Comparative Performance Evaluation of Bayesian-Optimized Random Forest and Conventional Machine Learning Techniques for Predictive Maintenance of Nigerian Aging Transmission Lines

Ganiyu Adedayo Ajenikoko[1], Abeega Andrew Aboityo[2], Abdulbasit Oladayo Ajenikoko[3], Moshood Kola Omotosho[4], Abdul-Muezz Adesina Shittu[5], Israel Ogooluwa Akinpade[6]

[1,2,4,5,6]Department of Electronic and Electrical Engineering
Ladoke Akintola University of Technology
P.M.B, 4000, Ogbomoso, Nigeria.
[3]Department of Electrical and Electronics Engineering
Federal Polytechnic, Ayede, Ogbomoso, Nigeria.
Corresponding Author: ajeedollar@gmail.com
https://doi.org/10.56201/ijcsmt.vol.12.no10.2026.pg154.166

Abstract

Maintaining aging high-voltage transmission assets requires dependable forecasts of failure risk rather than responses based solely on fixed inspection intervals. This paper evaluates Bayesian-optimized Random Forest (BO-RF) alongside conventional Random Forest (RF), gradient boosting, artificial neural network (ANN), support vector machine (SVM) and decision tree classifiers for predicting failures in a subsequent 30-day interval. A reproducible simulation generated 7,200 line-month observations representing 100 hypothetical transmission corridors between 2020 and 2025. Nine condition and exposure variables described asset age, loading, ambient temperature, lightning, corrosion, insulator condition, vegetation clearance, conductor hotspot temperature and preceding faults. The investigation separated the initial training period (2020–2022), Bayesian hyperparameter selection (2023), maintenance-alert threshold determination (2024) and final assessment (2025). Eighteen optimization trials selected RF settings by validation average precision, while alarm thresholds reflected an assumed missed-failure penalty four times the cost of a false alarm. In the 1,200-observation test year, BO-RF attained average precision of 0.316, F1-score of 0.361, recall of 0.465 and ROC-AUC of 0.695. Conventional RF attained average precision of 0.281 and F1-score of 0.335; the ANN achieved the highest ROC-AUC (0.700). At its selected threshold, BO-RF identified 73 failure events, issued 174 false alerts and missed 84 events. Its decision loss was 510 normalized units, compared with 529 for RF. Corridor-level bootstrap analysis supported a positive difference in average precision, although the confidence interval for the reduction in decision loss crossed zero. The investigation shows that evaluation of rare failures depends on temporal validation, false-alarm consequences and probability quality, not classification accuracy alone. Because the observations arise from a controlled simulation rather than utility maintenance records, the findings establish a testable methodology, not field-verified performance in Nigeria. Future applications require timestamped Nigerian transmission-line records, calibrated intervention costs and prospective testing. Attention should also address regional climate differences, measurement quality, and maintenance resource constraints.

Keywords: *Bayesian optimization; Random Forest; predictive maintenance; Nigerian transmission grid; failure prediction; machine learning; reliability assessment.*

1. Introduction

Transmission corridors operate in demanding conditions that expose conductors, towers, insulators and associated fittings to gradual wear as well as sudden disturbances. Thermal cycling, corrosion, lightning, vegetation growth, environmental contamination and sustained loading can progressively weaken asset condition. These influences often reinforce one another: the consequences of high loading, for example, may be more severe where insulation or conductor health has already deteriorated. Maintenance programmes based on fixed calendars cannot fully respond to such interactions. An alternative is to estimate the likelihood of an impending failure and direct scarce inspection capacity towards the circuits where preventive action may yield the greatest benefit.

The case for more systematic asset surveillance is particularly relevant to Nigeria, where grid expansion, asset condition and operating constraints must be managed together. The Nigerian Electricity Regulatory Commission reported three total grid collapses and two partial collapses during the fourth quarter of 2024 (Nigerian Electricity Regulatory Commission [NERC], 2025). Such statistics describe system-wide disturbances and cannot, without a fault investigation, be attributed to deteriorated transmission conductors. The Commission's subsequent annual reporting also highlights the importance of reliable sector-wide performance evidence (NERC, 2026). A predictive maintenance study should therefore rely on time-stamped records of individual line assets rather than assigning transmission-line causation to every major network interruption. Machine-learning models can combine operational measurements and condition indicators even where the underlying relationships are nonlinear. Shakiba et al. (2023) reviewed learning algorithms for transmission-line fault identification, while Anwar et al. (2025) investigated ensemble models for detecting electrical faults. Olojede et al. (2025) examined how learning-based diagnostics might support maintenance activities in power networks. Their work provides an important foundation, but fault identification and predictive maintenance are distinct tasks. The former establishes what has occurred or is occurring; the latter uses information already available to estimate whether a future event is likely within a defined planning period.

