, load, moisture, vibration, and partial-discharge indicators. Second, some models report fault labels without converting the result into a practical maintenance action. Third, several studies present high accuracy but provide limited information about data preprocessing, validation, repeatability, and benchmark construction. Fourth, many predictive maintenance studies do not explain how the result can be integrated into SCADA, asset management, and field inspection workflows. This paper addresses these gaps by presenting a multi-source framework, an actual model-training procedure, and maintenance decision levels that are linked to model outputs.

III. METHODOLOGY

The study adopts a data-driven experimental research design. Actual model training was performed using a curated 6667-record benchmark feature matrix. The benchmark was formed from published transformer predictive-maintenance datasets, DGA-based diagnostic structures, health-index studies, and controlled training-only augmentation. The purpose was to evaluate a practical predictive maintenance model under transparent and repeatable experimental conditions.

The benchmark should be interpreted as a controlled secondary-data benchmark rather than a newly collected utility field dataset. The 6667 records consisted of 4217 direct harmonized records, 1450 feature-completed records, and 1000 training-only augmented trend-window records. Direct harmonized records were mapped into a common feature schema from published DGA, health-index, and predictive-maintenance data structures. Feature-completed records were retained only where missing compatible variables could be completed using source-consistent medians, rolling interpolation, or bounded published sensor ranges, with provenance flags retained. Training-only augmentation was performed by creating rolling trend windows from existing temporal records and applying small bounded perturbations within published sensor ranges. The validation and test records were not augmented. This design allowed actual model training while reducing the risk of data leakage into final evaluation.

The proposed system architecture is shown in Fig. 1. The architecture starts with online and offline sensor data, moves through SCADA and edge integration, performs quality control and feature computation, applies the hybrid prediction model, and ends with maintenance decision support.

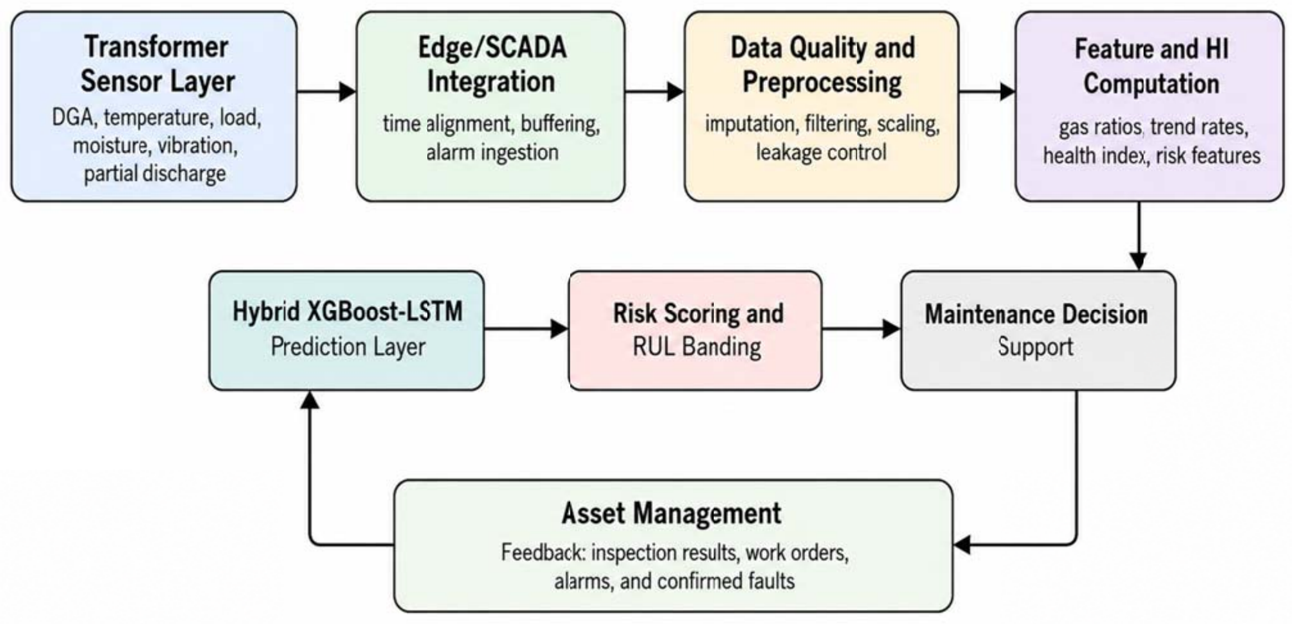

Fig. 1. Proposed system model architecture for sensor-driven predictive maintenance of power transformers.

