Communication in Physical Sciences, 2026, 13(10): 1557-1575

Implementation of Random Forest and Support Vector Machine Models for Predictive Failure Detection in Power Transmission Equipment

Lambe Mutalub Adesina, Joseph Ibrahim Kuta Joshua*, Ogunbiyi Olalekan, Abdul-Waheed Musa, Monsurat Balogun, and Jimada-Ojuolape Bilkisu
Receieved: 17 June 2026/ Accepted: 16 September 2026/Published: 24 September 2026
https://doi.org/10.4314/cps.v13i10.1

Abstract: *Unplanned failures of power transmission equipment cause costly outages and safety hazards, motivating a shift from reactive and time-based maintenance to predictive maintenance (PdM). This study implements and compares two supervised machine learning models, Random Forest (RF) and a Radial-Basis-Function Support Vector Machine (SVM), to predict the likelihood of failure over a 30-day horizon across five classes of transmission equipment: transformers, circuit breakers, insulators, towers, and conductors. We assembled a condition-monitoring dataset of 6,000 records from four representative Transmission Company of Nigeria (TCN) substations, capturing electrical, thermal, mechanical, dissolved-gas, and environmental parameters. We derived failure labels from a composite risk score grounded in established engineering thresholds (IEEE C57.104/IEC 60599 dissolved-gas limits, partial-discharge severity, aging curves, and thermal loading), yielding a realistic 23.22% failure prevalence (3.31:1 class imbalance). Realistic sensor dropout (2.91% missingness), calibration glitches, and duplicate telemetry were simulated and resolved through a seven-step pre-processing pipeline combining group-wise imputation, missingness-indicator flags, and conservative winsorization. Historical failure count, equipment age, corrosion index, and dissolved-gas concentrations emerged as the dominant predictors, consistent with established transmission asset degradation theory. Both models were implemented in scikit-learn, validated with 5-fold stratified cross-validation, and evaluated on a held-out test set of 1,200 records. Random Forest outperformed SVM on every metric: accuracy (81.3% vs 75.2%), precision (62.7% vs 46.8%), F1-score (54.8% vs 48.4%), ROC-AUC (79.8% vs 72.3%), and average precision (59.1% vs 46.7%). The study concludes that Random Forest is the more suitable of the two models for transmission equipment failure prediction and recommends its deployment as a risk-ranking and maintenance prioritization tool.*

Keywords *: Condition monitoring, Dissolved gas analysis, Machine learning, Power transmission equipment, Predictive maintenance, Random Forest, Support Vector Machine.*

Lambe Mutalub Adesina
Department of Electrical and Computer Engineering, Faculty of Engineering and Technology, Kwara State University, Malete, Nigeria
Email: lambe.adesina@kwasu.edu.ng —

Joseph Ibrahim Kuta Joshua
Department of Electrical and Computer Engineering, Faculty of Engineering and Technology, Kwara State University, Malete, Nigeria
Email: ibrahimkuta74@gmail.com —

Ogunbiyi Olalekan
Department of Electrical and Computer Engineering, Faculty of Engineering and Technology, Kwara State University, Malete, Nigeria
Email: Ogunbiyi.olalekan@kwasu.edu.ng. —

Abdul-Waheed Musa
Department of Electrical and Computer Engineering, Faculty of Engineering and

Technology, Kwara State University, Malete, Nigeria
Email: Waheed.musa@kwasu.edu.ng —

Monsurat Balogun

Department of Electrical and Computer Engineering, Faculty of Engineering and Technology, Kwara State University, Malete, Nigeria
Email: Monsurat.balogun@kwasu.edu.ng —

Jimada-Ojuolape Bilkisu
Department of Electrical and Computer Engineering, Faculty of Engineering and Technology, Kwara State University, Malete, Nigeria
Email: Bilkisu.jimada@kwasu.edu.ng —

1.0 Introduction

The modern electrical power system is the backbone of contemporary economic and social life, and its reliability underpins economic stability, public safety, and national security. Yet transmission infrastructure remains persistently susceptible to degradation and failure. Traditional maintenance strategies, chiefly reactive maintenance (repairing equipment only after it fails) and time-based preventive maintenance (servicing at fixed intervals irrespective of actual condition), are increasingly recognised as inefficient and insufficient for maximising system availability (García *et al*., 2020; Jardine *et al*., 2006; Kumar *et al*., 2018; Davies, 2012; Moleda *et al*., 2023). Reactive maintenance produces costly unplanned downtime and collateral damage, while preventive maintenance wastes resources on unnecessary interventions and can even introduce faults during servicing that would otherwise not have been needed.

These shortcomings have catalysed a shift toward Predictive Maintenance (PdM), a data-driven strategy that uses continuous condition monitoring and analytics to forecast failures before they occur, so that intervention happens only when justified by actual asset condition. PdM is a cornerstone of Industry 4.0 and smart-grid initiatives, aiming to maximise asset availability, extend operational life, optimise spare-parts inventory and improve overall reliability (Lee *at al*., 2025; Folorunso et al., 2025). The volume, velocity and variety of data generated by SCADA systems, phasor measurement units and dissolved-gas analysis (DGA) sensors, however, exceed the analytical reach of conventional statistical methods. Machine learning (ML) offers a practical route to extracting predictive signals from this high-dimensional, non-linear data (García *et al*., 2020).

Despite this potential, effective PdM deployment in power systems is hampered by data that is high-dimensional, noisy, and heavily imbalanced, since failure events are, by nature, rare relative to normal operation. Many classical models struggle to extract reliable patterns under these conditions, producing high false-alarm rates or poor interpretability. At the same time, ageing infrastructure and rising demand are intensifying the strain on transmission networks, particularly in developing country contexts such as Nigeria, where funding constraints and heterogeneous equipment vintages complicate maintenance planning (Ukiwe *et al*., 2023).

In Nigeria, these challenges are particularly important because transmission assets comprise equipment of different ages, manufacturers, operating histories, and maintenance conditions, making uniform time-based maintenance schedules less effective for identifying assets at imminent risk of failure.

Machine learning has become the dominant analytical engine for PdM because it can model the nonlinear relationships between stressors (thermal, electrical, chemical, mechanical, and environmental) and failure outcomes without extensive manual feature engineering. Supervised algorithms most commonly reported for transmission and substation equipment include Random Forest, Support Vector Machines, Naive Bayes, k-Nearest Neighbours, and Neural Networks (Achouch *et al*., 2023; Agrawal and Patel, 2021; Sun *et al*., 2024; Idowu

et al., 2025). Random Forest, an ensemble of decision trees trained on bootstrap samples with random feature subsets, is favoured for its robustness to noise and missing values, its ability to model feature interactions, and the interpretable feature-importance rankings it produces (Rokach and Maimon, 2005). Support Vector Machines, particularly with an RBF kernel, are valued for constructing maximum-margin decision boundaries on moderate-sized, high-dimensional tabular data (Cervantes *et al*., 2023). The choice of an appropriate learning algorithm is therefore important because transmission-equipment failure prediction involves heterogeneous features, nonlinear relationships, missing observations, and an inherently imbalanced distribution between failure and non-failure events.

A substantial body of research has applied dissolved gas analysis and infrared thermography to transformer condition monitoring, while other studies have investigated machine-learning approaches for fault detection and remaining-useful-life prediction in individual asset classes such as transformers, wind turbines, and industrial equipment(Qin *et al*., 2017; Achouch *et al*., 2023; Ouadah *et al*,, 2022). . However, much of the existing literature treats individual equipment categories independently and relies on asset-specific diagnostic indicators. Consequently, there remains limited evidence on the comparative performance of general-purpose classification models when heterogeneous electrical, thermal, mechanical, chemical, and environmental indicators are jointly used to predict failures across multiple transmission-equipment classes.(Qin *et al*., 2017; Achouch *et al*., 2023; Ouadah *et al*,, 2022). Against this background and in response to the identified research gap, this study develops and evaluates two supervised ML algorithms, Random Forest and an RBF-kernel Support Vector Machine, for predicting transmission-equipment failure within a 30-day horizon The specific objectives are to: assemble and analyse a realistic, multi-parameter condition monitoring dataset spanning five equipment classes; implement and validate both models on historical and simulated real-time data; and evaluate and compare their predictive performance using metrics appropriate to an imbalanced classification problem, to recommend a model suitable for deployment as a maintenance prioritisation tool. The significance of the study lies in its potential to support condition-based maintenance planning by identifying high-risk transmission assets before failure occurs. By integrating multiple condition indicators across heterogeneous equipment classes, the proposed approach provides a basis for risk ranking and maintenance prioritisation, which may assist transmission-system operators in allocating limited maintenance resources, reducing unplanned outages, and improving asset reliability.