Random Forest provides a practical starting point because its ensemble structure accommodates interactions among engineering variables without imposing a single linear risk equation. Nevertheless, results depend on the number and depth of trees, feature sampling, leaf constraints and class weights. Bayesian optimization offers a sequential way to tune these choices from validation feedback rather than examining every possible combination. Ajenikoko et al. (2025) have already investigated Bayesian-enhanced Random Forest in the Nigerian transmission-maintenance setting. The novelty pursued here is therefore a disciplined comparison: identical chronological test observations, clearly separated parameter and threshold selection, multiple evaluation criteria, and uncertainty analysis of any apparent advantage over an untuned RF baseline. The analysis addresses four related objectives: to compare the ability of six classifiers to rank line-months by future failure risk; to examine how their rankings change when precision, recall, calibration and intervention cost are considered; to identify the features most useful in the specified modelling environment; and to set out the verification steps needed for application to Nigerian 132-kV and 330-kV transmission assets. The study is designed as an openly reproducible modelling experiment. It does not claim access to operational event histories from the Transmission Company of Nigeria or imply that a simulated result establishes the actual failure rate of the national grid.

2. Related Works

Transmission-line machine learning has traditionally emphasized detection, classification and location of faults using disturbance measurements. In a broad survey, Shakiba et al. (2023) documented the performance of decision-tree, ensemble, SVM and neural approaches on

these diagnostic problems. Shukla and Deepa (2024) extended comparative analysis to deep-learning architectures for fault classification. Both lines of work demonstrate the value of learned patterns, but the diagnostic labels often arise during or after an electrical event. Forecasting a failure several weeks before it occurs is a different problem: fault events are relatively scarce, exposures vary over time, and only measurements available before the outcome can legitimately enter the model.

Recent work also illustrates how algorithm design influences protection-related decisions. Yousaf et al. (2024) incorporated Bayesian optimization into a high-voltage direct-current fault-detection approach. Anwar et al. (2025) evaluated ensemble fault-classification models, and Li et al. (2025) combined neural representations with tree-based prediction to investigate transmission-line anomalies. These studies establish useful modelling techniques but do not automatically establish performance for monthly condition-based maintenance of aging AC transmission corridors. In particular, unusually high discrimination obtained from post-fault waveforms should not be used as an expectation for forecasts based on routine asset-condition and weather observations. Ajenikoko et al. (2025) directly examined a hybrid Bayesian–Random Forest maintenance framework for the Nigerian setting. Building on that direction, the present research asks whether gains persist under a broader and more transparent benchmark. Its key departures are a shared forward-in-time evaluation period, separate optimization and alert-threshold stages, attention to minority-event measures, and comparison of decisions under an explicit penalty for missed failures. By resampling whole corridors, the uncertainty analysis also acknowledges that repeated monthly measurements from one asset are not statistically independent. The wider predictive-maintenance literature offers further lessons about data quality and deployment. Faisal et al. (2025) surveyed automated power-line inspection and the challenges of combining heterogeneous observations. Gillioz et al. (2025) demonstrated the methodological utility of synthetic electricity-network datasets, while emphasizing that generated scenarios need careful validation before being used to support field claims. Tsallis et al. (2025) reviewed machine-learning maintenance applications and barriers to implementation. In an asset-related application, Bazi et al. (2026) showed why class imbalance and the practical consequences of classification errors should be considered when studying transformer failures. Olojede et al. (2025) likewise drew attention to the connection between prediction and actionable maintenance planning. Three methodological gaps motivate this paper. First, accuracy is frequently presented without enough information about the timing of the forecast, the proportion of failures, or the workload created by false alerts. Second, hyperparameter tuning, alarm calibration and test assessment are not always separated, increasing the risk of optimistic evaluation. Third, modest differences in score are sometimes interpreted as economic improvements without confidence intervals or justified maintenance costs. The adopted benchmark treats these issues as part of model assessment and provides an adaptable protocol for subsequent studies using authenticated Nigerian transmission-line data.

3. Methodology



Figure 1. Workflow for chronological model training, optimization and testing.

Table 1. Chronological allocation of training, validation and test records.

Period	Records	Function
2020–2022	3,600	Initial training
2023	1,200	Hyperparameter tuning
2024	1,200	Alarm threshold validation
2025	1,200	Unseen test

3.1 Prediction task and analysis unit. Each observation describes one transmission corridor at the start of a calendar month, with the target defined as at least one reportable failure during the subsequent 30-day period. The model therefore predicts a future event using only information available beforehand. For this controlled experiment, 100 hypothetical corridors are followed for 72 months from January 2020 to December 2025, producing 7,200 line-month records. Corridor identifiers support grouped uncertainty calculations but do not enter the predictive feature set. The benchmark is an asset-risk classification experiment; it is not a load-flow simulation or a digital reconstruction of the Nigerian 330-kV grid topology.

3.2 Construction of the modelling dataset. Nine predictors are included: line age in years, per-unit loading, ambient temperature, number of lightning exposures, corrosion severity, insulator deterioration, vegetation clearance in metres, conductor hotspot temperature and the count of faults recorded in the previous twelve months. The latter is updated through time to avoid using future outcomes. Variables are generated with bounded distributions, seasonal influences and age-dependent deterioration, using random seed 20261015. All 7,200 records are computer-generated; they do not originate from TCN inspection logs or measured failure databases. Event labels are sampled from a Bernoulli distribution with probability $p = 1/[1 + \exp(-\eta)]$, where η is the specified latent-risk function.