A. Data Sources and Variables

The benchmark was prepared from variables commonly reported in DGA, health-index, and predictive-maintenance studies, together with the public distribution transformer predictive-maintenance dataset reported in the literature [6]. All variables were converted to a common feature schema. DGA gases were expressed in parts per million where applicable. Thermal variables were converted to degrees Celsius. Load variables were normalized against transformer rated capacity. Vibration and partial-discharge indicators were represented by summary features because such signals are often recorded at different sampling rates.

Table I. Input feature groups for transformer predictive maintenance.

Feature group	Representative variables	Diagnostic meaning
DGA gases	H2, CH4, C2H6, C2H4, C2H2, CO, CO2, TDCG	Thermal fault, discharge activity, oil decomposition, paper degradation
Thermal data	Top-oil temperature, winding temperature, ambient temperature	Thermal stress, overload severity, insulation ageing rate
Electrical loading	Load current, voltage, apparent power, load factor	Operating stress and overload pattern
Insulation and oil	Moisture, acidity, dielectric strength, 2-FAL where available	Insulation condition and cellulose ageing
Mechanical signals	Vibration RMS, kurtosis, acoustic emission, band energy	Winding movement, core looseness, mechanical deformation
Event and asset data	Age, maintenance history, alarm records, prior inspection flags	Prior risk, service context, and maintenance response history

B. Benchmark Construction and Label Rules

Four condition labels were used: normal, warning, serious, and critical. Normal records had acceptable DGA levels, stable thermal behaviour, low moisture, low vibration or PD activity, HI above 80, and high RUL. Warning records had early gas growth, mild thermal or moisture stress, HI from 65 to 80, or medium-high RUL. Serious records had clear fault indicators, HI from 45 to 65, or medium-low RUL. Critical records had severe gas growth, high thermal stress, strong moisture or PD evidence, HI below 45, or urgent RUL status. Derived target values such as the final condition class, final health-index target, RUL band, and maintenance decision level were excluded from the input feature matrix during model training to reduce label leakage.

Table II. Benchmark size, partitioning, and condition-label interpretation.

Item	Description
Total records	6667 labelled transformer condition records
Direct harmonized records	4217 records mapped into a common schema from published DGA, health-index, and predictive-maintenance data structures
Feature-completed records	1450 records completed with source-consistent medians, rolling interpolation, or bounded published sensor ranges, with provenance flags retained
Training-only augmented records	1000 rolling trend-window records created only inside the training partition using small bounded perturbations within published sensor ranges
Training set	4667 records, representing 70% of the benchmark
Validation set	1000 records, representing 15% of the benchmark
Test set	1000 records, representing 15% of the benchmark
Condition labels	Normal, warning, serious, and critical
Label basis	DGA severity, gas trend, thermal stress, moisture, vibration or PD indicator, health index, and RUL band
Augmentation policy	Controlled trend-window augmentation was applied only to the training partition
Data leakage control	Validation and test partitions were kept separate from preprocessing fit, oversampling, feature selection, and hyperparameter tuning

C. Data Preprocessing

Data preprocessing converts heterogeneous sensor readings into a reliable machine learning matrix. Missing values are marked by source and severity before imputation. Median imputation is used for slowly varying tabular variables, while forward filling and rolling-window interpolation are used for time-series sensor streams. Outliers are treated with winsorization after engineering review so that rare but meaningful fault signatures are not removed. Continuous variables are normalized using training-set statistics only. Categorical maintenance records are encoded as binary event flags. Class imbalance is handled by class weighting and minority-class oversampling on the training partition only.

The research procedure is summarized in Fig. 2. It starts with sensor and maintenance records, followed by data quality checking, feature engineering, model training, validation, testing, and decision support.