1.1 Literature Review

Maintenance strategies for power system assets are generally classified into reactive, preventive, and predictive approaches. Reactive maintenance runs equipment to failure, minimising servicing costs at the expense of frequent and potentially expensive downtime, whereas preventive maintenance services equipment at fixed intervals, thereby avoiding some failures but potentially replacing components before the end of their useful life. Predictive maintenance, in contrast, uses condition data, sensor readings, historical records, and environmental information to time interventions according to actual asset health, combining some of the cost advantages of reactive maintenance with the reliability benefits of preventive maintenance (Moleda et al., 2023; Iduwu et al., 2026). Recent advances in machine learning have further strengthened predictive maintenance by enabling the identification of nonlinear relationships between equipment condition indicators and subsequent failure events. However, the effectiveness of these approaches depends strongly on the quality, diversity, and representativeness of the data used for model development and validation. This section

critically reviews previous applications of machine-learning methods to power-system asset diagnosis and predictive maintenance, with particular attention to the equipment classes, input variables, modelling approaches, validation strategies, and performance reported in studies relevant to the present investigation.

Studies have demonstrated the application of Random Forest (RF) and Support Vector Machine (SVM) algorithms to dissolved gas analysis (DGA) signals for diagnosing faults in oil-immersed transformers**.** Adu-Gyamfi (2026) benchmarked Random Forest against k-Nearest Neighbors (KNN), SVM, and an Artificial Neural Network (ANN) using 589 field-verified DGA records spanning six IEC 60599 fault classes. Using identical stratified splits and grid-search tuning to facilitate a fair comparison, Random Forest achieved the highest held-out accuracy of 83.76%, compared with 76.07% for both KNN and SVM and 70.94% for ANN. However, the substantial reduction in Random Forest accuracy to 66.48% and 67.12% when evaluated on two independent externally sourced DGA datasets demonstrates that strong performance on a single dataset does not necessarily translate into equivalent performance on unseen data. This finding highlights the importance of external validation and dataset diversity when assessing the generalisability of machine-learning models for equipment-failure prediction.

Ensemble- and resampling-based refinements of Random Forest have also produced high classification accuracies in transformer DGA studies**.** Guan *et al.* (2023) combined Tomek-link-based data cleaning, Adaptive Synthetic Sampling (ADASYN), and Snake-Optimisation-tuned Random Forest (SO-RF) to achieve 97.06% diagnostic accuracy on a public DGA dataset, demonstrating the potential influence of data preprocessing and class-imbalance treatment on model performance. Balabantaraya *et al.* (2026) likewise reported that Random Forest and Gradient Boosting achieved accuracies of up to 98% on real service-transformer DGA datasets, compared with 81.4% for SVM and 76% for ANN baselines reported from previous studies. Although these results demonstrate the strong diagnostic capability of ensemble learning, the studies remain largely focused on transformer fault classification and DGA-specific indicators. Consequently, their findings provide limited evidence concerning whether the same algorithms can maintain reliable performance when electrical, thermal, mechanical, chemical, and environmental variables are jointly used across different types of transmission equipment.

Transmission-line and insulator fault studies typically employ electrical features, such as voltage and current waveforms, or image-derived features rather than chemical indicators. Several studies have reported high predictive accuracies using these approaches. Anwar et al. (2025) evaluated Random Forest, KNN, and LSTM models using transmission-line phase-voltage and current patterns and proposed an RF-LSTM-Tuned-KNN ensemble that achieved 99.96% accuracy on a multi-label fault dataset, compared with 97.50% for standalone Random Forest and 96.55% for KNN. For binary fault classification, KNN and Random Forest achieved 99.85% and 99.72%, respectively. Najafzadeh et al. (2024) combined fuzzy thresholding with optimised machine-learning classifiers for fault detection, classification, and localisation along a power-grid line, reporting 98.1% accuracy for a decision tree and 100% for Random Forest, with a localisation error below 154 m. These high accuracy values should, however, be interpreted in relation to the nature of the datasets, because simulation-derived or highly separable fault datasets may not reproduce the noise, missing observations, measurement uncertainty, and class imbalance encountered in operational condition-monitoring data. In a more recent comparative study involving ten machine-learning and deep-learning algorithms for transmission-line fault-type classification and fault-location prediction, Random Forest produced an accuracy of 93.20%, compared with

92.45% for decision tree, 86.58% for CatBoost, and 86.40% for XGBoost. Random-Forest-based approaches to insulator condition assessment, including prediction of lightning-flashover criteria on 110 kV insulators and transfer-learning-based visual defect detection, further extend machine-learning applications to transmission-asset condition assessment.

Despite these advances, the existing evidence remains fragmented across individual equipment types, diagnostic modalities, and datasets. Transformer studies have predominantly concentrated on DGA-based diagnosis, while transmission-line and insulator studies have often relied on electrical waveforms, simulated fault conditions, or image-based features. Comparatively less attention has been given to a unified predictive-failure framework that accommodates heterogeneous transmission assets and combines multiple condition-monitoring domains within a common modelling and validation procedure.

This limitation is particularly relevant because practical transmission networks comprise interconnected assets with different degradation mechanisms and observable condition indicators, including transformers, circuit breakers, insulators, towers, and conductors. A model developed exclusively for transformer DGA or transmission-line electrical faults may therefore not adequately represent the broader failure-risk profile of a heterogeneous transmission network. Furthermore, high classification accuracy reported on individual datasets does not by itself establish generalisability, particularly where failure events are relatively rare and the underlying observations contain missing values, outliers, measurement errors, and temporal or regional variation.

Accordingly, there remains a need for comparative evaluation of machine-learning algorithms using a common analytical framework that incorporates electrical, thermal, mechanical, chemical, environmental, and asset-history variables across multiple transmission-equipment classes. The present study addresses this gap by considering transformers, circuit breakers, insulators, towers, and conductors within a unified predictive-failure framework and by directly comparing Random Forest with an RBF-kernel Support Vector Machine under the same preprocessing, data-splitting, validation, and performance-evaluation conditions. This approach is intended to provide a more comprehensive assessment of model suitability for transmission-equipment risk ranking and maintenance prioritisation than studies restricted to a single asset type or diagnostic modality.

2.0 Materials and Methods

This section describes the study area, dataset construction, monitoring variables, data preprocessing procedures, and machine-learning implementation used to develop and compare Random Forest (RF) and Support Vector Machine (SVM) models for predictive failure detection in power transmission equipment. The methodology comprises dataset preparation, data-quality assessment, preprocessing, feature transformation, stratified data partitioning, model development, validation, and performance evaluation.

2.1 Study Area

This study was based on the Transmission Company of Nigeria (TCN) network, using four representative substations to provide regional coverage of the northern, eastern, southern, and western parts of the network. The selected substations were Kaduna Substation (North, 330 kV), Onitsha Substation (East, 330 kV), Benin Substation (South, 330 kV), and Osogbo Substation (West, 330 kV). These substations were used to provide regional labels and contextual variation in the dataset rather than to imply complete geographical representation of the TCN network. The monitored variables were organised into electrical, thermal, mechanical, chemical, environmental, and asset-history domains to capture the principal factors associated with transmission-equipment condition and failure risk. Table 1 summarises

the monitoring domains and their representative variables.

Table 1:Monitoring feature domains and representative variables

Domain	Representative Features	Count
Electrical	Voltage, load current, load percentage, partial-discharge magnitude	4
Thermal	Surface temperature, ambient temperature	2
Mechanical	Vibration, corrosion index, oil level	3
Chemical	Dissolved H_2, CH_4, C_2H_2, and SF_6 pressure where applicable	4
Environmental	Humidity, wind speed, rainfall, lightning strikes (30-day)	4
Asset history/context	Age, maintenance recency, prior failure count, equipment type, region, season	6

Note: SF_6 pressure is treated as an equipment-specific gas-related condition indicator and applies only to equipment for which SF_6 is relevant; it should not be interpreted as a conventional transformer DGA variable.

2.2 Dataset Description

The complete computational implementation, including data preprocessing, model training, validation, and evaluation, was carried out in Python. A condition-monitoring dataset comprising 6,000 records was assembled to represent five transmission-equipment classes: transformers, circuit breakers, insulators, towers, and conductors. The dataset covered an 18-month monitoring period from January 2023 to June 2024 and incorporated four seasonal conditions (summer, winter, monsoon, and spring) to represent regional and seasonal variations in operating conditions.

Each record contained electrical, thermal, mechanical, chemical, and environmental condition indicators together with asset-history information, including equipment age, maintenance recency, and prior failure count. The response variable was a binary failure indicator representing whether the equipment was classified as likely to fail within the subsequent 30-day prediction horizon.