3.3 Event-generation model. Let L denote line loading, C the corrosion index, I the insulator degradation index, K the lightning count, V the vegetation clearance, H the hotspot temperature, F the previous fault count and W the wet-season indicator. A corridor-specific effect u_c captures persistent unobserved variation. The simulated failure propensity is determined from the following log-odds equation:

$$\eta = -5.10 + 2.40 \max(L-0.64,0) + 1.70\,C + 1.70\,I + 0.105\,K + 0.30 \max(3-V,0) + 0.12\,F + 1.90 \max(L-0.80,0)I + 0.025 \max(H-65,0) + u_c + 0.20\,W. \quad (1)$$

The corridor effect u_c follows a zero-mean normal distribution with standard deviation 0.33. This formulation introduces nonlinear load–insulation interactions and gives deteriorated assets a greater probability of experiencing a failure. It is an explicit modelling assumption, not an equation fitted to Nigerian asset records. The generator produces annual event counts of 134, 113, 119, 119, 155 and 157 for 2020 through 2025, respectively. In the final year, the event prevalence is 157/1,200 = 13.08%; this minority-event structure makes a seemingly strong classification accuracy potentially compatible with poor failure detection.

3.4 Forward-in-time evaluation. Data splitting followed calendar order so that later observations could not influence earlier model selection. The initial training period comprises 3,600 line-months from 2020–2022; 1,200 observations from 2023 are used to tune BO-RF hyperparameters. Models then trained on 2020–2023 produce scores for 2024, from which separate alert thresholds are selected. Final versions are fitted on data through 2024, giving 6,000 training observations, and assessed once on the untouched 2025 block of 1,200 observations. Standardization for ANN and SVM is fitted within each training pipeline. This test examines future months on familiar simulated corridors; performance on new geographic corridors remains untested.

3.5 Learning algorithms. The baseline RF uses 240 trees with a minimum of three observations per leaf. BO-RF considers tree number and depth, split and leaf constraints, feature sampling and class weighting. The comparator set includes histogram-based gradient boosting with 160 iterations, an ANN with two hidden layers containing 24 and 12 neurons, an RBF-kernel SVM, and a decision tree with maximum depth seven and minimum leaf size twelve. ANN and SVM use standardized inputs, while ANN training includes early stopping. These fixed baseline settings provide a reproducible demonstration; they should not be described as equally tuned alternatives to BO-RF.

3.6 Ensemble prediction and Bayesian search. Random Forest estimates the probability of a failure by averaging positive-class probabilities from M component trees h_m, as shown in Equation (2). Rather than choose settings manually, Bayesian optimization searches the hyperparameter vector θ to maximize average precision on the 2023 validation observations, according to Equation (3). Eighteen seeded trials use a tree-structured Parzen estimator to guide successive proposals. The selected BO-RF configuration contains 140 trees, maximum depth 14, minimum leaf size 11, minimum split size 14, square-root feature sampling and balanced-subsample class weights. The configuration is specific to the investigated dataset and evaluation criterion.

$$p_{RF}(x) = (1/M)\,\Sigma_m\, h_m(x). \quad (2)$$

$$\theta^* = \arg\max_{\{\theta \in \Theta\}} AP_{validation}(RF_\theta). \quad (3)$$

3.7 Converting risk scores into maintenance alerts. An estimated probability is not by itself a dispatch instruction. For each model, a threshold τ is selected using only 2024 scores by minimizing the decision loss in Equation (4). The penalty is set at four units for each missed failure and one unit for each false alarm; candidate thresholds range from 0.08 to 0.80. These normalized penalties permit a controlled comparison of alert trade-offs. They are not observed Nigerian maintenance expenditures and do not incorporate line criticality, field crew travel, parts replacement, restoration time or cascading outage consequences.

$$L(\tau) = 4\,FN(\tau) + FP(\tau); \quad \tau^* = \arg\min_\tau L_{validation}(\tau). \quad (4)$$

3.8 Evaluation and uncertainty assessment. From true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), accuracy equals (TP+TN)/N, precision equals TP/(TP+FP), recall equals TP/(TP+FN), and F1 equals 2TP/(2TP+FP+FN). ROC-AUC summarizes discrimination across operating points; average precision (AP) summarizes the precision–recall relationship and is particularly informative when failures are uncommon. The Brier score, $B = (1/N)\Sigma_i(p_i - y_i)^2$, evaluates the quality of predicted probabilities. For paired comparison of BO-RF and RF, 500 bootstrap iterations resample complete corridors, retaining the temporal dependence among monthly observations for a given asset.

3.9 Software implementation. The comparative experiment was executed using Python, NumPy, pandas, scikit-learn and Optuna with fixed random seeds. The corresponding design could be implemented in MATLAB R2023a using TreeBagger, fitcensemble, fitcsvm and bayesopt. The source code, generated line-month dataset and plotted results are included in the accompanying reproducibility materials. Importantly, the output values reported below result from this computational experiment and must be validated on operational records before informing Nigerian transmission-maintenance decisions.

4. Results and Discussion

Table 2. Comparative predictive performance on the 2025 simulation test set.