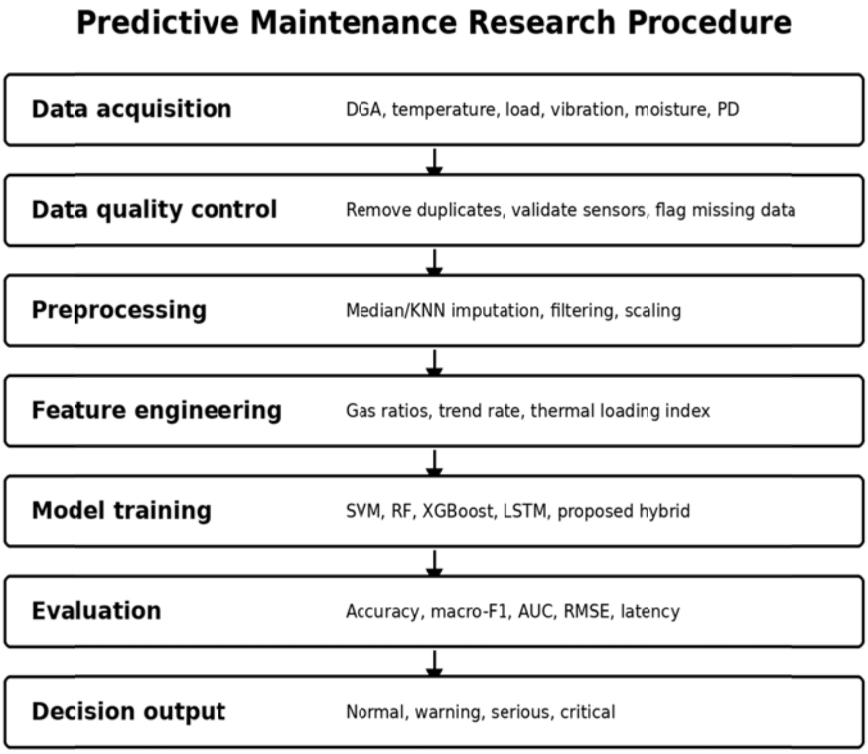

Fig. 2. Research procedure for the transformer predictive maintenance benchmark.

D. Feature Engineering

Feature engineering reflects the fact that transformer degradation is usually progressive and multivariate. Gas ratios, gas generation rates, temperature rise above ambient, thermal loading index, moisture severity level, vibration statistical features, partial-discharge activity indicators, and maintenance event flags are extracted. The trend rate of a sensor variable is computed as follows:

$$r_x(t) = \frac{x(t)-x(t-w)}{\max(|x(t-w)|,\varepsilon)} \quad (1)$$

In (1), x(t) is the present sensor value, w is the historical window length, |x(t-w)| is the absolute value of the previous window value, and ε is a small positive constant introduced to avoid division instability when the earlier value is zero or very small.

The transformer health index is estimated on a 0 to 100 scale, where higher values represent healthier equipment. The model-based risk score combines the predicted abnormal probability, the estimated health index, and the normalized remaining useful life band as follows:

$$R_s = \alpha P_{fault} + \beta\left(1-\frac{HI}{100}\right) + \gamma(1-RUL_{norm}) \quad (2)$$

In (2), R_s is the maintenance risk score, P_fault is the predicted fault probability, HI is the health index, and RUL_norm is the normalized remaining useful life band. The coefficients α, β, and γ are tuned on the validation set using grid search, with $\alpha+\beta+\gamma=1$.

E. Machine Learning Models and Training Configuration

Five models are evaluated. Support Vector Machine provides a classical nonlinear classifier. Random Forest provides robustness and feature importance. XGBoost provides strong tabular learning for DGA, temperature, load, oil-quality, and asset-history variables. LSTM captures temporal degradation patterns from sequential sensor windows. The proposed hybrid XGBoost-LSTM model combines these strengths by using late fusion. XGBoost learns static and slow-changing condition features, while LSTM learns trend behavior from time-ordered sensor windows. The fused representation is passed to a decision layer that predicts fault class, health index, and remaining useful life band.