The dataset was computationally constructed to represent realistic transmission-equipment condition-monitoring observations, with engineering-based operating ranges and failure-risk relationships used to generate the observations. Controlled data-quality imperfections, including sensor missingness, calibration irregularities, and duplicate telemetry, were subsequently introduced to reproduce realistic challenges encountered in condition-monitoring datasets.

The number of observations assigned to each equipment class and the resulting class distribution should be reported explicitly to permit assessment of equipment-level representation and reproducibility of the modelling procedure.

The computational workflow used Python libraries appropriate to each analytical stage. Pandas and NumPy were used for data handling, cleaning, transformation, and numerical operations. Scikit-learn was used for preprocessing, model development, data partitioning, cross-validation, and performance evaluation. The RandomForestClassifier was used to implement the RF model, while SVC with an RBF kernel was used for the SVM model. StandardScaler and OneHotEncoder were used for numerical standardisation and categorical encoding, respectively. StratifiedKFold and cross-validation procedures were used to maintain the class distribution during model

validation. Performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, confusion matrices, and precision-recall analysis. Joblib was used to serialise trained model pipelines, while Matplotlib and Seaborn were used to generate the graphical outputs.

2.0 Materials and Methods

This section describes the study area, dataset construction, monitoring variables, data preprocessing procedures, and machine-learning implementation used to develop and compare Random Forest (RF) and Support Vector Machine (SVM) models for predictive failure detection in power transmission equipment. The methodology comprises dataset preparation, data-quality assessment, preprocessing, feature transformation, stratified data partitioning, model development, validation, and performance evaluation.

2.1 Study Area

This study was based on the Transmission Company of Nigeria (TCN) network, using four representative substations to provide regional coverage of the northern, eastern, southern, and western parts of the network. The selected substations were Kaduna Substation (North, 330 kV), Onitsha Substation (East, 330 kV), Benin Substation (South, 330 kV), and Osogbo Substation (West, 330 kV). These substations were used to provide regional labels and contextual variation in the dataset rather than to imply complete geographical representation of the TCN network. The monitored variables were organised into electrical, thermal, mechanical, chemical, environmental, and asset-history domains to capture the principal factors associated with transmission-equipment condition and failure risk. Table 1 summarises the monitoring domains and their representative variables.

2.2 Dataset Description

The complete computational implementation, including data preprocessing, model training, validation, and evaluation, was carried out in Python. A condition-monitoring dataset comprising 6,000 records was assembled to represent five transmission-equipment classes: transformers, circuit breakers, insulators, towers, and conductors. The dataset covered an 18-month monitoring period from January 2023 to June 2024 and incorporated four seasonal conditions (summer, winter, monsoon, and spring) to represent regional and seasonal variations in operating conditions.

Table 1. Monitoring feature domains and representative variables

Domain	Representative Features	Count
Electrical	Voltage, load current, load percentage, partial-discharge magnitude	4
Thermal	Surface temperature, ambient temperature	2
Mechanical	Vibration, corrosion index, oil level	3
Chemical	Dissolved H_2, CH_4, C_2H_2, and SF_6 pressure where applicable	4
Environmental	Humidity, wind speed, rainfall, lightning strikes (30-day)	4
Asset history/context	Age, maintenance recency, prior failure count, equipment type, region, season	6

Note: SF_6 pressure is treated as an equipment-specific gas-related condition indicator and is applicable only to equipment for which SF_6 is relevant; it should not be interpreted as a conventional transformer DGA variable.

Each record contained electrical, thermal, mechanical, chemical, and environmental condition indicators together with asset-history information, including equipment age, maintenance recency, and prior failure count. The response variable was a binary failure indicator representing whether the equipment was classified as likely to fail within the subsequent 30-day prediction horizon.

The dataset was computationally constructed to represent realistic transmission-equipment condition-monitoring observations, with engineering-based operating ranges and failure-risk relationships used to generate the observations. Controlled data-quality imperfections, including sensor missingness, calibration irregularities, and duplicate telemetry, were subsequently introduced to reproduce realistic challenges encountered in condition-monitoring datasets.

The number of observations assigned to each equipment class and the resulting class distribution should be reported explicitly to permit assessment of equipment-level representation and reproducibility of the modelling procedure.

The computational workflow used Python libraries appropriate to each analytical stage. Pandas and NumPy were used for data handling, cleaning, transformation, and numerical operations. Scikit-learn was used for preprocessing, model development, data partitioning, cross-validation, and performance evaluation. The RandomForestClassifier was used to implement the RF model, while SVC with an RBF kernel was used for the SVM model. StandardScaler and OneHotEncoder were used for numerical standardisation and categorical encoding, respectively. StratifiedKFold and cross-validation procedures were used to maintain the class distribution during model validation. Performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, confusion matrices, and precision-recall analysis. Joblib was used to serialise trained model pipelines, while Matplotlib and Seaborn were used to generate the graphical outputs.

2.3 Data Pre-processing

The binary failure label was not assigned arbitrarily. It was derived from a composite latent risk score built from condition-monitoring indicators with documented links to transmission-equipment failure, including IEEE C57.104 / IEC 60599 dissolved-gas thresholds, partial-discharge severity, ageing curves, and thermal loading, passed through a logistic function with stochastic noise. This produced a realistic 23.22% failure prevalence (3.31:1 class imbalance), avoiding the unrealistically balanced classes common in synthetic PdM datasets.

Realistic field imperfections, sensor dropout (2.91% missingness across numeric fields, ranging from 0.33% for equipment age to 8.47% for partial-discharge magnitude), calibration glitches and duplicate telemetry, were deliberately simulated and then resolved through a seven-step pipeline comprising duplicate removal, group-wise (equipment-type-conditioned) imputation, creation of twelve missingness-indicator flags, and conservative outlier winsorization at 3×IQR rather than the conventional 1.5×IQR, a threshold better suited to the skewed distributions typical of engineering monitoring variables. Non-predictive identifiers (equipment ID, timestamp) were dropped, with a derived 'days-since-epoch' feature retained to preserve temporal trend information. Numeric features, including the missingness indicator flags, were standardised to zero mean and unit variance, and categorical features (equipment type, region, season) were one-hot encoded. The resulting 33-variable, model-ready dataset was split 80:20 into training (4,800 records) and held-out test (1,200 records) sets using stratified sampling to preserve the 23.22% failure ratio in both partitions.

2.4 Mathematical formulation for pre-processing

Formula applied during the seven-step pre-processing pipeline are:

Missingness Rate (Missingness Audit)

$$Missing\ Rate = (\frac{\mathrm{n}_{missing}}{N}) \times 100\% \quad (1)$$

Domain-Range Validity Check (Glitch Detection)

$$\begin{cases} x_{flagged} = missing, & if\ x < lo\ or\ x > hi \\ x_{flagged} = x, & otherwise \end{cases} \quad (2)$$

Where [lo, hi] is the physically valid range for a given sensor field (e.g. humidity pct ∈ [0, 100];

corrosion index ∈ [0, 10]; vibration mms ∈ [0, 20]). Values outside this range are converted to missing before imputation rather than used directly.

Group-Wise Median Imputation

$$\hat{x}_i = \text{median}(\, x_j \mid \text{equipment\_type}_j = t,\ \ x_j \ observed\,) \quad (3)$$

For a missing value in feature x belonging to equipment type t, the imputed value $\hat{x}_i$ is the median of all observed values of that feature within the same equipment type group. The median (rather than the mean) is used because sensor distributions are right-skewed and the median is robust to residual outliers at this stage of the pipeline.

Mode Imputation (Categorical Fields)

$$\hat{c}_i = \text{mode}(\, \text{equipment\_type}, \text{region}, \text{season}\,) \quad (4)$$

The most frequently occurring category fills any missing categorical value.

Interquartile Range (Winsorization)

$$\text{IQR} = Q_3 - Q_1 (5)$$

Where Q_1 is the 25th percentile and Q_3 is the 75th percentile of a given numeric feature.

Outlier Fences and Winsorization

$$\text{Lower fence} = Q_1 - k \times IQR \qquad \text{Upper fence} = Q_3 + k \times IQR \quad (6)$$

$$x_{winsorized} = \min(\, \max(\, x, Lower\ fence\,), Upper\ fence\,)$$

With k = 3 (a conservative choice versus the standard k = 1.5), since several monitoring variables (dissolved gases, vibration, partial discharge) are naturally right-skewed by engineering process and a tighter fence would misclassify legitimate high-risk readings as noise.

Z-Score Standardisation (Model-Ready Transform)

$$z = \frac{x - \mu}{\sigma} \quad (7)$$

where μ and σ are the training-set mean and standard deviation of each numeric feature, fitted only on the training split and then applied to the test split (avoiding data leakage). Standardisation is required for SVM's distance-based margin optimisation and improves numerical stability for Random Forest splits.

