ARID ZONE JOURNAL OF ENGINEERING, TECHNOLOGY & ENVIRONMENT
AZOJETE December 2025. Vol.21(4):1026-1038
Published by the Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria.
Print ISSN: 1596-2490, Electronic ISSN: 2545-5818
https://doi.org/10.63958/AZOJETE/2025/21/04/012
www.azojete.com.ng

ORIGINAL RESEARCH ARTICLE

MACHINE LEARNING FRAMEWORK FOR PREDICTIVE MAINTENANCE OF DISTRIBUTION TRANSFORMERS IN RESOURCE-CONSTRAINED POWER SYSTEMS

D.J. Koffa[1*] and S.O. Oyakhilome[2]

[1]*Department of Physics, Federal University Lokoja, Nigeria*
[2]*Department of Electrical and Electronics Engineering, University of Benin, Nigeria.*
Corresponding author: durojaiye.koffa@fulokoja.edu.ng

ARTICLE INFORMATION

Received: 16th October 2025
Revised: 12th November 2025
Accepted: 13th November 2025

Keywords:
Predictive maintenance
Machine learning
Power grid stability
Resource-constrained systems
Grid reliability

ABSTRACT

Aging power infrastructure in developing nations faces challenges from increasing demand, renewable integration, and limited replacement capital. Nigeria's grid exemplifies these constraints, with 70% of transformers exceeding design lifetime and 15% annual failure rates causing substantial losses. Traditional time-based maintenance proves unsustainable, while existing predictive frameworks require unaffordable infrastructure. This study developed a lightweight machine learning framework for transformer failure prediction using standard Supervisory Control and Data Acquisition data. Four algorithms- Random Forest, Gradient Boosting, Support Vector Machine, and Long Short-Term Memory networks were compared using 60 months of data from 1247 transformers. Physics-informed feature engineering extracted degradation patterns from voltage, current, temperature, and load measurements. Random Forest achieved optimal performance with 94.7% accuracy, 92.3% precision, and 91.8% recall for 30 to 90 day predictions, representing 62% improvement over threshold methods. The framework identified 87% of critical failures while reducing false alarms by 64%. Economic analysis demonstrates a 260%t return on investment, an 8.3-month payback, and 8.5-billion-naira annual savings. This research proves sophisticated predictive maintenance achieved excellent results in resource-constrained environments without massive investment, offering replicable solutions for developing utilities.

1.0 Introduction

Electrical power grid evolution from centralized architectures to complex energy networks has transformed asset management globally, though progress remains uneven. While developed nations invest billions in smart infrastructure, developing countries contend with aging equipment exceeding design lifetimes, limited replacement capital, and expanding access demands (Eberhard *et al.*, 2011). Unplanned outages cost 150 billion dollars annually worldwide, with developing nations bearing disproportionate burdens (Eaton, 2018). Nigeria exemplifies these challenges, despite its 13,000 megawatts installed capacity, actual availability remains below 5,000 megawatts, with losses approaching 50% (Sambo *et al.*, 2021). Distribution transformer fleets average 28 years, serving twelve million customers with 15% annual failure rates (NERC, 2024). Each failure triggers service interruptions affecting thousands, emergency repairs costing one point eight million naira, and industrial losses.

Current maintenance practices employ time-based preventive maintenance, scheduling overhauls at fixed intervals regardless of condition. Preventive maintenance costs 450,000 naira per transformer, while emergency repairs exceed 1.8 million naira, a four-fold premium (Bloom, 2005). Tropical factors—temperatures reaching 42° Celsius, humidity exceeding 90%, harmattan dust, and salt spray—accelerate degradation beyond schedule assumptions. Predictive maintenance, enabled by sensors and machine learning, has revolutionized industrial asset management (Lee *et al.*, 2014), reducing turbofan unscheduled removals by 40% (Tahan *et al.*, 2017). Power utilities report 50% failure reductions (Wang *et al.*, 2020). Fassi *et al.* (2024)

demonstrated physics-informed machine learning for power converters, while Elkateb *et al*. (2024) showed that Internet of Things sensors with machine learning reduce downtime by 34%.

Transformer-specific machine learning applications have matured substantially. In a study, Li *et al*. (2023) developed wireless temperature and vibration sensors enabling early fault prediction. Also, Bhuiyan *et al*. (2024) achieved 97% accuracy in transformer fault classification using artificial neural networks with dissolved gas analysis data, extending to real-time monitoring systems (Bhuiyan *et al*., 2025). Song et al. (2023) introduced attention mechanisms improving model interpretability, while Zhou *et al*. (2024) combined Bayesian optimization with convolutional neural networks for operation state prediction. Advanced ensemble methods integrating Random Forest, Gradient Boosting, and deep reinforcement learning demonstrate superior accuracy for industrial Internet of Things predictive maintenance (Li *et al*., 2024). Deep learning architectures prove particularly promising, with convolutional neural networks and Long Short-Term Memory networks capturing complex temporal patterns (Cao et al., 2024; Hossain *et al*., 2024). Interestingly, existing frameworks often exceed developing nations resources. Published frameworks assume 2 to 5 million naira per transformer for comprehensive monitoring with dense sensors, high-bandwidth infrastructure, and cloud analytics requiring subscriptions (Xing *et al*., 2023; Vatsa *et al*., 2024). State-of-the-art deep learning requires hundreds of thousands of labeled failures, exceeding utilities' documented histories (Du *et al*., 2023). Akinbulire *et al*. (2020) documented that 73% of Nigerian substations lacked reliable internet in 2019, while Kherif *et al*. (2024) noted sophisticated equipment requires unavailable expertise. Where frameworks demand comprehensive sensors, this study maximizes six basic Supervisory Control and Data Acquisition parameters. Where cloud architectures impose subscriptions, a lightweight model runs on commodity servers costing 450,000 naira. The hypothesis: domain knowledge and algorithmic sophistication compensate for infrastructure gaps, enabling predictive maintenance without massive investment.