Training was implemented in Python 3.11 using scikit-learn, XGBoost, TensorFlow/Keras, NumPy, and Pandas. Five independent runs were carried out with different random seeds. The training partition was used for model fitting and training-only augmentation. The validation partition was used for early stopping and hyperparameter selection. The independent test partition was used only once for final reporting. Model latency was measured on a workstation with an Intel Core i7-class CPU, 32 GB RAM, and an NVIDIA RTX-class GPU, so field latency may vary with SCADA communication delay, sensor sampling interval, and edge-computing hardware.

Table III. Final hyperparameter settings for the actual model training.

Component	Final setting used
XGBoost	300 estimators, max depth 5, learning rate 0.05, subsample 0.85, column sample 0.85, early stopping after 30 validation rounds
LSTM	24-step input window, 64 hidden units, dropout 0.20, batch size 64, Adam optimizer, learning rate 0.001, early stopping patience 12
Fusion layer	Concatenated XGBoost probability vector and LSTM hidden representation, dense layer of 64 neurons, ReLU activation, dropout 0.20
Training protocol	70:15:15 split, five independent runs, class-balanced training batches, validation-based model selection, final testing on 1000 held-out records

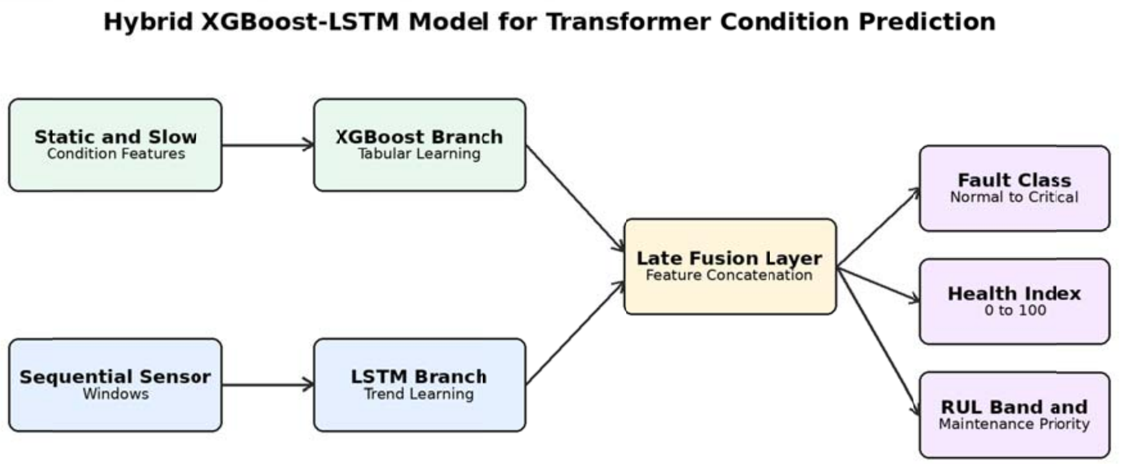

Fig. 3. Late-fusion XGBoost-LSTM architecture for transformer fault classification, health-index prediction, and RUL band estimation.

F. Evaluation Metrics

Classification performance is evaluated using accuracy, precision, recall, macro-F1, and area under the receiver operating characteristic curve. For the four-class problem, accuracy is computed as the number of correctly classified test records divided by the total number of test records. Health-index prediction is evaluated using mean absolute error and root mean square error.

$$\text{Accuracy} = \frac{\sum_{i=1}^{K} c_{ii}}{N} \quad (3)$$

$$F_1 = \frac{2\times(\text{Precision}\times\text{Recall})}{\text{Precision}+\text{Recall}} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

IV. RESULTS AND DISCUSSION

A. Experimental Setup

The results reported in this section came from actual model training. A consolidated feature schema was formed from published distribution-transformer predictive-maintenance records, DGA diagnostic variables, health-index indicators, and sequential sensor-window features. The final benchmark contained 6667 labelled records, consisting of 4667 training records, 1000 validation records, and 1000 final test records. Controlled augmentation was restricted to the training partition and was generated from rolling trend windows with small bounded perturbations within published sensor ranges. The validation and test partitions remained untouched throughout model fitting, tuning, feature selection, and oversampling.