One-Hot Encoding (Categorical Transform)

$$x_{encoded} = [\, \mathbb{1}(x = c_1), \mathbb{1}(x = c_2), \ldots, 1(x = c_k)\,] \quad (8)$$

Where 1 is the indicator function (1 if true, 0 otherwise) and $\{c_1, \ldots, c_k\}$ are the categories of a given categorical feature. Applied to equipment type, region and season.

Missingness Indicator Flag

$$\text{flag}_i = 1(\, x_i \text{ was originally missing}\,) \quad (9)$$

2.5 Random Forest Classifier Implementation

A Random Forest ensemble of 400 decision trees (maximum depth 12, minimum leaf size 3, balanced class weighting) was implemented in scikit-learn. Each tree is trained on a bootstrap sample of the training data, and at every split only a random subset of features (approximately √p of the 33 available) is considered, decorrelating the trees so that averaging their votes cancels out the tendency of any single tree to overfit. Each split is chosen to minimise Gini impurity,

$$\text{Gini} = 1 - \Sigma\, p_i^2 \quad (10)$$

where p_i is the class proportion at a node; the tree splits until nodes are pure or the depth/leaf-size stopping criteria are reached. Random Forest was selected for its ability to model non-linear interactions between electrical, thermal and chemical stressors without explicit feature-interaction engineering, and for the interpretable feature-importance rankings it yields for maintenance engineers.

2.5 Support Vector Machine Implementation

A Support Vector Machine with a Radial Basis Function (RBF) kernel (C = 5.0, gamma = 'scale', balanced class weighting) was implemented as the comparator model. SVM seeks the widest possible margin between classes, defined by a hyperplane $w \cdot x + b = 0$. For non-separable data, a soft margin is used, introducing slack variables ξ_i penalised by cost parameter C:

$$\text{minimise } \tfrac{1}{2}\|w\|^2 + C \cdot \Sigma \xi_i \quad (11)$$

Because the failure/no-failure boundary in this dataset is non-linear (risk rises sharply beyond certain age or dissolved-gas thresholds), the RBF kernel,

$$K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2) \quad (12)$$

was used to compute similarity as though points had been projected into a higher-dimensional space, without the computational cost of the explicit transformation. Because SVM decision functions produce a signed distance rather than a probability, Platt scaling (an internal logistic regression fitted to the decision-function output) was used to obtain the calibrated probabilities needed for ROC and precision-recall analysis.

The complete workflow used for data preprocessing, model development, validation, evaluation, and maintenance-risk prioritisation is shown in Fig. 1.

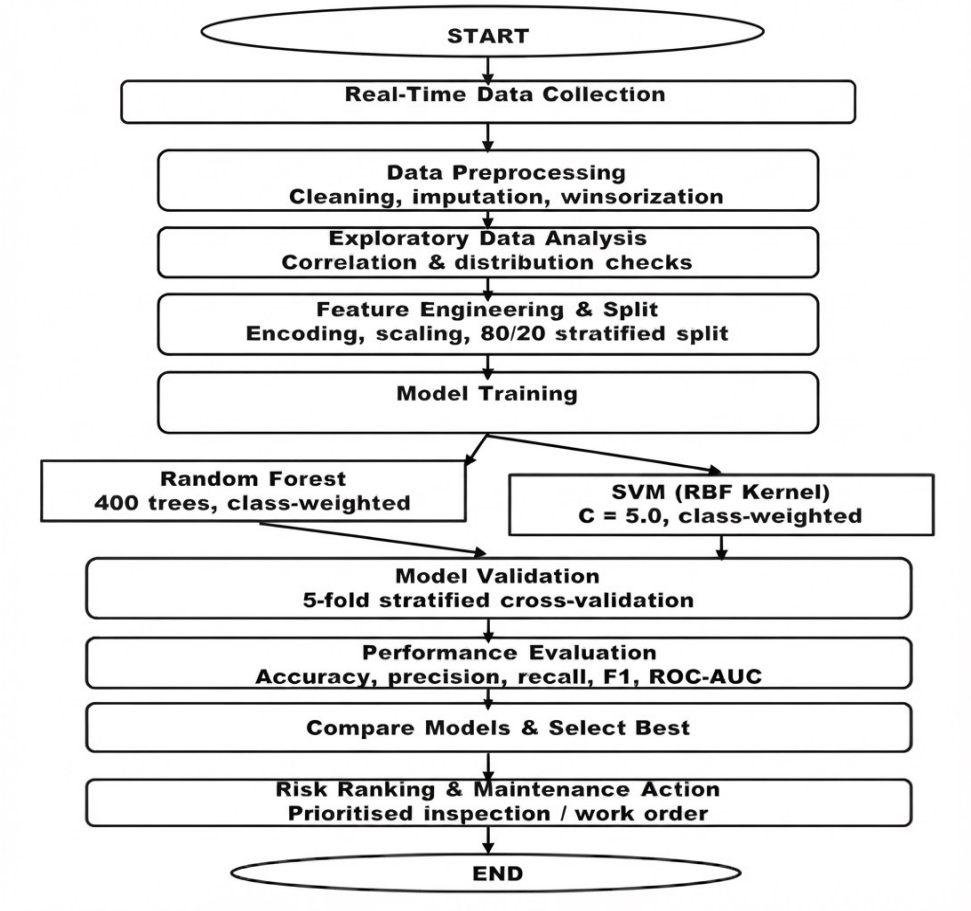

Fig. 1: Implementation Flowchart

2.6 Performance Evaluation Metrics

This study establishes the use of performance evaluation metrics such as accuracy, precision, recall, and F1-score to evaluate the efficiency of models. The formulas of evaluation are briefly defined as follows:

Accuracy is a performance metric that quantifies the overall effectiveness of a classification model.

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

Precision serves as a measure of the model's exactness in predicting the positive class.

$$\textbf{Precision} = \frac{TP}{TP+FP} \quad (14)$$

Recall quantifies the proportion of actual positive samples that are correctly classified as positive

$$\textbf{Recall (Sensitivity)} = \frac{TP}{TP+FN} \quad (15)$$

The F1-score, which harmonically averages precision and recall, proves to be an effective performance measure for classification models.

$$\textbf{F1-Score} = 2\,\frac{Precision\ x\ Recall}{Precision+Recall} \quad 16)$$

$$\textbf{False Positive Rate (FPR)} = \frac{FP}{FP+TN} \quad (17)$$

$$\textbf{False Negative Rate (FNR)} = \frac{FN}{TP+FN} \quad (18)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

3.0 Results and Discussion

3.1 Exploratory Data Analy

The exploratory data analysis was conducted to assess the quality, distribution, and characteristics of the condition-monitoring dataset before model development. Particular attention was given to missing observations, equipment-specific failure patterns, relationships among monitoring variables, and variations in failure rates across seasons and geographical regions.

Fig. 2 presents the missing-value (sensor-dropout) rate for the monitoring features in the raw dataset. The overall sensor-dropout rate across the numerical fields was 2.91%, corresponding to 3,662 missing cells out of 126,000 numerical observations. This level of missingness indicates that the dataset contained measurable but manageable data-quality imperfections, providing a realistic basis for evaluating the preprocessing and imputation procedures described in Section 2.3. The variation in missingness among individual monitoring variables also demonstrates that sensor reliability was not uniform across the monitored features. Following the preprocessing procedure, the missing observations were treated using the specified imputation strategy and missingness indicators.