Model	Acc. %	Prec. %	Recall %	F1 %	ROC-AUC	AP	Brier
BO-RF	78.50	29.55	46.50	36.14	0.695	0.316	0.144
RF	80.17	29.85	38.22	33.52	0.685	0.281	0.106

Gradient Boosting	78.67	28.19	40.76	33.33	0.683	0.293	0.106
ANN	84.42	32.56	17.83	23.05	0.700	0.277	0.107
SVM	69.50	21.05	48.41	29.34	0.649	0.207	0.110
Decision Tree	82.67	29.92	24.20	26.76	0.631	0.246	0.115

Table 3. Alert-level confusion matrix values and decision loss on the holdout set.

Model	Threshold	TP	FP	FN	TN	Loss units
BO-RF	0.45	73	174	84	869	510
RF	0.22	60	141	97	902	529
Gradient Boosting	0.17	64	163	93	880	535
ANN	0.37	28	58	129	985	574
SVM	0.15	76	285	81	758	609
Decision Tree	0.34	38	89	119	954	565

4.1 Composition of the test period. The 2025 holdout includes 157 failure-labelled and 1,043 non-failure line-months. Merely classifying every observation as a non-failure would yield 86.92% accuracy, despite identifying none of the events that a preventive maintenance team is trying to avoid. This reference point demonstrates why overall accuracy is not a sufficient indicator of usefulness. All six algorithms were evaluated using the same 1,200 test observations, and no threshold decisions were made using test labels. Table 2 presents the predictive metrics; Table 3 shows the actual alert counts used to assess intervention-related loss.

4.2 Ability to distinguish higher-risk months. BO-RF achieved AP = 0.316 and ROC-AUC = 0.695, compared with AP = 0.281 and ROC-AUC = 0.685 for the conventional RF model. Gradient boosting yielded AP = 0.293 and ROC-AUC = 0.683. Although ANN reached the highest ROC-AUC (0.700), its AP was 0.277; the corresponding values for SVM and the decision tree were 0.207 and 0.246. Figures 2 and 3 illustrate that model rankings depend on the chosen discrimination metric. In a population where failures form a small share of all observations, the precision–recall perspective more directly reflects the value of prioritizing high-risk assets. This is a property of the evaluated simulation rather than an established ranking on Nigerian field data.

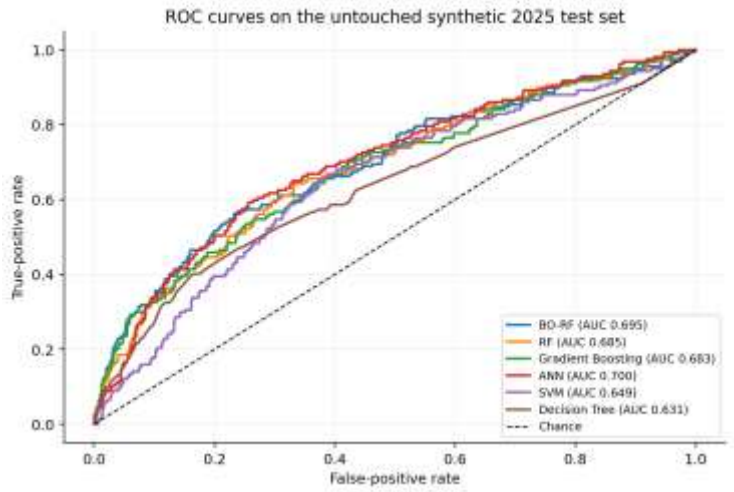


Figure 2. ROC discrimination curves of all six models on the 2025 holdout.

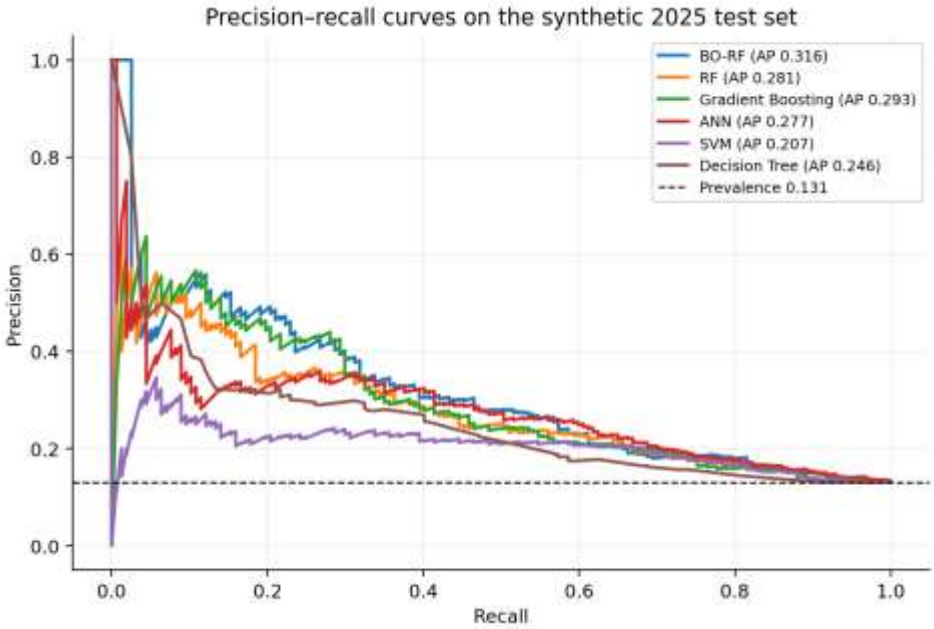


Figure 3. Precision–recall curves; dashed line denotes test-set failure prevalence.