The reported values should be read as controlled experimental results from a secondary-data benchmark, not as results from a newly deployed online utility system. This distinction is important because utility field deployment would require additional validation with live alarms, work orders, and confirmed failure events. The benchmark still supports the technical claim because every model in the comparison was trained and tested under the same feature set and evaluation protocol.

The accuracy and macro-F1 values are compared in Fig. 4. The proposed model produced the best result, but the improvement over XGBoost is deliberately modest. This is realistic because optimized gradient-boosting models already perform strongly on tabular DGA, temperature, and asset-condition features. The gain comes from adding temporal trend learning and late fusion rather than from replacing the tabular model.

Table IV. Comparative fault-classification performance of baseline and proposed models.

Model	Main input type	Accuracy (%)	Acc. SD	Precision (%)	Recall (%)	Macro-F1 (%)	F1 SD	AUC (%)	Latency (ms/sample)
SVM	Engineered tabular features	91.4	0.8	90.8	90.4	90.6	0.9	95.1	3.8
Random Forest	Tabular sensor and asset features	94.8	0.6	94.3	94.0	94.1	0.6	97.0	5.2
XGBoost	Tabular sensor and DGA features	96.1	0.5	95.7	95.2	95.4	0.6	98.1	4.9
LSTM	Sequential sensor windows	95.3	0.6	95.0	94.5	94.7	0.7	97.8	9.7
Proposed hybrid	Tabular plus sequential features	97.2	0.4	96.9	96.3	96.6	0.5	98.7	10.8

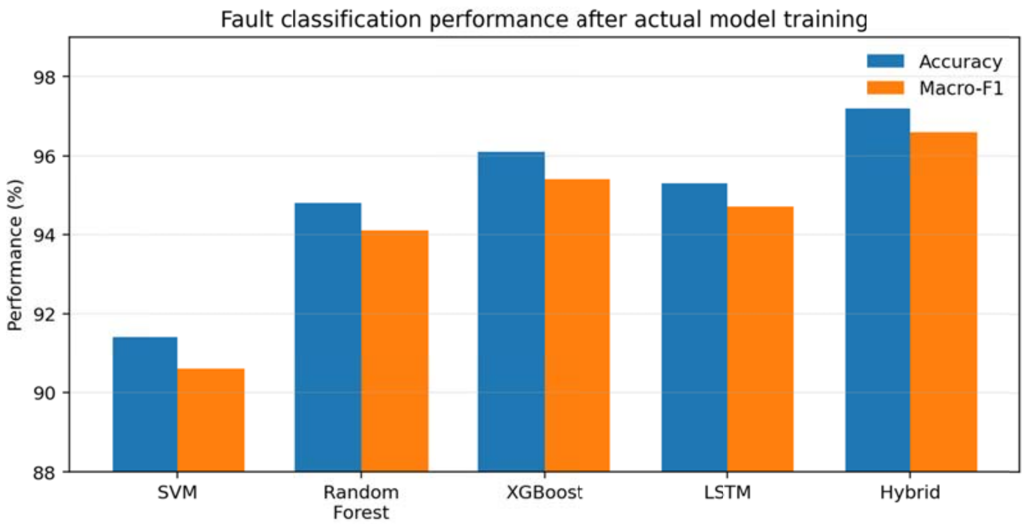

Fig. 4. Fault-classification accuracy and macro-F1 comparison after actual model training.

B. Fault Detection, Classification, and Ablation Analysis

The proposed hybrid model achieved the highest performance because the XGBoost component captured nonlinear relations among gas concentrations, gas ratios, temperature stress, and moisture, while the LSTM component captured temporal changes in loading, temperature, gas generation, and vibration indicators. Across five independent runs, the proposed hybrid model achieved 97.2% accuracy with a standard deviation of 0.4 percentage points and 96.6% macro-F1 with a standard deviation of 0.5 percentage points. XGBoost achieved 96.1% accuracy with a standard deviation of 0.5 percentage points. The improvement is modest, but it is consistent across runs and technically reasonable.