sis

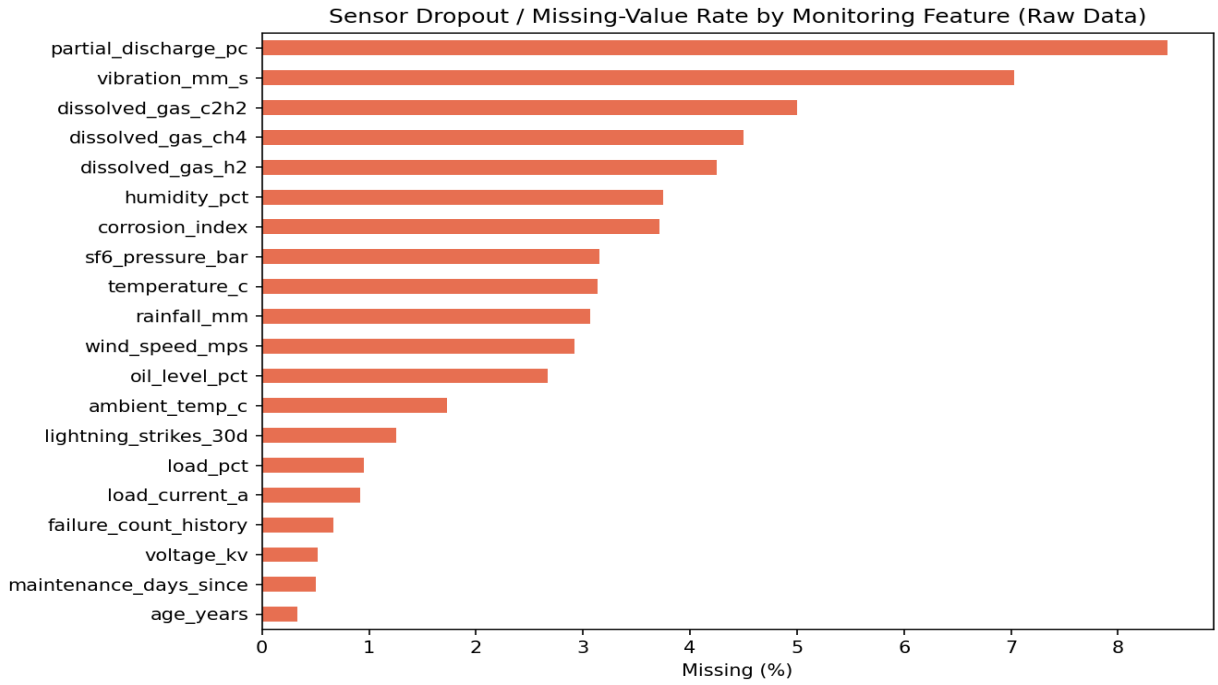

Fig. 2: Missing-Value (sensor dropout) Rate by Monitoring Feature in the Raw data

Fig. 3 shows the failure rate across the five equipment classes considered in the study. Transformers exhibited the highest failure rate and accounted for approximately 28% of all recorded failures, followed by conductors and circuit breakers. The observed differences among equipment classes indicate that failure risk was not uniformly distributed across the transmission assets and therefore support the need to include equipment type as an explanatory variable in the predictive models. The result also demonstrates the importance of evaluating a predictive-maintenance framework across heterogeneous equipment rather than restricting the analysis to a single asset category.

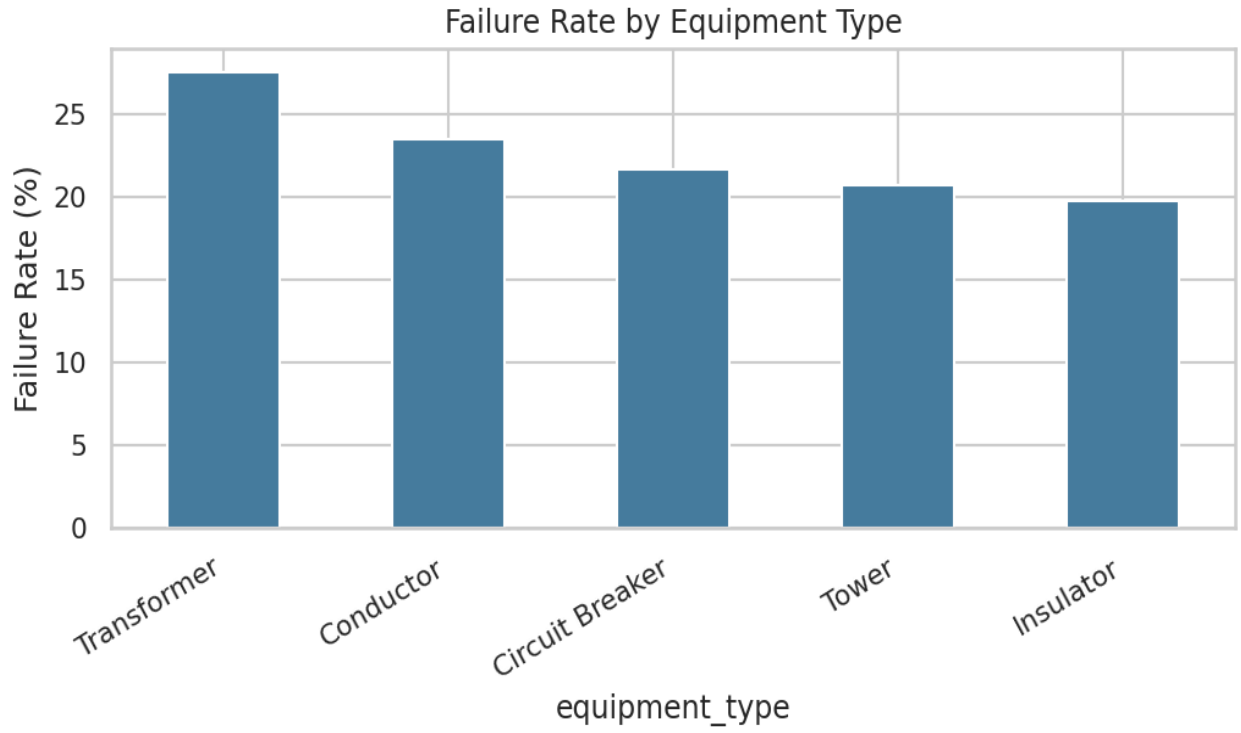

Fig. 3: Failure Rate by Equipment Type

Fig. 4 presents the correlation matrix of the monitoring features. Partial-discharge magnitude, dissolved acetylene (C_2H_2) concentration, and surface temperature exhibited some of the strongest bivariate associations with the failure outcome. These relationships are physically plausible because partial discharge can indicate insulation deterioration, elevated dissolved-gas concentrations can be associated with abnormal electrical or thermal activity, and increased surface temperature can reflect excessive loading or deteriorating equipment condition. However, correlation represents a bivariate association and should not be interpreted as evidence that any single variable independently causes equipment failure. The correlation analysis therefore provided an exploratory basis for subsequent multivariable machine-learning modelling rather than serving as the final feature-selection procedure.

Fig. 5 illustrates the variation in failure rate across seasons and geographical regions. The monsoon period (June–September) recorded the highest average failure rate at 24.8%, with the East region reaching a peak of approximately 29%. The observed pattern is consistent with the possibility that rainfall, humidity, and associated environmental stresses may contribute to adverse equipment operating conditions. However, these observations represent associations within the

study dataset and should not be interpreted as establishing rainfall, humidity, or flooding as independent causal determinants of failure without further statistical or field investigation. The October–November period recorded the lowest observed failure rate of 21.8%. Rather than describing this period as the "safest" maintenance window, it is more appropriate to state that it exhibited the lowest observed failure rate within the analysed dataset; operational maintenance scheduling would require consideration of additional system constraints and field conditions.