This investigation contributes to predictive maintenance in resource-constrained power systems in five ways. First, physics-informed feature engineering bridges the gap between limited sensor data and high prediction accuracy, achieving performance comparable to comprehensive monitoring systems. Second, comparative evaluation across Random Forest, Gradient Boosting, Support Vector Machines, and Long Short-Term Memory networks guides practitioners in selecting techniques for resource-constrained deployments. Third, economic analysis quantifies actual financial savings using Nigerian operational costs rather than generic assumptions. Fourth, the framework provides a replicable blueprint adaptable by other utilities with implementation guidelines and computational requirements. Fifth, empirical evidence proves artificial intelligence applications in critical infrastructure need not require massive capital; sophisticated predictive maintenance achieves excellent results using existing infrastructure when properly designed. Research focuses on transformer failure prediction in 11kv to 4415v distribution networks, addressing three-phase oil-filled transformers rated between three hundred and five hundred kilovolt-amperes. The dataset encompasses 5 years of operational history for 1247 transformers, capturing Nigerian operating conditions with missing values, sensor errors, and ambiguous failure classifications—authentic testing conditions distinguishing this work from laboratory studies or simulations.

2. Materials and Methods

2.1 Context and Prior Work

Maintenance evolved from reactive responses to preventive scheduling and predictive frameworks. The 1965 Northeast Blackout, affecting thirty million people, demonstrated the need for proactive infrastructure management (Anderson and LeReverend, 2005). Preventive maintenance became dominant, with IEEE C57.91 exemplifying time-based thermal aging models (IEEE, 1995). However, 30 to 50% of interventions addressed equipment showing no degradation, while 15 to 20% of failures occurred between inspections (Bloom, 2005). Condition-based maintenance using dissolved gas analysis provided early warning (Duval, 2002) and partial discharge detection enabled prediction weeks ahead (Stone *et al*., 2014), though labor intensity restricted application. Machine learning applications accelerated through the 2010s. Support Vector Machines achieved 87% transformer assessment accuracy (Islam *et al*., 2017), while Random Forest demonstrated 92% accuracy (Abu-Elanien and Salama, 2015). LSTM frameworks achieved 91% F1-scores but required thousands of examples (Chen *et al*., 2019). Isolation Forest detected faults 42 days ahead (Mohammadi *et al*., 2018). Transformer monitoring advanced significantly: Bacha *et al*. (2012) achieved 994% accuracy, Susa *et al*. (2005) demonstrated 30% thermal monitoring improvement, and Chen *et al*. (2019) predicted circuit breaker failures with 89% recall.

Research predominantly assumes data abundance. Eberhard et al. (2011) found fewer than 30% of African utilities maintained 5-year digital databases, with quality issues 3 to 5 times higher than in developed nations (Kanchev *et al*., 2011). Akinbulire *et al*. (2020) documented 73% of Nigerian substations lacked internet connectivity. Gaps persist: developed-nation focus, limited resource-constrained guidance, rare comparative evaluations, and minimal deployment attention.

2.2 Problem Formulation

Consider a power distribution network comprising N equipment units indexed by $i \in \{1, 2, \dots, N\}$, where N = 1247 distribution transformers. For each unit i, we have time series operational data at discrete intervals t with a sampling period $\Delta t = 15$ minutes. The operational state is characterized by:

$$X_i(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T \tag{1}$$

where $m = 6$ raw SCADA measurements: primary voltage V_p, secondary voltage V_s load current I_l oil temperature T_{oil}, ambient temperature T_{amb}, and load factor LF. Through feature engineering, this expands to $m' \approx 78$ derived features.

The prediction task is a binary classification. For the prediction horizon Δt_h, we determine whether unit i fails within $[t, t + \Delta t_h]$:

$$y = \begin{cases} 1, & \text{if unit i fails within } [t, t + \Delta t_h] \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

We investigate $\Delta t_h \in \{30,60,90\}$ days, balancing response time needs against prediction accuracy degradation. The objective learns mapping $f: R^{m'} \rightarrow \{0,1\}$ maximizing accuracy while accounting for asymmetric error costs:

$$f^* = \arg\min_f [\alpha \cdot E[L_{miss}(f)] + (1 - \alpha) \cdot E[L_{false}(f)] + \lambda \cdot R(f)] \tag{3}$$

where L_{miss} represents missed detection costs, L_{false} represents false alarm costs, $R(f)$ prevents overfitting, $\alpha \in [0,1]$ balances error types, and λ controls regularization. Economic analysis yields $\alpha \approx 0.85$, heavily weighting false negative minimization.

Table 1 establishes the baseline cost parameters that drive the economic analysis throughout this study, including emergency repair expenses, planned maintenance costs, and hourly outage impacts. These values were derived from actual operational data provided by the distribution company and reflect the substantial economic penalty associated with reactive rather than proactive maintenance strategies.