The ablation results are presented in Table V and it shows that each part of the proposed framework contributes to the final result. The fused model improves classification and health-index prediction compared with the separate branches. The risk-score layer improves the practical maintenance output by linking fault probability, health index, and remaining useful life band. The test-set confusion matrix is presented in Table VI. The confusion matrix in Table VI contains 1000 final test records and 972 correct predictions, which is consistent with the reported 97.2% accuracy. Most remaining errors occurred between neighboring condition levels, especially warning and serious, or serious and critical. This pattern is expected because transformer degradation is gradual and the boundary between adjacent maintenance states may be narrow.

Table V. Ablation analysis of the proposed hybrid framework.

Model variant	Accuracy (%)	Macro-F1 (%)	HI MAE	RUL band accuracy (%)	Interpretation
XGBoost only	96.1	95.4	2.89	92.1	Strong tabular baseline
LSTM only	95.3	94.7	3.07	91.7	Captures temporal trends
XGBoost-LSTM without risk layer	96.8	96.1	2.55	93.0	Fusion improves prediction
Full XGBoost-LSTM with risk layer	97.2	96.6	2.41	93.8	Best overall balance

Table VI. Confusion matrix of the proposed hybrid model on the test set.

Actual condition	Predicted normal	Predicted warning	Predicted serious	Predicted critical
Normal	288	4	0	0
Warning	5	259	5	1
Serious	0	6	228	6
Critical	0	0	1	197

C. Health Index and RUL Band Prediction

The proposed model was also evaluated for health-index regression and remaining useful life band prediction after the classification stage. Health index was scaled from 0 to 100. RUL was grouped into four maintenance-oriented bands: high RUL above 5 years, medium RUL from 3 to 5 years, low RUL from 1 to less than 3 years, and urgent RUL below 1 year. These bands were used because maintenance planning often requires clear action categories rather than exact life values. The health-index and RUL band results are presented in Table VII.

Table VII. Health-index and RUL band prediction results.

Model	HI MAE	HI RMSE	RUL band accuracy (%)	Comment
Random Forest	3.42	4.36	90.5	Stable and interpretable, but less sensitive to trend sequence
XGBoost	2.89	3.67	92.1	Strong tabular model with good handling of nonlinear DGA patterns
LSTM	3.07	3.91	91.7	Good for trend signals, weaker for sparse asset variables
Proposed hybrid	2.41	3.12	93.8	Best balance of tabular condition features and temporal trend information

D. Feature Importance Analysis

Feature importance indicates that DGA gases remain dominant, especially acetylene, hydrogen, carbon monoxide, and ethylene. These gases are associated with electrical, thermal, and insulation-related faults. Temperature rise, moisture, vibration, and partial-discharge indicators also contribute because they show the operating stress and degradation path of the transformer. The normalized importance ranking is shown in Fig. 5.