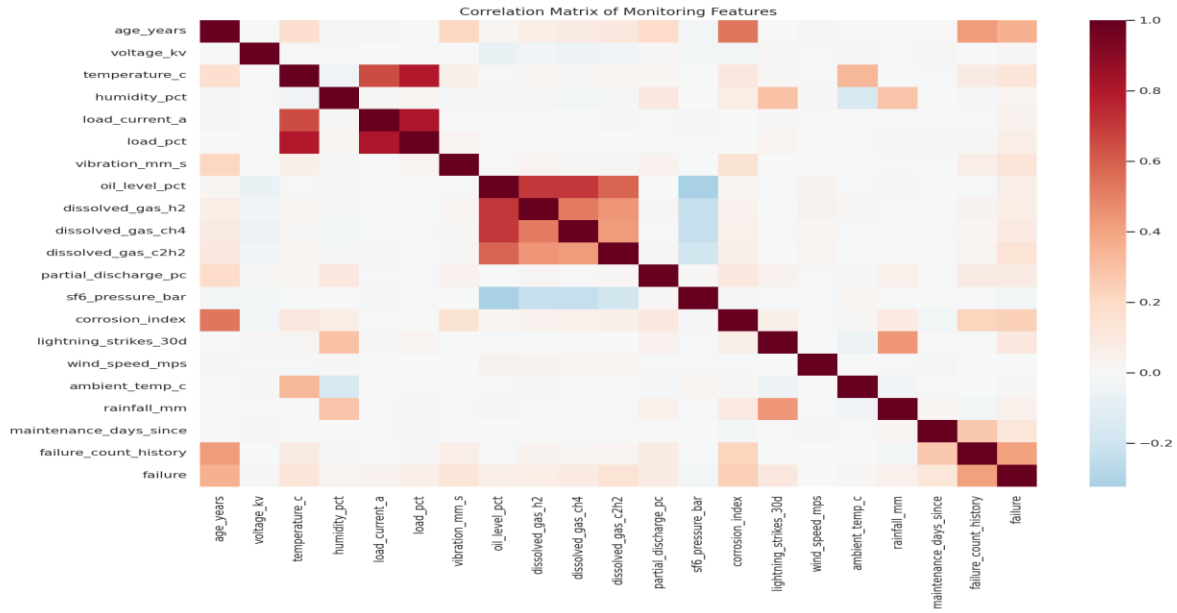

Fig. 4: Correlation Matrix of Monitoring Features

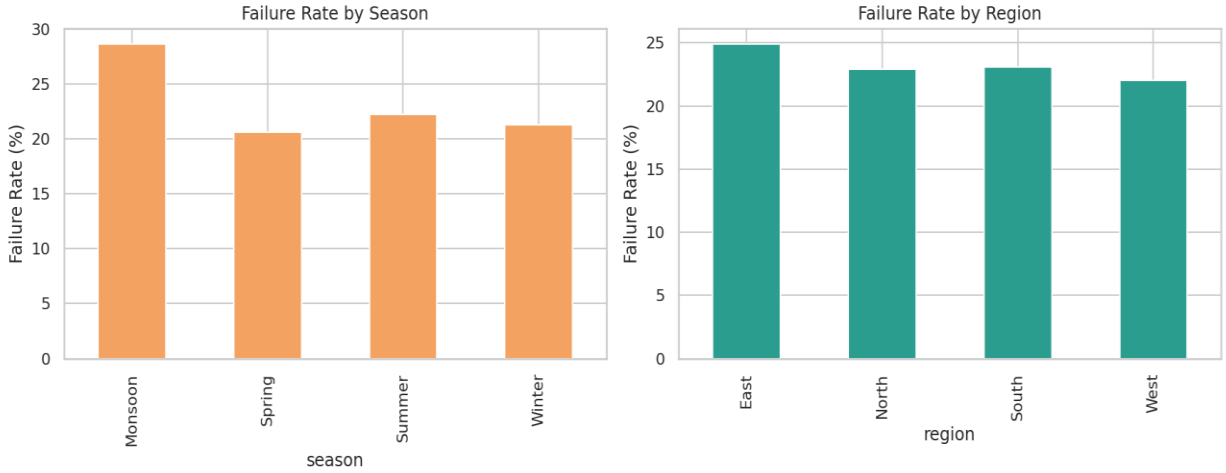

Fig. 5: Failure Rate by Season and Geographic Region

3.2 Model Validation

Five-fold stratified cross-validation was performed on the training dataset to assess the stability of the RF and SVM models before evaluation on the independent held-out test set. Random Forest achieved a mean F1-score of 0.513 ± 0.017 across the five folds, whereas SVM achieved 0.465 ± 0.007. The relatively small variation among the fold-level scores indicates consistent performance across the

validation partitions. The results provide evidence of stable cross-validation performance, although the absence of overfitting should ultimately be assessed by comparing cross-validation results with performance on the independent held-out test set rather than inferred from cross-validation variability alone.

3.3 Confusion Matrices for Failure Identification

Fig. 6 presents the confusion matrices obtained for Random Forest and SVM on the held-out test dataset. Random Forest correctly identified 136 of the 279 actual failure cases, while producing 81 false-positive prediction. SVM correctly identified 140 of the 279 actual failures but generated 159 false-positive predictions.

The confusion matrices demonstrate an important difference between the two models. Although SVM identified four more actual failures than Random Forest, it did so at the expense of substantially more false alarms. Random Forest therefore produced fewer false-positive maintenance alerts in the evaluated test set, whereas SVM showed a slightly greater ability to capture the positive failure class. These differences are consistent with the precision and recall values presented in Table 2**.**

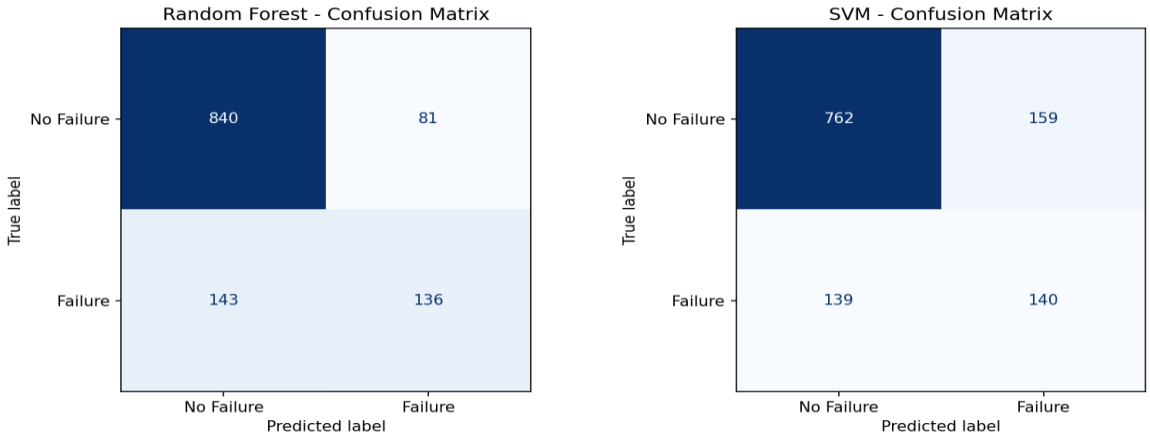

Fig. 6: Confusion Matrices for Random Forest (left) and SVM (right)

3.4 Receiver Operating Characteristic Curve for Both Models

Fig. 7 presents the receiver operating characteristic (ROC) curves for Random Forest and SVM. Random Forest achieved an area under the ROC curve (ROC-AUC) of 0.798, whereas SVM achieved 0.723. The higher ROC-AUC obtained by Random Forest indicates stronger overall discrimination between failure and non-failure cases across the range of classification thresholds evaluated. Nevertheless, the ROC-AUC values also indicate that neither model provides perfect separation between the two classes, demonstrating the inherent difficulty of predicting equipment failure from the available condition-monitoring variables.

Fig. 7 presents the receiver operating characteristic (ROC) curves for Random Forest and SVM. Random Forest achieved an area under the ROC curve (ROC-AUC) of 0.798, whereas SVM achieved 0.723. The higher ROC-AUC obtained by Random Forest indicates stronger overall discrimination between failure and non-failure cases across the range of classification thresholds evaluated. Nevertheless, the ROC-AUC values also indicate that neither model provides perfect separation between the two classes, demonstrating the inherent difficulty of predicting equipment failure from the available condition-monitoring variables.