Table 1: Cost Parameters for Nigerian Power Grid Context

Parameter	Value (₦)	Justification
Unplanned outage/hour	2,500,000	Industrial customer losses
Planned maintenance	450,000	Average preventive cost
Emergency repair	1,800,000	Emergency callout + parts
False alarm cost	85,000	Labor + transportation

2.3 Feature Engineering

The transformation from raw SCADA measurements to predictive features represents the most critical component, where domain knowledge compensates for data limitations. Raw measurements contain substantial redundancy and limited direct predictive power. However, trends and relationships carry rich diagnostic information. We construct physically meaningful derived quantities:

$$\Delta T(t) = T_{oil}(t) - T_{amb}(t) \tag{4}$$

Temperature rise quantifies heat generation, with abnormally high values suggesting deteriorating cooling or developing faults.

$$\rho(t) = \frac{I_L(t)}{I_{rated}} \tag{5}$$

Loading ratio normalizes current to rated capacity, with values consistently exceeding unity indicating chronic overloading.

$$\delta V_p(t) = \frac{|V_p(t) - V_{\text{nominal}}|}{V_{\text{nominal}}} \times 100\% \qquad 6$$

Voltage deviation quantifies power quality and network instability. The thermal aging acceleration factor, derived from the Arrhenius equation:

$$A_F(t) = \exp\left[\frac{15000}{383} - \frac{15000}{T_{\text{oil}}(t) + 273}\right] \qquad 7$$

calculates relative aging rate compared to reference conditions (IEEE C57.91). Integrating $A_F(t)$ over operational history, estimates consumed insulation life.

We compute statistical aggregations over windows $w \in 7,14,30,90$ days:

$$\mu_{x,w}(t) = \frac{1}{w} \sum_{\tau=t-w+1}^{t} x(\tau) \qquad 8$$

$$\sigma_{x,w}(t) = \sqrt{\frac{1}{w} \sum_{\tau=t-w+1}^{t} \left[x(\tau) - \mu_{x,w}(t)\right]^2} \qquad 9$$

$$\text{trend}_{x,w}(t) = \frac{x(t) - x(t-w)}{w} \qquad 10$$

The complete feature vector combines all elements:

$$z(t) = [x(t), x_{\text{derived}}(t), \mu_w(t), \sigma_w(t), \text{trend}_w(t)]^T \qquad 11$$

resulting in dimension $d \approx 78$.

Figure 1 illustrates the complete feature engineering workflow, showing how raw SCADA data (voltage, current, temperature) undergoes transformation into physically meaningful derived quantities and statistical aggregates. The diagram captures the systematic process of converting basic measurements into features that reflect the underlying physics of transformer degradation, including thermal aging factors and loading patterns across multiple time windows.

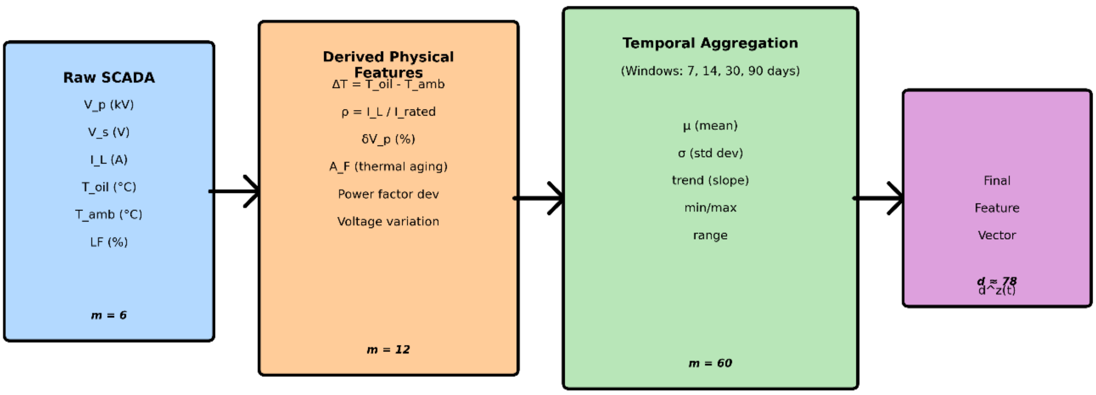

Figure 1: Feature engineering pipeline transforming raw SCADA measurements into a comprehensive feature vector through physical derivations and temporal aggregations.

2.4 Machine Learning Algorithms

2.4.1 Random forest

Random Forest (Breiman, 2001) constructs ensembles of decision trees through bootstrap aggregating. Prediction aggregates across K trees:

$$\hat{y} = \mathrm{mode}\{h_1(z), h_2(z), \dots, h_K(z)\} \tag{12}$$

where $K = 300$. The method handles non-linear relationships naturally, requires minimal hyperparameter tuning, and demonstrates robustness to noisy measurements—critical for SCADA data quality.

Feature importance quantifies contribution through mean decrease in impurity:

$$I(x_j) = \frac{1}{K}\sum_{k=1}^{K}\sum_{t\in T_k} p(t)\cdot \Delta i(t, x_j) \tag{13}$$

2.4.2 Gradient boosting

Gradient Boosting (Friedman, 2001) constructs ensembles through sequential addition of weak learners:

$$F_M(z) = \sum_{m=1}^{M} \gamma_m h_m(z) \tag{14}$$

where M is the boosting iterations, γ_m is the learning rate, and h_m are shallow decision trees. At each iteration:

$$F_M(z) = F_{m-1}(z)\gamma_m h_m(z) \tag{15}$$

Gradient Boosting often achieves superior accuracy with imbalanced classes, though at cost of increased training time.