4.3 Performance at selected intervention thresholds. The BO-RF threshold of 0.45 yielded 73 correct failure alerts, 174 false alerts and 84 missed failures, corresponding to precision of 29.55%, recall of 46.50% and F1-score of 36.14%. At its separately selected threshold of 0.22, ordinary RF correctly flagged 60 failures, issued 141 false alerts and missed 97 failures, with F1-score of 33.52%. Thus, BO-RF captured 13 additional failure intervals but also created 33 additional unnecessary alerts. Figure 4 compares precision, recall and F1, while Figure 5 displays the complete BO-RF confusion matrix. The result reinforces a practical point: higher sensitivity is useful only if the organization can manage the associated inspection workload.

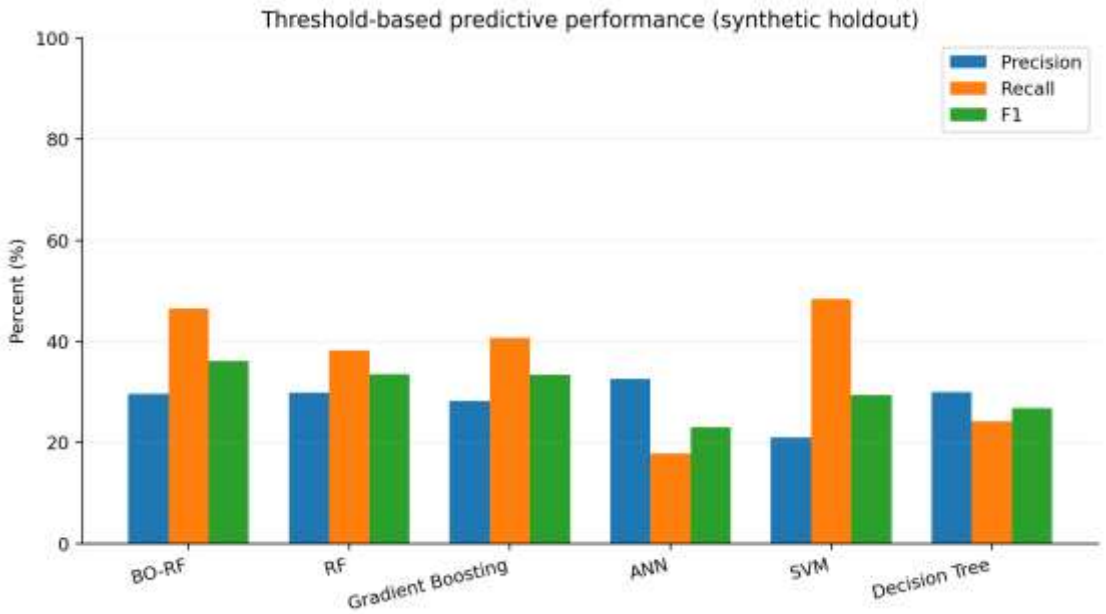


Figure 4. Precision, recall and F1 at validation-chosen alarm thresholds.

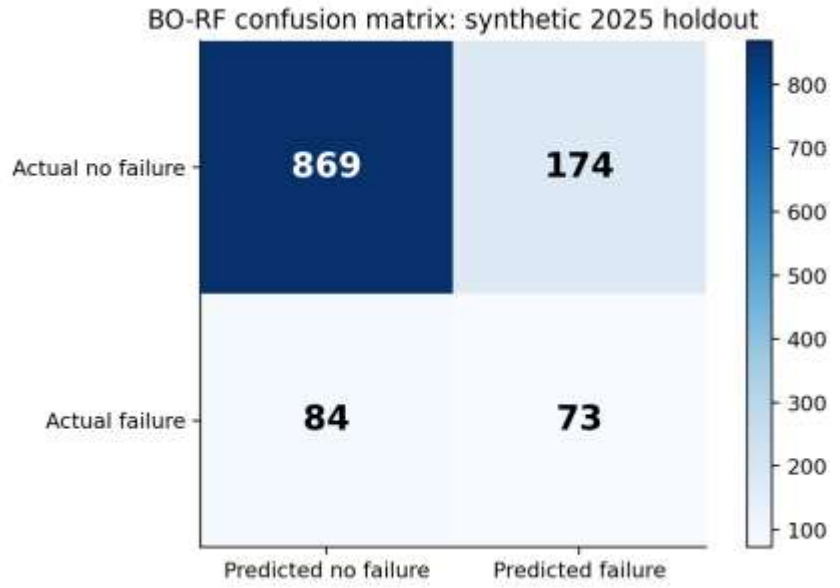


Figure 5. BO-RF holdout confusion matrix (no failure/failure).

4.4 Interpreting conflicting accuracy rankings. The ANN attained 84.42% accuracy and the decision tree 82.67%, both exceeding BO-RF at 78.50%. Yet their low recall meant that ANN detected only 28 failures and the decision tree 38, out of 157. A maintenance team concerned primarily with missed high-consequence events might reject these models despite their higher accuracy. A team operating under severe limits on inspection resources might reach another conclusion. It follows that classifier selection cannot be reduced to a single score; the decision must reflect the operational consequences of both types of error and the capacity to undertake preventive work.