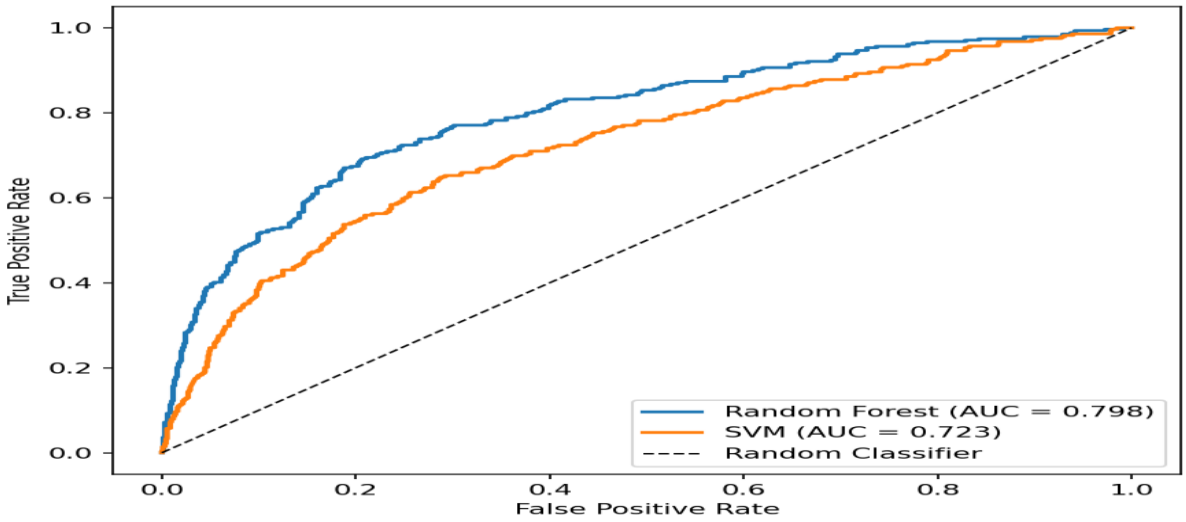

Fig. : ROC curves for Random Forest Versus SVM

3.4 Receiver Operating Characteristic Curve for Both Models

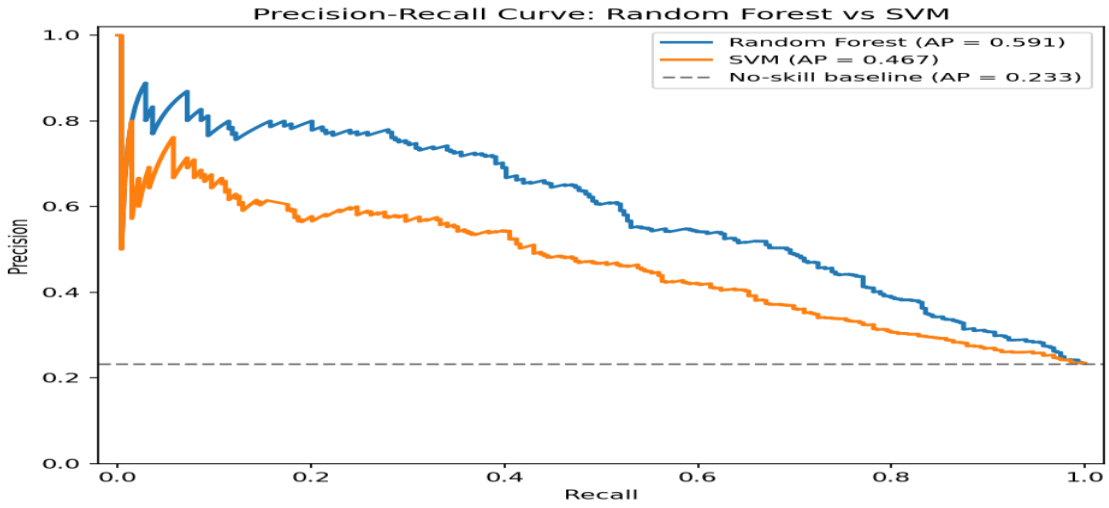

Fig. 7: Precision-Recall curve for Random Forest Versus SVM

Precision-recall curve shown in Fig. 7 establishes that both RF and SVM converged toward approximately 0.23 as recall approached 1.0, matching the dataset's 23.22% failure prevalence, the expected behaviour when a classifier is forced to flag nearly every case as a failure. Across almost the entire recall range, the Random Forest curve dominated the SVM curve, meaning that for any given fraction of failures captured, Random Forest generated fewer false alarms. Feature importance analysis further identified historical failure count, equipment age, corrosion index and dissolved-gas concentrations as the dominant predictors in the Random Forest model, in agreement with established transmission asset degradation theory and the correlation patterns observed during exploratory analysis.

3.6 Performance Evaluation Comparison

Table 2 summarises the performance of Random Forest and SVM on the independent 1,200-record held-out test set, while Fig. 8 provides a graphical comparison of the corresponding performance metrics

Table 2: Model performance comparison on the 1,200-record held-out test set

Metric	Random Forest	Support Vector Machine
Accuracy	0.813	0.752
Precision	0.627	0.468
Recall (Sensitivity)	0.488	0.502
F1-Score	0.548	0.484
ROC-AUC	0.798	0.723
Average Precision (PR-AUC)	0.591	0.467

Fig. 8 provides a graphical comparison of the performance of the Random Forest and SVM models across accuracy, precision, recall, F1-score, and ROC-AUC. The figure shows that Random Forest achieved higher accuracy (0.813), precision (0.627), F1-score (0.548), and ROC-AUC (0.798) than SVM, which recorded corresponding values of 0.752, 0.468, 0.484, and 0.723, respectively. In contrast, SVM produced a slightly higher recall (0.502) than Random Forest (0.488), indicating that it identified a marginally larger proportion of the actual failure cases at the selected classification threshold. The visual comparison therefore confirms the numerical results presented in Table 2 and highlights the different performance characteristics of the two classifiers.As shown in Table 2 and Figure 9, Random Forest achieved higher accuracy (81.3% versus 75.2%), precision (62.7% versus 46.8%), F1-score (54.8% versus 48.4%), ROC-AUC (79.8% versus 72.3%), and average precision (59.1% versus 46.7%) than SVM. SVM, however, achieved a slightly higher recall of 50.2% compared with 48.8% for Random Forest. Thus, the two models exhibited different error characteristics: Random Forest generated more precise failure predictions, whereas SVM identified a slightly larger proportion of the actual failures at the selected classification threshold.

The difference in recall is also reflected in the confusion matrices in Fig. 6, where SVM correctly identified 140 of the 279 actual failures compared with 136 identified by Random Forest.

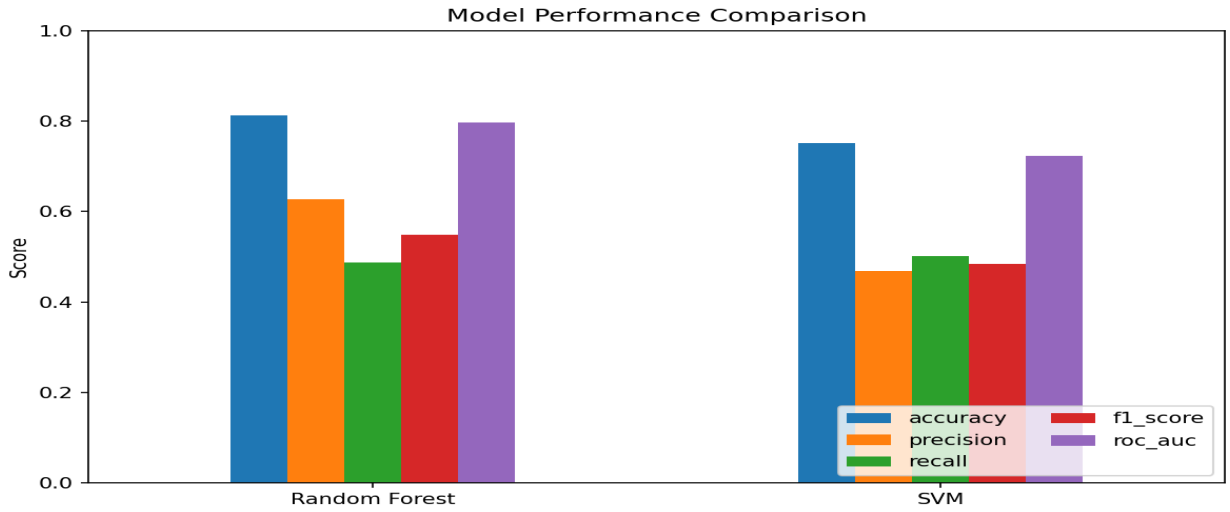

Fig. 8: Comparative performance of Random Forest and SVM across all metrics

3However, SVM generated 159 false-positive predictions compared with 81 for Random Forest, indicating a substantially greater false-alarm burden at the evaluated threshold. The combined evidence from Table 2 and Figs. 6–8 therefore demonstrates that model performance depends on the operational objective and the relative consequences assigned to missed failures and false alarms.

3.7 Summary of Key Findings

The results demonstrate the performance and practical characteristics of the RF and SVM models for predictive failure detection in power transmission equipment. The principal findings are summarised below.

Data and exploratory findings

i. The dataset contained 6,000 records with a failure prevalence of 23.22%, corresponding to an approximately 3.31:1 non-failure-to-failure class ratio.

ii. The raw numerical monitoring variables exhibited an overall sensor-dropout rate of 2.91%, equivalent to 3,662 missing cells out of 126,000 observations, as shown in Fig. 2.

iii. Transformers accounted for approximately 28% of the recorded failures and exhibited the highest observed failure rate among the equipment classes, followed by conductors and circuit breakers (Fig. 3).

iv. Partial-discharge magnitude, dissolved acetylene (C_2H_2), and surface temperature exhibited strong bivariate associations with failure in the exploratory correlation analysis (Fig. 4).

v. The monsoon period recorded the highest average failure rate of 24.8%, reaching approximately 29% in the East region, whereas October–November recorded the lowest observed failure rate of 21.8% (Fig. 5).

vi. Historical failure count, equipment age, corrosion index, and dissolved-gas concentrations were among the dominant predictors identified by the Random Forest model.

i. Five-fold stratified cross-validation produced a mean F1-score of 0.513 ± 0.017 for Random Forest and 0.465 ± 0.007 for SVM, indicating relatively stable performance across the validation folds.

ii. On the independent 1,200-record held-out test set, Random Forest achieved higher accuracy, precision, F1-score, ROC-AUC, and average precision than SVM (Table 2 and Fig. 8).

iii. SVM achieved slightly higher recall than Random Forest (50.2% versus 48.8%) and correctly identified 140 of the 279 actual failures compared with 136 identified by Random Forest (Figure 6).

iv. SVM generated substantially more false-positive predictions than Random Forest (159 versus 81), indicating a greater false-alarm burden at the evaluated classification threshold (Fig. 6).

v. The ROC curves demonstrated higher overall discriminative performance for Random Forest (AUC = 0.798) than SVM (AUC = 0.723) (Fig. 7).

vi. The precision–recall analysis showed that Random Forest generally maintained higher precision across the evaluated recall range (Fig. 8).

Both models were reported to train and generate predictions within seconds on standard CPU hardware without requiring GPU acceleration or cloud infrastructure. This computational characteristic suggests that both approaches could be implemented in environments where computational resources are limited, although actual deployment performance would depend on the hardware, data-ingestion architecture, retraining frequency, and operational requirements of the utility.