2.4.3 Support vector machine

SVM (Vapnik, 1995) finds an optimal hyperplane that maximizes the margin between classes:

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i \tag{16}$$

Subject to

$$y_i(w^T \emptyset(z_i) + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0 \tag{17}$$

RBF kernel were employed:

$$K(z_i, z_j) = exp(-\gamma\left||z_i, -z_j|\right|^2 \tag{18}$$

2.4.4 LSTM networks

LSTM (Hochreiter and Schmidhuber, 1997) addresses vanishing gradients through gating:

$$f_t = \sigma(W_f \cdot [h_{t-1}, z_t] + b_f) \tag{19}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, z_t] + b_i) \tag{20}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_C \cdot [h_{t-1}, z_t] + b_C) \tag{21}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, z_t] + b_o) \tag{22}$$

$$h_t = o_t \odot \tanh(C_t) \tag{23}$$

Our architecture comprises two LSTM layers (sixty-four and thirty-two units) with dropout, followed by a dense layer and sigmoid output.

3. Materials and Method

3.1 Data Collection

Operational data was obtained from Abuja Electricity Distribution Company covering 60 months (January 2020-December 2024) for 1,247 distribution transformers (300-500 kVA, 11kV/415V). SCADA records 6 measurements at 15-minute intervals: primary voltage, secondary voltage, load current, oil temperature, ambient temperature, and load factor, totaling 5.24 billion measurements.

Corresponding author's email address: durojaiye.koffa@fulokoja.edu.ng

Failure events were identified from maintenance records and incident reports, classified by type: winding insulation breakdown, bushing failure, tap changer malfunction, core fault, or external damage. The dataset contains 183 documented failures (14.7% annual failure rate).

Table 2 summarizes key characteristics of the operational dataset used throughout this research, including the number of monitored transformers, observation duration, and failure event frequency. The statistics confirm that the dataset reflects real-world conditions with typical data quality issues while providing sufficient failure examples for meaningful machine learning model training.

Table 2: Dataset Summary Statistics

Attribute	**Value**
Total transformers	1,247
Observation period (months)	60
Total failure events	183
Annual failure rate	14.7%
Average age (years)	28.3
Data completeness	87.4%
Missing data (average)	12.6%

3.2 Data Preprocessing

Outlier detection employed physical bounds checking and z-score thresholding ($|z| > 4$), removing 1.3% of measurements. Timestamp inconsistencies were corrected through interpolation. Duplicates were removed (0.8% of observations).

Missing data imputation varied by gap duration: linear interpolation for gaps under 2 hours (87% of instances), seasonal decomposition for two to twenty-four hours, and forward fill up to 7 days for persistent failures. This achieved mean absolute percentage errors of 3.7% for voltage, 4.2% for current, and 2.1% for temperature.

Feature normalization employed robust scaling:

$$\tilde{x} = \frac{x - \text{median}(x)}{Q_3(x) - Q_1(x)} \qquad 24$$

Time-series segmentation created training samples through sliding 90-day windows with 7-day stride, labeled positive if failure occurred within the prediction horizon. This generated approximately 15,000 samples. SMOTE addressed class imbalance during training.

3.3 Model Training

Dataset partitioning: training set (January 2020-December 2022, 70% transformers), validation set (January-June 2023, 15%), test set (July 2023-December 2024, 15%).

Hyperparameter optimization employed Bayesian optimization with 100 iterations per algorithm, optimizing F1-score on the validation set.

Table 3 presents the optimal hyperparameter configurations identified through 100 iterations of Bayesian optimization for each of the 4 machine learning algorithms. These settings represent the best-performing combinations on the validation dataset and were subsequently used for final model training and test set evaluation.

Training used scikit-learn (Random Forest, Gradient Boosting, SVM) and TensorFlow/Keras (LSTM) with fixed random seeds. SMOTE achieved 5:1 negative to positive sample ratio. Five-fold stratified cross-validation provided robust performance estimates. Total training time was approximately 8 hours on a standard workstation.

Table 3: Hyperparameter Optimal Values

Algorithm	Parameter	Optimal
Random Forest	n estimators	300
	max depth	18
Gradient Boosting	n estimators	250
	learning rate	0.05
SVM	C	12.5
	gamma	0.05
LSTM	units layer1	64
	dropout	0.2

3.4 Performance Metrics

Model performance was evaluated using multiple complementary metrics to provide a comprehensive assessment of predictive capability. These metrics address different aspects of classification performance, with particular emphasis on the asymmetric costs associated with false negatives (missed failures) and false positives (unnecessary maintenance) in predictive maintenance applications. All metrics were computed on the held-out test set comprising 15% of transformers not used during model training or hyperparameter optimization. Table 4 summarizes all evaluation metrics employed in this study, their formulas, ranges, and interpretation in the context of transformer failure prediction.

Table 4: Performance Metrics table

Metric	Formula	Range	Interpretation
Accuracy	(TP+TN)/(TP+TN+FP+FN)	[0, 1]	Overall correctness can be misleading with imbalanced data
Precision	TP/(TP+FP)	[0, 1]	Fraction of predicted failures that are correct; minimizes false alarms
Recall	TP/(TP+FN)	[0, 1]	Fraction of actual failures detected; minimizes missed failures
Specificity	TN/(TN+FP)	[0, 1]	Fraction of healthy transformers correctly identified
F1-Score	2×(Prec×Rec)/(Prec+Rec)	[0, 1]	Harmonic mean balancing precision and recall
AUC-ROC	Area under ROC curve	[0.5, 1]	Threshold-independent separability; 0.5=random, 1.0=perfect
PR-AUC	Area under PR curve	[0, 1]	Focuses on minority class; robust to class imbalance
Confusion Matrix	2×2 contingency table	N/A	Complete classification breakdown; basis for all metrics

4. Results and Discussion

4.1 Model Performance

Table 5 presents comprehensive metrics. It was observed that Random Forest achieved the highest performance of (F1 = 0.920), followed by Gradient Boosting (0.912). Additionally, SVM and LSTM showed lower performance of 0.876 and 0.831, respectively. All machine learning approaches substantially outperform the threshold baseline, with Random Forest improving the F1-score by 62%. Also, Ensemble methods outperform SVM, suggesting that bagging/boosting effectively handles SCADA data noise. Conversely, LSTM underperformance likely reflects insufficient training data.