4.5 Comparative maintenance-decision loss. Under the stipulated cost ratio of four units per missed event and one per false alert, total test losses were 510 for BO-RF, 529 for conventional RF, 535 for gradient boosting, 574 for ANN, 609 for SVM and 565 for the decision tree. The BO-RF–RF difference was 19 normalized units, equivalent to a reduction

of 3.59% relative to the RF result. Figure 6 shows the six-model comparison. The values quantify only the decision assumptions in this experiment; they are neither naira-denominated expenditures nor demonstrated financial savings. Any practical costing framework would need records on inspection time, equipment criticality, site accessibility, materials, planned outages and the consequences of failure.

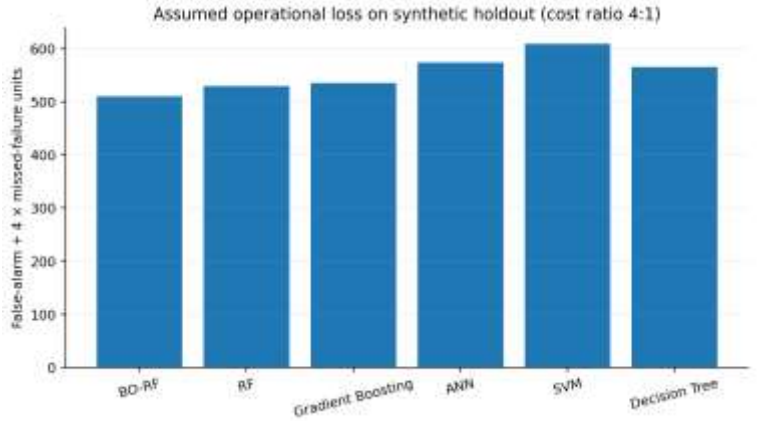


Figure 6. Classifier decision losses under the 4FN + FP criterion.

4.6 Reliability of the predicted probabilities. Notwithstanding its advantage in AP and F1-score, BO-RF had a Brier score of 0.144, worse than the 0.106 obtained by conventional RF and gradient boosting. This outcome distinguishes ranking ability from probability calibration. In particular, class-weighted classifiers can order high-risk observations reasonably well even when their numerical probability outputs do not correspond closely to event frequencies. The selected threshold of 0.45 is therefore a model operating point, not a verified statement that each flagged asset has a 45% chance of failing. A field deployment would need calibration on later measured observations and periodic review as maintenance actions change the observed failure process.

4.7 Engineering interpretation of predictive features. Permutation importance for BO-RF, measured by the reduction in holdout AP, was greatest for vegetation clearance (0.059), line age (0.039), hotspot temperature (0.033), loading (0.028) and insulator degradation (0.026). Corrosion, previous faults and lightning exposure also contributed. Figure 7 illustrates the ordering. Because the generator explicitly embeds several of these relationships, the importance scores indicate what the algorithm learned under the assumed conditions; they do not establish causes of failures on Nigerian lines. Even so, the exercise identifies useful categories of evidence for a future asset database, including clearance surveys, thermographic assessments, insulation inspections and historical fault records.

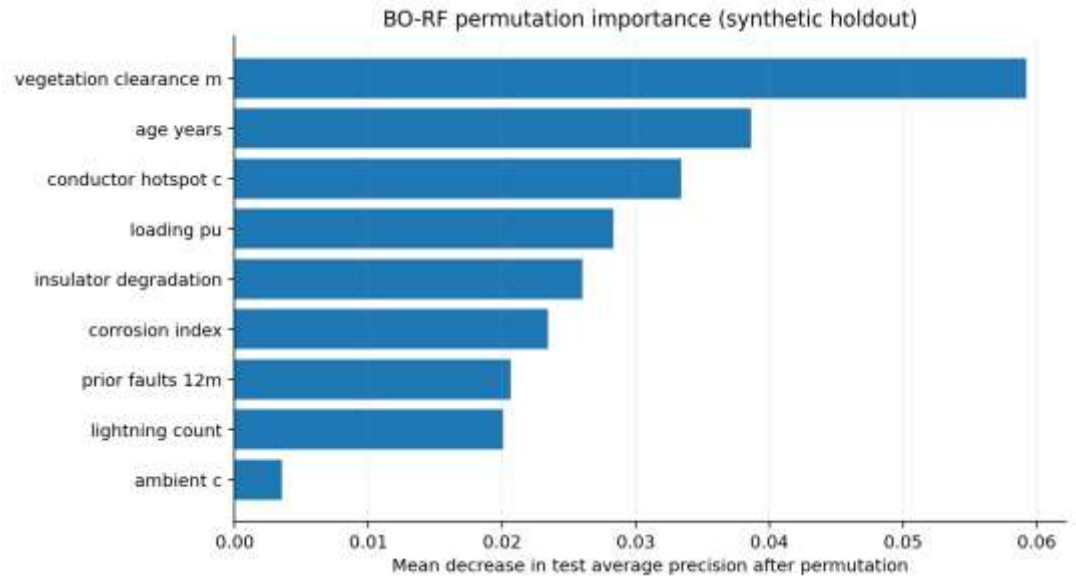


Figure 7. BO-RF permutation importance assessed using holdout average precision.