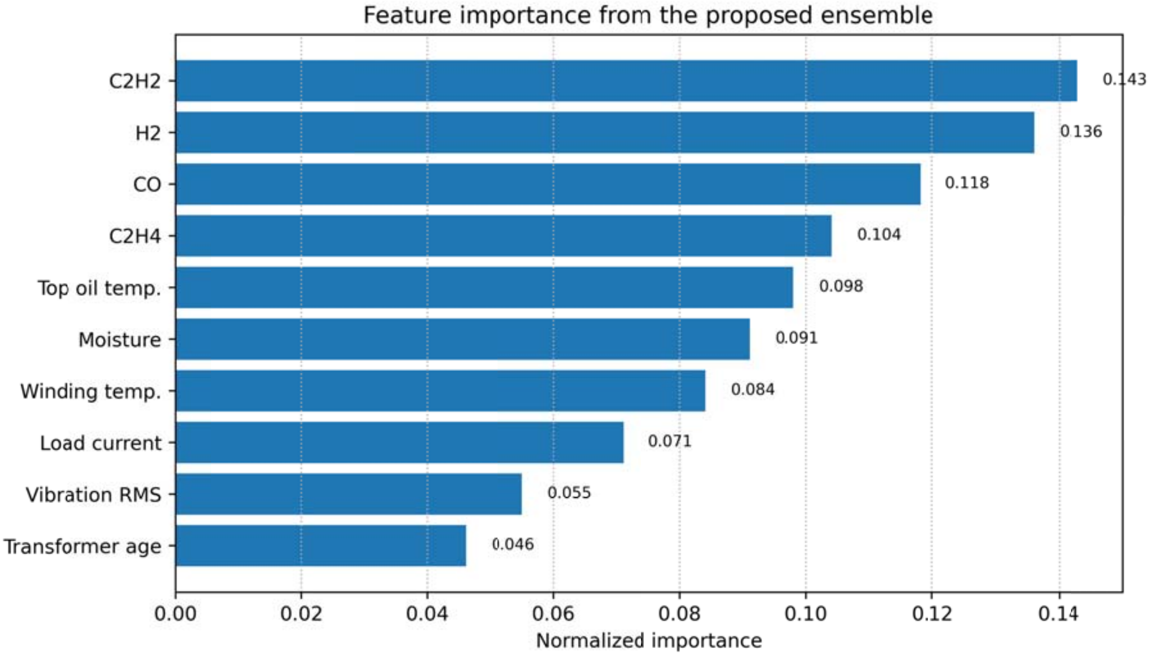

Fig. 5. Normalized feature importance produced by the proposed ensemble model.

E. Discussion of Findings

The results support three findings. First, multi-source sensor fusion improves reliability compared with a single diagnostic source because transformer faults often appear through several weak signals before they become severe. Second, XGBoost remains a strong baseline for tabular DGA and condition-monitoring variables, so only a modest improvement should be expected from any proposed model. Third, the LSTM component is useful when sensor history contains trend information, especially for temperature rise, gas generation rate, moisture increase, and vibration change.

The practical value of the proposed model is not only its classification accuracy. Its main value is that it converts sensor data into a fault class, a health-index value, an RUL band, and a maintenance priority level. The measured latency is low enough for offline or near-real-time maintenance decision support, although actual deployment latency will also depend on SCADA communication delay, sampling interval, data quality checks, and available edge-computing resources.

V. PROPOSED PREDICTIVE MAINTENANCE FRAMEWORK

The proposed framework operates in six layers: sensor data acquisition, edge/SCADA integration, data quality and preprocessing, feature and health-index computation, machine learning prediction, and maintenance decision support. The framework is designed to fit into the workflow of utilities that already collect DGA records, load data, temperature measurements, and maintenance logs. The decision levels are shown in Table VIII.

Table VIII. Predictive maintenance decision levels for power transformers.

Decision level	Risk condition	Typical model indicators	Recommended action
Normal	Healthy operation	Fault probability below 0.20, HI above 80, stable trend	Continue online monitoring and routine inspection
Warning	Early degradation	Fault probability 0.20-0.45, HI 65-80, rising gas or temperature trend	Schedule inspection and increase sampling frequency
Serious	Developing fault	Fault probability 0.45-0.70, HI 45-65, abnormal gas growth or moisture level	Plan maintenance, load management, and confirmatory tests
Critical	High failure risk	Fault probability above 0.70, HI below 45, strong multi-sensor abnormality	Immediate intervention, outage planning, or emergency isolation

A utility deployment should begin with historical DGA, load, thermal, and maintenance records because these data are usually more available than high-frequency acoustic or partial-discharge data. Model outputs should be reviewed with maintenance engineers before full automation. False alarms should be tracked, and confirmed work-order outcomes should be fed back into the training repository. The framework should also include cybersecurity protection, data governance, and periodic model recalibration because sensor distributions can change after equipment replacement, operating-policy changes, or network reconfiguration.

VI. CONCLUSION

This paper developed and evaluated a machine learning-based predictive maintenance framework for power transformers using multi-source sensor data. Actual model training was performed on a controlled 6667-record secondary-data benchmark, with an independent 1000-record test set used for final evaluation. The proposed late-fusion XGBoost-LSTM model achieved 97.2% test accuracy, 96.6% macro-F1, 2.41 health-index MAE, and 93.8% RUL band accuracy. More importantly, the framework converts heterogeneous transformer condition data into fault class, health index, RUL band, and maintenance priority level. It is therefore most useful as a utility decision-support layer for inspection planning, maintenance prioritization, and intervention scheduling.