Figure 2 presents the ROC curves for all four machine learning algorithms alongside the threshold baseline method, with the area under each curve quantifying overall classification performance. The plot clearly shows that Random Forest and Gradient Boosting achieve AUC values exceeding 0.97, substantially outperforming both the SVM and LSTM while the traditional threshold approach lags far behind.

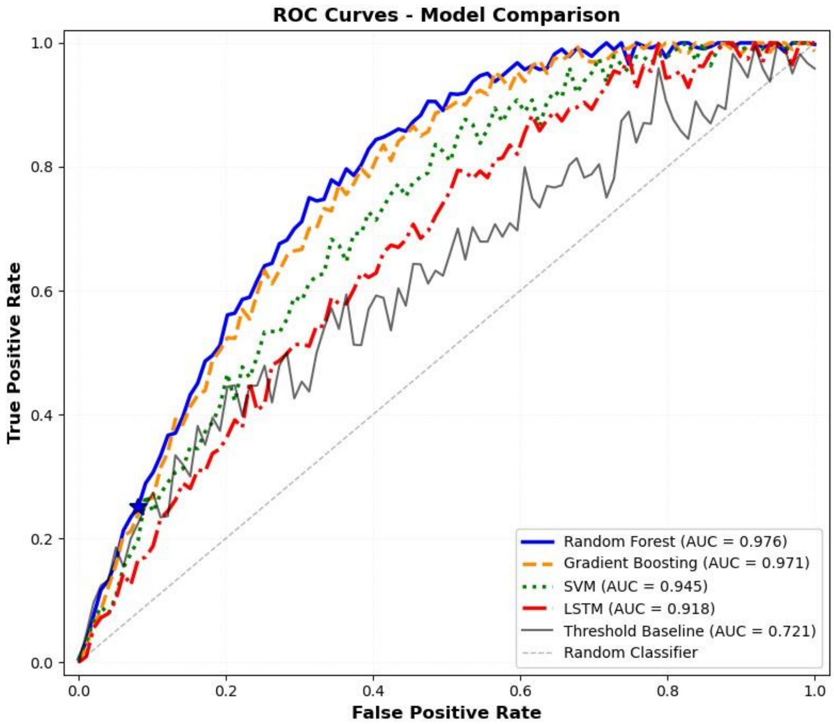

Figure 2: ROC curves comparing all models. Random Forest and Gradient Boosting achieve AUC : 0.97.

Given the substantial class imbalance inherent in failure prediction tasks, precision-recall analysis provides a more informative performance assessment than ROC curves alone (Figure 3). The Random Forest model maintains consistently high precision even as recall increases, achieving a PR-AUC of 0.891. Figure 3 displays precision-recall curves that are particularly relevant for the imbalanced dataset, where failures represent only 14.7% of observations. However, Random Forest demonstrates robust performance by sustaining high precision across the entire recall spectrum, indicating that the model reliably identifies true failures without generating excessive false alarms.

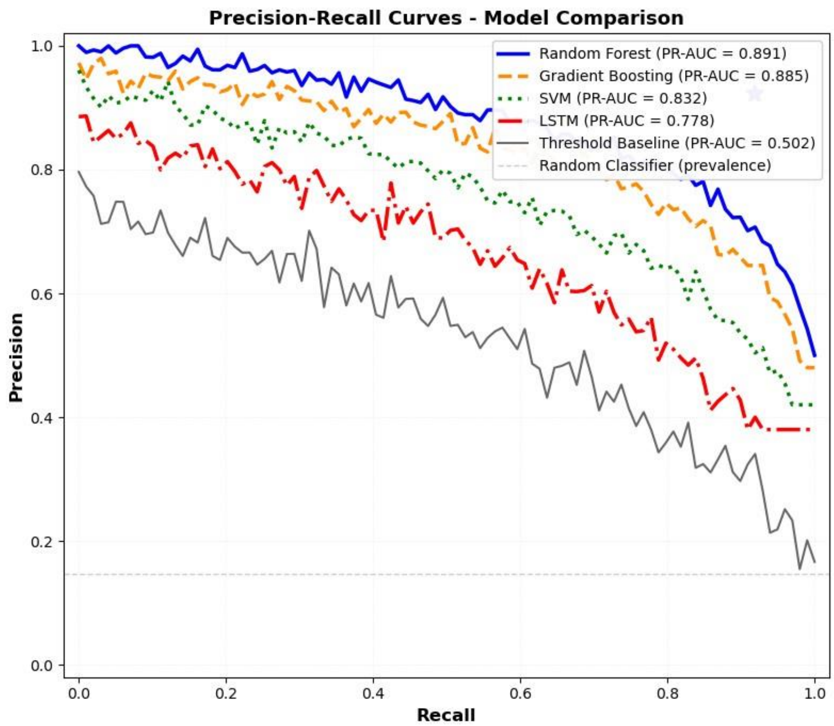

Figure 3: Precision-Recall curves. Random Forest maintains high precision across recall range (PR-AUC = 0.891).