4.8 Statistical uncertainty in BO-RF improvement. Paired corridor-level resampling produced an AP difference (BO-RF minus RF) of +0.035, with a 95% bootstrap interval of +0.003 to +0.056. This result is consistent with a modest discrimination advantage under the modelled conditions. The corresponding change in maintenance-decision loss was −19 units, but its interval extended from −61.6 to +18.5 units. Because this range includes zero, the experiment does not establish a reliable reduction in expected intervention cost. Table 4 reports both intervals. It would be misleading to convert the point estimate into a confident operational or financial claim, especially given the single data-generation mechanism and assumed loss ratio.

Table 4. Paired corridor-cluster bootstrap uncertainty (500 repetitions).

Bootstrap quantity	Estimate	95% corridor-resample interval
Δ average precision (BO-RF − RF)	+0.035	+0.003 to +0.056
Δ decision loss (BO-RF − RF)	−19.0 units	−61.6 to +18.5 units

4.9 Implications for power-network reliability. Better failure classification does not directly prove higher availability, lower interruption duration or reduced energy not supplied. Those outcomes require evidence that alerts produce successful interventions and that prevented failures affect the network's operating performance. A credible rollout would rank lines by calibrated risk, review recommendations against engineering constraints, schedule inspections or replacements, and monitor actual event rates after action. Its effectiveness should then be evaluated against established maintenance practice using asset failures, intervention effort, restoration time and outage consequences, supported where necessary by a network reliability model.

4.10 Relationship to previous studies. The observed benefit of tuning is consistent with the motivation for Bayesian-enhanced tree ensembles discussed by Ajenikoko et al. (2025), but a direct numerical comparison would be inappropriate because the datasets, targets and

performance definitions are not necessarily the same. The pronounced differences between accuracy and minority-event recall are consistent with the concerns addressed in class-aware maintenance research such as Bazi et al. (2026). Olojede et al. (2025) likewise emphasize the challenge of translating diagnostic predictions into operational interventions. The predictive scores reported here should not be compared uncritically with high classification accuracy for post-fault signals, as investigated by Anwar et al. (2025): identifying an existing disturbance is intrinsically different from anticipating an uncertain event week in advance.

4.11 Limitations and required sensitivity tests. The dataset contains hypothetical corridors, generated weather and condition variables, and event outcomes sampled from one stipulated risk function. Furthermore, comparisons use a specially tuned BO-RF against baseline alternatives that have not received matching optimization budgets. The monthly time resolution, fixed 4:1 decision penalty and reuse of the same corridors in testing restrict generalization. Next-stage experiments should examine changes in event prevalence, missing sensor values, alert costs, class weighting, time-varying model performance and genuinely unseen corridors. Studies using Nigerian operational records should also evaluate 132-kV and 330-kV assets separately if their physical characteristics and incident-reporting practices differ.

5. Conclusions

The study established a transparent framework for comparing Bayesian-optimized Random Forest with conventional RF, gradient boosting, ANN, SVM and decision tree models for advance identification of transmission-line failure risk. The forecast was defined at the start of each month for failure occurrence over the next 30 days. The numerical investigation used 7,200 simulated observations across 100 hypothetical corridors, permitting direct scrutiny of model selection, alarm decisions and error costs without implying that utility records had been collected. Its findings are therefore methodological and computational, rather than field-established conclusions about the Nigerian transmission system.

A principal lesson is the importance of controlling the timing of every modelling decision. Parameter selection used earlier observations, alarm thresholds were chosen in a later validation period, and the final test remained untouched until evaluation. The study also demonstrated the need to combine discrimination, sensitivity, precision, F1-score, calibration and decision loss when failures are less common than normal operation. Finally, resampling whole transmission corridors offered a more suitable uncertainty check than treating successive measurements on the same asset as unrelated observations.

For the stated simulation, BO-RF ranked highest on average precision (0.316), delivered the largest F1-score (0.361) and incurred the smallest loss under the assumed 4:1 penalty (510 units). The ordinary RF counterpart achieved AP = 0.281 and an estimated loss of 529 units. These advantages were not universal: ANN produced the best ROC-AUC, while ordinary RF gave much better-calibrated probabilities. In addition, the bootstrap interval for the cost improvement crossed zero. Accordingly, the results favour further investigation of BO-RF but do not warrant a declaration of guaranteed economic or reliability superiority.

For Nigerian transmission operations, a practical recommendation is to begin with a verifiable line-level database that joins asset age and condition reports to time-stamped load, weather, fault and maintenance histories. Prospective model evaluation should then compare risk-guided inspection with existing schedules, measure false-dispatch burden, and test whether targeted interventions actually reduce future failures and service interruption. Calibrated cost information and line criticality should inform threshold selection. A method should be recommended for operational adoption only after field trials demonstrate stable predictive accuracy and measurable maintenance benefits under real network conditions.