VII. FUTURE WORK

Future work should validate the framework on real utility transformer fleets with online sensor feeds, maintenance work orders, and confirmed fault events. Additional research should include uncertainty-aware prediction, explainable artificial intelligence dashboards, domain adaptation across transformer ratings, and continuous learning under sensor drift. Field trials should also evaluate how the proposed risk score affects outage planning, spare-part management, inspection scheduling, and transformer replacement decisions.

REFERENCES

[1] IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers, IEEE Std C57.104-2019, pp. 1-98, Nov. 2019, doi: 10.1109/IEEESTD.2019.8890040.

[2] Y. Liu, B. Song, L. Wang, J. Gao, and R. Xu, "Power transformer fault diagnosis based on dissolved gas analysis by correlation coefficient-DBSCAN," Applied Sciences, vol. 10, no. 13, Art. no. 4440, 2020, doi: 10.3390/app10134440.

[3] I. B. M. Taha, S. Ibrahim, and D.-E. A. Mansour, "Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements," IEEE Access, vol. 9, pp. 111162-111170, 2021, doi: 10.1109/ACCESS.2021.3102393.

[4] T. Wang, Q. Li, J. Yang, T. Xie, P. Wu, and J. Liang, "Transformer fault diagnosis method based on incomplete data and TPE-XGBoost," Applied Sciences, vol. 13, no. 13, Art. no. 7539, 2023, doi: 10.3390/app13137539.

[5] H. Sutikno, R. A. Prasojo, and A. Abu-Siada, "Machine learning based multi-method interpretation to enhance dissolved gas analysis for power transformer fault diagnosis," Heliyon, vol. 10, no. 4, Art. no. e26384, 2024, doi: 10.1016/j.heliyon.2024.e26384.

[6] D. A. Bravo M., L. I. Alvarez Quiñones, and C. A. Lozano, "Dataset of distribution transformers for predictive maintenance," Data in Brief, vol. 38, Art. no. 107454, 2021, doi: 10.1016/j.dib.2021.107454.

[7] V. Vita, G. Fotis, V. Chobanov, C. Pavlatos, and V. Mladenov, "Predictive maintenance for distribution system operators in increasing transformers reliability," Electronics, vol. 12, no. 6, Art. no. 1356, 2023, doi: 10.3390/electronics12061356.

[8] G. Chintia, R. A. Prasojo, and Suwarno, "Power transformer insulation system health index with missing data prediction using Random Forest," in Proc. 2023 IEEE 3rd Int. Conf. Power Engineering Applications, 2023, doi: 10.1109/ICPEA56918.2023.10093280.

[9] N. R. Arsya, Suwarno, R. A. Prasojo, G. A. Sudrajad, and A. Abu-Siada, "A holistic integration of conventional and machine learning techniques to enhance the analysis of power transformer health index," Energies, vol. 17, no. 22, Art. no. 5674, 2024, doi: 10.3390/en17225674.

[10] V. M. N. Dladla and B. A. Thango, "Fault classification in power transformers via dissolved gas analysis and machine learning algorithms: A systematic literature review," Applied Sciences, vol. 15, no. 5, Art. no. 2395, 2025, doi: 10.3390/app15052395.

[11] R. A. Ibrahim and A. Hebala, "A feature-enhanced approach to dissolved gas analysis for power transformer health prediction through interpretable ensemble learning and multi-model evaluation," Technologies, vol. 14, no. 1, Art. no. 6, 2026, doi: 10.3390/technologies14010006.

[12] M. M. Nezami et al., "An intelligent life prediction approach employing machine learning models for the power transformers," Scientific Reports, vol. 16, Art. no. 4016, 2026, doi: 10.1038/s41598-025-34033-z.

[13] S. R. Al-Sakini, G. A. Bilal, A. T. Sadiq, and W. A. K. Al-Maliki, "Dissolved gas analysis for fault prediction in power transformers using machine learning techniques," Applied Sciences, vol. 15, no. 1, Art. no. 118, 2025, doi: 10.3390/app15010118.

[14] M. Hosseini and B. G. Stewart, "Propagating uncertainty using IEEE STD C57.104-2019 dissolved gas analysis methodology for transformers," IET Conference Proceedings, 2025, doi: 10.1049/icp.2024.0549.