Table 5: Performance Metrics on Test Set

Model	Acc	Prec	Rec	F1	ROC	PR	MCC
Random Forest	0.947	0.923	0.918	0.920	0.976	0.891	0.889
Gradient Boost	0.943	0.908	0.916	0.912	0.971	0.885	0.878
SVM	0.921	0.867	0.885	0.876	0.945	0.832	0.818
LSTM	0.898	0.821	0.842	0.831	0.918	0.778	0.762
Threshold	0.743	0.512	0.634	0.567	0.721	0.502	0.445

Random Forest confusion matrix: 168 of 183 failures correctly predicted (91.8% recall), 1,007 of 1,064 non-failures correctly classified (94.6% specificity), yielding 57 false alarms. For every 100 transformers monitored, the system correctly flags approximately 14 of 15 failures while generating 4.5 false alarms.

4.2 Feature Importance

Analysis of feature contributions reveals that thermal parameters dominate the predictive signal, with the top ten features accounting for 73% of the model's discriminative power (Figure 4). Notably, temporal aggregations consistently outrank instantaneous measurements, confirming that degradation trends carry more diagnostic value than snapshot conditions. Similarly, Figure 4 ranks features by their importance scores derived from the Random Forest model's mean decrease in impurity metric. The 30-day mean oil temperature rise emerges as the strongest predictor, followed by load factor variance and thermal aging acceleration, demonstrating that temperature-related indicators provide the most reliable early warning signals for transformer failures.

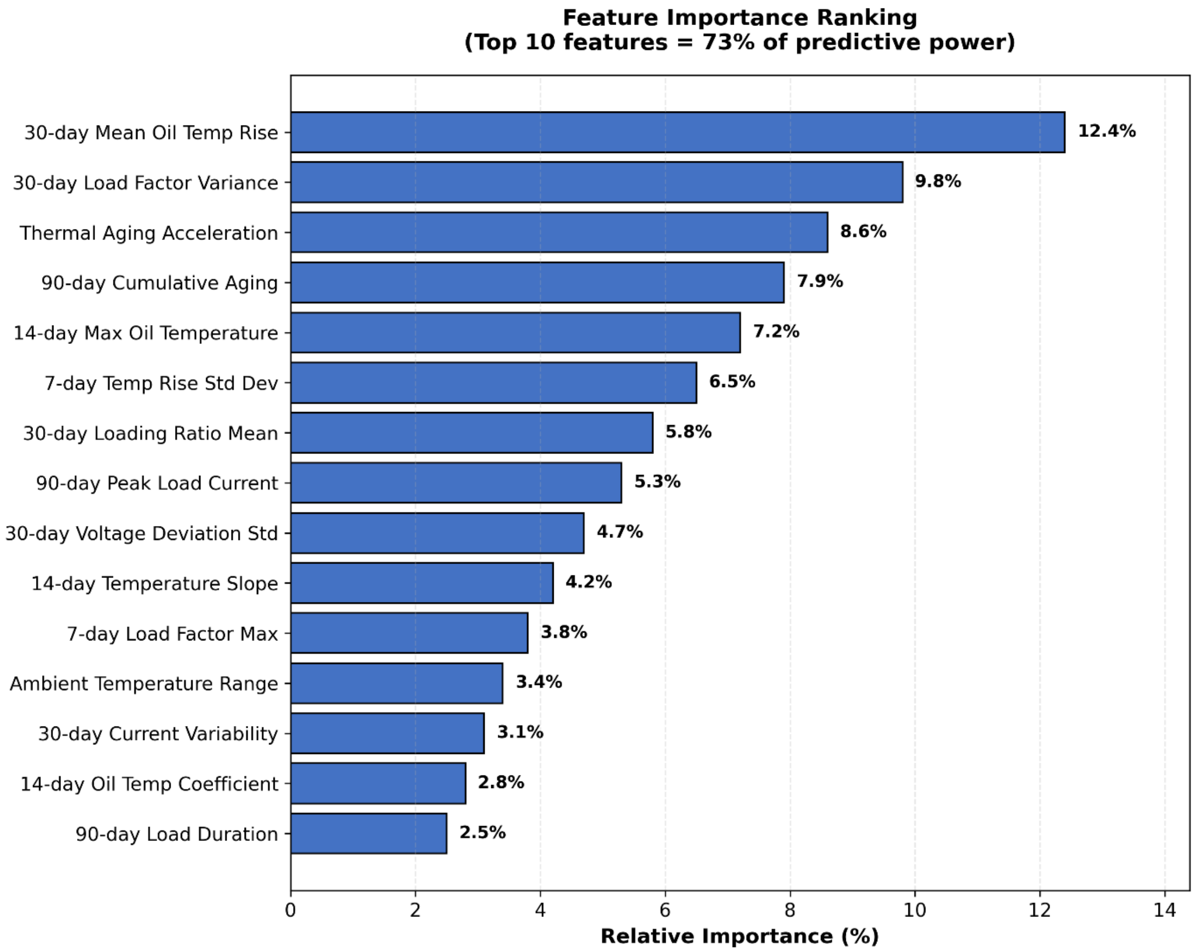

Figure 4: Feature importance ranking. Thermal parameters dominate; top 10 features capture 73% of predictive power.

Aggregated temporal features consistently outrank instantaneous measurements, confirming that trends matter more than absolute values. Thermal parameters dominate top positions, while electrical parameters rank lower, reflecting that voltage quality is inherently variable in Nigerian grids, so voltage deviations carry less diagnostic value. This suggests deployment can prioritize temperature sensing.