The results indicate that Random Forest provides a useful basis for risk ranking and maintenance prioritisation because it achieved higher precision, F1-score, ROC-AUC, and average precision than SVM on the held-out test set. However, its recall of 48.8% also indicates that a substantial proportion of actual failures remained undetected at the selected classification

threshold. Consequently, the model should be used as a decision-support and prioritisation tool rather than as a substitute for engineering inspection, condition assessment, and maintenance judgement. SVM may remain useful in applications where increasing failure detection sensitivity is prioritised, although the associated increase in false-positive predictions would need to be considered in operational planning.

4.0 Conclusion

This study implemented and compared Random Forest and RBF-kernel Support Vector Machine models for predicting failure in power transmission equipment within a 30 days horizon, using a realistic, multi-parameter condition monitoring dataset spanning five heterogeneous equipment classes and four Nigerian transmission regions.

Random Forest consistently outperformed SVM across all metric measured: accuracy (81.3%), precision (62.7%), F1-score (54.8%), ROC-AUC (79.8%) and average precision (59.1%) confirming primary predictive suitable model for deployment, while both models generalised stably under 5-fold cross validation and trained efficiently on standard CPU hardware. Historical failure count, equipment age, corrosion index and dissolved-gas concentrations emerged as the leading predictors of failure, corroborating established asset degradation theory.

Acknowledgement

The author sincerely acknowledges consistent encouragement, constructive criticism, and technical support of the Academic Staff of the Department of Electrical and Computer Engineering, Kwara State University, Malete, for making the research a successful one.

5.0 References

Achouch, M., Dimitrova, M., Dhouib, R., Ibrahim, H., Adda, M., Sattarpanah Karganroudi, S., Ziane, K., & Aminzadeh, A. (2023). Predictive maintenance and fault monitoring enabled by machine learning: Experimental analysis of a TA-48 multistage centrifugal plant compressor. *Applied Sciences*, 13, 3, 1790, https://doi.org/10.3390/app13031790

Adebowale, O. J. (2026). Design and simulation of a predictive maintenance model for a power transformer: A case study of a transmission substation in Nigeria. *Malaysian Journal of Science and Advanced Technology*, DOI: 10.56532/mjsat.v6i2.679

Adu-Gyamfi, K. (2026). Dissolved gas analysis-based power transformer fault diagnosis using random forest: A comparative evaluation against K-nearest neighbors, support vector machine, and artificial neural network classifiers *Engineer.ing Archive*. https://doi.org/10.31224/8150

Agrawal, S., & Patel, A. (2021). SAG cluster: An unsupervised graph clustering based on collaborative similarity for community detection in complex networks. *Physica A: Statistical Mechanics and Its Applications*, 563, 125459, https://doi.org/10.1016/j.physa.2020.125459

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, 1, 2549, https://doi.org/10.1038/s41598-025-86554-2

Aydın, C., & Evrentuğ, B. (2025). Evaluation of predictive maintenance efficiency with the comparison of machine learning models in machining production process in brake industry. *PeerJ Computer Science*, DOI: 10.7717/peerj-cs.2999

Balabantaraya, R., Chatterjee, A., Sahoo, A. K., Raj, A., & Khandual, B. (2026). A machine learning framework for improved fault diagnosis in service transformers using dissolved gas analysis. *Electrica*, 26, DOI: 10.5152/electrica.2026.25270

Cervantes-Bobadilla, M., García-Morales, J., Saavedra-Benítez, Y., Hernández-Pérez, J.

A., Adam-Medina, M., Guerrero-Ramírez, G. V., & Escobar-Jímenez, R. F. (2023). Multiple fault detection and isolation using artificial neural networks in sensors of an internal combustion engine. *Engineering Applications of Artificial Intelligence*, 117, 105524, https://doi.org/10.1016/j.engappai.2022.105524.

Davies, A. (2012). *Handbook of condition monitoring: Techniques and methodology*. Springer.

Faria, H., Jr., Costa, J. G. S., & Olivas, J. L. M. (2015). A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renewable and Sustainable Energy Reviews*, 46, pp. 201–209.

Folorunso, T. N., Idowu, K., & Balogun, A. (2025). AI-driven financial risk forecasting: A multi-model ensemble approach for enterprise decision intelligence. *Journal of Computational Analysis and Applications*, 34, 3. https://eudoxuspress.com/index.php/pub/article/view/5080

García, L., Parra, L., Jimenez, J. M., Lloret, J., & Lorenz, P. (2020). IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. *Sensors*, 20, 4, 1042, https://doi.org/10.3390/s20041042

Guan, S., Yang, H., & Wu, T. (2023). Transformer fault diagnosis method based on TLR-ADASYN balanced dataset. *Scientific Reports*, 13, 22957, https://doi.org/10.1038/s41598-023-49901-9

Huda, A. N., & Taib, S. (2013). Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment. *Applied Thermal Engineering*, 61, 2, pp. 220–227.

Idowu, K., Cherotich, V., Aniebonam, E. E., Odozor, L., & During, A. (2025). Integrating AI and machine learning into cyber risk management for critical utility systems. *International Journal of Artificial Intelligence Engineering and Technology*. https://doi.org/10.54660/IJAIET.2025.6.2.32-48

Idowu, K., Leslie, A., Twayana, S., Nitzan, Y., & Waheed, B. O. (2026). Real-time detection of DDoS and phishing attacks for enhanced banking security. *International Journal of Scientific Research in Science Engineering and Technology*. https://doi.org/10.32628/IJSRSET2613313

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20, 7, pp. 1483–1510.

Koffa, D., et al. (2025). Machine learning framework for predictive maintenance of power transformers in resource-constrained Nigerian distribution networks. *Arid Zone Journal of Engineering, Technology and Environment*, 21, 4, pp. 1026–1038.

Kumar, A., Shankar, R., & Thakur, L. S. (2018). A big data driven sustainable manufacturing framework for condition-based maintenance prediction. *Journal of Computational Science*, 27, pp. 428–439.

Lee, J., Park, J., & Park, H. (2025). Optimizing cable replacement schedules in power grids: A data-driven approach with machine learning and power flow analysis. *IEEJ Transactions on Electrical and Electronic Engineering*, 20, 9, pp. 1497–1502.

Moleda, M., Małysiak-Mrozek, B., Ding, W., Sunderam, V., & Mrozek, D. (2023). From corrective to predictive maintenance—A review of maintenance approaches for the power industry. *Sensors*, 23, 13, 5970, https://doi.org/10.3390/s23135970

Najafzadeh, M., et al. (2024). Fault detection, classification and localization along the power grid line using optimized machine learning algorithms. *International Journal of Computational Intelligence Systems*, 17,

Article 49, https://doi.org/10.1007/s44196-024-00434-7

Ouadah, A., Zemmouchi-Ghomari, L., & Salhi, N. (2022). Selecting an appropriate supervised machine learning algorithm for predictive maintenance. *The International Journal of Advanced Manufacturing Technology*, 119, 7, pp. 4277–4301.

Qin, S., Wang, K., Ma, X., Wang, W., & Li, M. (2017). Ensemble learning-based wind turbine fault prediction method with adaptive feature selection. In B. Zou, Q. Han, G. Sun, W. Jing, X. Peng, & Z. Lu (Eds.), *Data Science: ICPCSEE 2017* (Communications in Computer and Information Science, 728, pp. 559–570). Springer. https://doi.org/10.1007/978-981-10-6388-6_49

Rokach, L., & Maimon, O. (2005). Top-down induction of decision trees classifiers—A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35, 4, pp. 476–487.

Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, 121549, https://doi.org/10.1016/j.eswa.2023.121549.

Ukiwe, E. K., Adeshina, S. A., & Tsado, J. (2023). Techniques of infrared thermography for condition monitoring of electrical power equipment. *Journal of Electrical Systems and Information Technology*, 10, 1, 49, https://doi.org/10.1186/s43067-023-00115-z

Yang, Y., & Iqbal, M. Z. (2025). Cost-optimised machine learning model comparison for predictive maintenance. *Electronics*, 14, 12, 2497, https://doi.org/10.3390/electronics14122497

Declaration

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors, based on the information supplied.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical Consideration

Not applicable

Author; contributions

Lambe Mutalub Adesina conceptualized the study, designed the methodology, and supervised the research. Joseph Ibrahim Kuta Joshua conducted data preparation, model implementation, and statistical analysis. Ogunbiyi Olalekan and Abdul-Waheed Musa contributed to literature review, data interpretation, and validation. Monsurat Balogun and Jimada-Ojuolape Bilkisu contributed to manuscript preparation, visualization, critical review, and editing. All authors reviewed and approved the final manuscript.