6. Contributions to Knowledge

The research contributes a comparative evaluation protocol rather than claiming the original invention of Bayesian-optimized Random Forest for electrical-asset maintenance. Prior applications, including Ajenikoko et al. (2025), already motivate the use of tuning for this problem. The present work makes the relationship between forecasting, uncertainty and maintenance decision-making more explicit, thereby producing a framework that can be transferred to audited operational datasets. First, the research specifies the forecast unit and timing in engineering terms: at the start of a line-month, information already available is used to predict an event in the following 30 days. This distinction discourages the common substitution of post-fault signal classification for predictive maintenance, and it provides an auditable definition of what an actionable warning should mean. Second, the investigation presents an ordered validation design in which initial training, Bayesian hyperparameter optimization, alert-threshold selection and final testing occupy different chronological periods. The approach reduces look-ahead bias and gives future investigators a clear basis for reproducing or challenging the findings. Accompanying code and the generated dataset permit the sequence to be tested independently and adapted to other asset populations.

Third, the analysis integrates statistical classification scores with a decision-oriented maintenance objective. By examining AP, ROC-AUC, confusion-matrix indicators, Brier score and normalized intervention loss, it demonstrates that no model is necessarily best on every criterion. In particular, the findings make clear why the highest accuracy is not sufficient when most records are non-failures and why improved ranking need not imply well-calibrated failure probabilities. Fourth, the corridor-level bootstrap provides an uncertainty-aware comparison between BO-RF and ordinary RF. The estimated improvement in average precision is positive for the investigated scenario, while the estimated reduction in decision loss remains uncertain. Reporting this distinction prevents small point-estimate differences from being overinterpreted and supports more defensible decisions when models are later assessed on measured utility records. Finally, the study links model inputs to a structured programme of evidence collection. The condition variables and importance analysis suggest measurable factors that utilities could prioritize in inspection records, while avoiding any assertion that simulated importance scores establish actual causal drivers. The combined result is a reproducible, field-validation-ready approach for research on risk-based maintenance of Nigerian high-voltage transmission assets, including future studies that integrate network consequences, verified costs and observed improvements in service reliability.

References

Ajenikoko, G. A., Aboityo, A. A., Ajenikoko, A. O., Omogoye, S., Ayeni, P. J., & Oladele, D. (2025). Development of a hybrid Bayesian-Random Forest framework for transmission line predictive maintenance in Nigeria. International Journal of Engineering and Modern Technology, 11(12), 22–31. https://doi.org/10.56201/ijemt.vol.11.no12.2025.pg22.31

Anwar, T., Mu, C., Yousaf, M. Z., Khan, W., Khalid, S., Hourani, A. O., & Zaitsev, I. (2025). Robust fault detection and classification in power transmission lines via ensemble machine learning models. Scientific Reports, 15, 2549. https://doi.org/10.1038/s41598-025-86554-2

Bazi, S., Nhaila, H., & El Khaili, M. (2026). Data-driven machine learning for enhancing predictive maintenance of distribution transformers. e-Prime – Nexus of Electrical, Electronic, and Intelligent Engineering, 17, 201164. https://doi.org/10.1016/j.eprime.2026.201164

Faisal, M. A. A., Mecheter, I., Qiblawey, Y., Hernandez Fernandez, J., Chowdhury, M. E. H., & Kiranyaz, S. (2025). Deep learning in automated power line inspection: A review. Applied Energy, 385, 125507. https://doi.org/10.1016/j.apenergy.2025.125507

Gillioz, M., Dubuis, G., & Jacquod, P. (2025). A large synthetic dataset for machine learning applications in power transmission grids. Scientific Data, 12, 168. https://doi.org/10.1038/s41597-025-04479-x

Li, Y., Ukwuoma, C. C., & Kemayou, L. N. M. (2025). Enhancing machine-learning-based anomaly detection in transmission lines models using robust features from neural networks. Progress in Engineering Science, 2(4), 100151. https://doi.org/10.1016/j.pes.2025.100151

Nigerian Electricity Regulatory Commission. (2025). Fourth quarter 2024 report. https://nerc.gov.ng/resources/nerc-fourth-quarter-2024-report/

Nigerian Electricity Regulatory Commission. (2026). 2025 annual report and accounts. https://nerc.gov.ng/resources/2025-annual-report-accounts/

Olojede, D., King, S., & Jennions, I. (2025). Application of machine learning in power grid fault detection and maintenance. Energy Informatics, 8, 119. https://doi.org/10.1186/s42162-025-00574-w

Shakiba, F. M., Azizi, S. M., Zhou, M., & Abusorrah, A. (2023). Application of machine learning methods in fault detection and classification of power transmission lines: A survey. Artificial Intelligence Review, 56(7), 5799–5836. https://doi.org/10.1007/s10462-022-10296-0

Shukla, P. K., & Deepa, K. (2024). Deep learning techniques for transmission line fault classification—A comparative study. Ain Shams Engineering Journal, 15(2), 102427. https://doi.org/10.1016/j.asej.2023.102427

Tsallis, C., Papageorgas, P., Piromalis, D., & Munteanu, R. A. (2025). Application-wise review of machine learning-based predictive maintenance: Trends, challenges, and future directions. Applied Sciences, 15(9), 4898. https://doi.org/10.3390/app15094898

Yousaf, M. Z., Singh, A. R., Khalid, S., Bajaj, M., Kumar, B. H., & Zaitsev, I. (2024). Enhancing HVDC transmission line fault detection using disjoint bagging and Bayesian optimization with artificial neural networks and scientometric insights. Scientific Reports, 14, 23610. https://doi.org/10.1038/s41598-024-74300-z