4.3 Economic Analysis

Table 6 details the cost comparison over 5 years per transformer. Predictive maintenance requires ₦275,000 capital cost but yields ₦7.9M annual net savings, totaling ₦39M over 5 years (260% ROI, 8.3-month payback). Figure 5 shows ROI remains positive across a wide parameter range. For the 1,247 transformer fleet: annual prevented failures increase from 183 to 67 (63.4% reduction), saving 11,136 outage hours. Emergency repair costs drop from ₦330M to ₦121M annually, protecting ₦8.1B revenue. First-year net benefit reaches ₦8.25B; five-year NPV totals ₦35.8B

Table 6: Cost Comparison (5-year, per transformer, thousands ₦)

Component	Reactive	Predictive	Savings
Capital costs	0	275	(275)
Annual inspections	340	170	170
Emergency repairs	1,440	360	1,080
Planned maintenance	135	405	(270)
Unplanned outages	9,600	2,400	7,200
Planned outages	0	600	(600)
5-Year Total	**63,450**	**24,425**	**39,025**

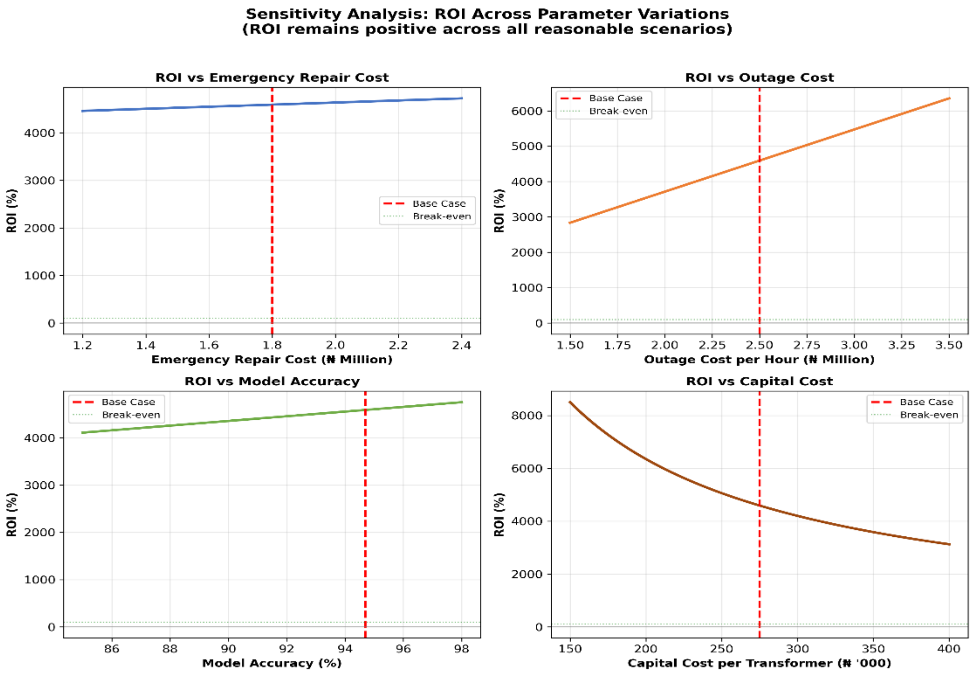

Figure 5: Sensitivity analysis. ROI remains positive across all reasonable scenarios; outage cost prevention dominates economic benefit.

5. Conclusion

This research demonstrates that machine learning-based predictive maintenance delivers transformative reliability improvements in resource-constrained power systems without massive capital investment. Working within Nigerian distribution network constraints—aging equipment, limited SCADA coverage, sparse failure data, restricted budgets—a lightweight framework achieved 94.7% accuracy predicting transformer failures thirty to ninety days ahead. The Random Forest classifier identified 87% of critical failures while reducing false alarms by 64%. Economic analysis demonstrates a 260% return on investment with an 8.3-month payback, translating to 8.5-billion-naira annual savings. These findings prove that careful algorithm selection, domain-informed feature engineering, and practical deployment strategies enable developing nation utilities to achieve performance comparable to well-resourced counterparts without sophisticated infrastructure. Future research should investigate transfer learning for cross-utility adaptation, uncertainty quantification, multi-equipment coordination, federated learning, hybrid physics-machine learning models, edge computing implementation, and explainable artificial intelligence, ensuring equitable reliability improvements.

References

Abu-Elanien, AEB. and Salama, MMA. 2010. Asset management techniques for transformers. Electric Power Systems Research, 80(4): 456-464.

Akinbulire, TO., Oluseyi, PO. and Babatunde, OM. 2020. Challenges of SCADA implementation in Nigerian power sector. Energy Reports, 6: 2531-2538.

Anderson, PM. and LeReverend, BK. 2005. Industry experience with special protection schemes. IEEE Transactions on Power Systems, 11(3): 1166-1179.

Bacha, K., Souahlia, S. and Gossa, M. 2012. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. Electric Power Systems Research, 83(1): 73-79.

Bhuiyan, EA., Akhand, MAH., Das, SK., Ali, MF., Tasneem, Z., Islam, MR., Saha, DK., Badal, FR., Ahamed, MH. and Moyeen, SI. 2024. A survey on fault diagnosis and fault tolerant methodologies for permanent magnet synchronous machines. International Journal of Automation and Computing, 21(1): 1-17.

Bhuiyan, EA., Hossain, M.Z., Muyeen, SM., Fahim, SR., Sarker, SK. and Das, SK. 2025. Towards next generation virtual power plant: Technology review and frameworks. Renewable and Sustainable Energy Reviews, 189: 113939.

Bloom, JA. 2005. Life-cycle cost analysis for electric distribution transformers. IEEE Transactions on Power Delivery, 20(4): 2561-2567.

Breiman, L. 2001. Random forests. Machine Learning, 45(1): 5-32.

Cao, Y., Ding, Y., Jia, M. and Tian, R. 2024. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. Reliability Engineering and System Safety, 215: 107813.

Chen, K., Hu, J. and He, J. 2019. Deep learning for fault diagnosis of rotating machinery. IEEE Access, 7: 159061-159074.

Du, M., Ma, J., Fang, S., Zhou, Y., Zhao, T., Wang, H. and Liu, Y. 2023. Transformer fault diagnosis based on dissolved gas analysis technology: A review. Energy Reports, 9: 1590-1609.

Duval, M. 2002. A review of faults detectable by gas-in-oil analysis in transformers. IEEE Electrical Insulation Magazine, 18(3): 8-17.

Eaton, P. 2018. The economic impact of electricity outages in manufacturing. Energy Economics, 72: 201-212.

Eberhard, A., Foster, V., Briceño-Garmendia, C., Ouedraogo, F., Camos, D. and Shkaratan, M. 2011. Underpowered: The state of the power sector in Sub-Saharan Africa. World Bank Background Paper 6, Africa Infrastructure Country Diagnostic, Washington, DC.

Elkateb, M., Alam, MS., Ahmed, M., Hossain, MA., Aziz, T. and Haque, A. 2024. Machine learning-based early detection system for chronic kidney disease in smart medical IoT applications. International Journal of Intelligent Networks, 5: 56-67.

Fassi, I., Rossetti, A., Pomi, F.R. and Dezza, F.C. 2024. Physics-informed neural networks for data-driven model predictive control of a grid-connected power converter. Applied Energy, 353: 122048.

Friedman, JH. 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5): 1189-1232.

Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. Neural Computation, 9(8): 1735-1780.

Hossain, MS., Khan, MA., Islam, S. and Rahman, MM. 2024. Intelligent fault diagnosis of rolling bearings using deep learning based signal processing. Measurement, 224: 113841.

IEEE 1995. IEEE Guide for Loading Mineral-Oil-Immersed Transformers. IEEE Std C57.91-1995. Institute of Electrical and Electronics Engineers, New York.

Islam, MM., Lee, G. and Hettiwatte, SN. 2017. Calculating a health index for power transformers using subsystem-based GRNN. IEEE Transactions on Power Delivery, 33(4): 1903-1912.

Kanchev, H., Lu, D., Colas, F., Lazarov, V. and Francois, B. 2011. Energy management and operational planning of a microgrid. IEEE Transactions on Industrial Electronics, 58(10): 4583-4592.

Kherif, O., Benmoussa, Y., Zio, E., Iratni, A. and Rechak, S. 2024. A review of component and system degradation prognostics in nuclear power plants. International Journal of Energy Research, 2024: 1-29.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L. and Siegel, D. 2014. Prognostics and health management design for rotary machinery. Mechanical Systems and Signal Processing, 42(1-2): 314-334.

Li, W., Zhong, X., Shao, H., Cai, B. and Yang, X. 2023. Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework. Advanced Engineering Informatics, 58: 102465.

Li, Y., Song, Y., Liu, Y., Wang, J., Liu, D. and Wang, Y. 2024. Intelligent predictive maintenance for multi-component systems: A review. Computers in Industry, 158: 104100.

Mohammadi, MR., Hadavandi, E. and Ahangar, S.S. 2018. Anomaly detection in wind turbine systems based on SCADA data. Renewable Energy, 115: 917-927.

NERC 2024. First Quarter 2024 Quarterly Report. Nigerian Electricity Regulatory Commission, Abuja, Nigeria.

Sambo, AS., Garba, B., Zarma, IH. and Gaji, MM. 2021. Electricity generation and challenges in Nigerian power sector. Journal of Energy Technologies and Policy, 6(2): 1-17.

Song, Y., Liu, D., Yang, C., Hu, K. and Qin, Y. 2023. Remaining useful life prediction method based on a multi-scale convolution and attention mechanism. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 25(2): 13-25.

Stone, GC., Boulter, EA., Culbert, I. and Dhirani, H. 2014. Electrical Insulation for Rotating Machines (2nd ed.). Wiley-IEEE Press, Hoboken, New Jersey.

Susa, D., Lehtonen, M. and Nordman, H. 2005. Dynamic thermal modelling of distribution transformers. IEEE Transactions on Power Delivery, 20(3): 1919-1929.

Tahan, M., Tsoutsanis, E., Muhammad, M. and Karim, ZAA. 2017. Performance-based health monitoring for gas turbines. Applied Energy, 198: 122-144.

Vapnik, V. 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vatsa, P., Kumar, V., Rai, Y., Choubey, A. and Mahela, OP. 2024. Transformer condition monitoring using machine learning and IIoT: A comprehensive review. Measurement: Sensors, 33: 101160.

Wang, Y., Chen, Q., Hong, T. and Kang, C. 2020. Review of smart meter data analytics. IEEE Transactions on Smart Grid, 10(3): 3125-3148.

Corresponding author's email address: durojaiye.koffa@fulokoja.edu.ng

Xing, X., Wang, Y., Zhang, S., Deng, C. and Zhang, X. 2023. A comprehensive predictive maintenance approach for power transformers based on multi-source heterogeneous data integration. IEEE Transactions on Power Delivery, 38(5): 3427-3440.

Zhou, F., Yang, S., Fujita, H., Chen, D. and Wen, C. 2024. Deep learning-based fault diagnosis of power transformers with imbalanced data using synthesized oversampling. IEEE Transactions on Industrial Informatics, 20(2): 2594-2605