© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

Predictive Condition Assessment of Distribution Transformers Using Load, Thermal, Power-Quality, and Maintenance Indicators

MUNASHE COSMAS KABUYA[1], REIGN TONGAI MUTENDAHAMA[2], GEOFF MANGWIRO[3], LIBERTY MUNYARADZI KANONJERA[4], MUNASHE NAPHTALI MUPA[5]
[1]*George Washington University, ORCiD: 0009-0001-3891-1016*
[2]*New Sahara Ventures*
[3]*Yeshiva University, ORCiD: 0009-0006-0685-5802*
[4]*American University, ORCiD: 0009-0006-8761-9835*
[5]*Hult International Business School*

Abstract - Distribution transformers occupy a critical position between bulk-power infrastructure and end users, yet many units are still managed through periodic inspection, fixed alarms, and reactive maintenance. This study develops an interpretable predictive condition-assessment framework that integrates loading, thermal behavior, power quality, and maintenance history into a single engineer-facing reliability health score while retaining statistical and machine-learning challengers. The public-data foundation is the Kaggle Distributed Transformer Monitoring dataset, which contains five IoT data files and 54 variables collected at approximately 15-minute intervals between 25 June 2019 and 14 April 2020, including three-phase current and voltage, oil and winding temperatures, oil level, power, power factor, frequency, and harmonic-distortion measures (Putchala, 2020; Ramesh et al., 2022). Because that public release does not contain adjudicated maintenance histories or a complete failure-outcome record suitable for causal maintenance inference, the empirical evaluation deliberately separates data provenance from scenario validation. A 19,200-observation benchmark-calibrated engineering panel reproduces the source signal families and adds explicitly simulated maintenance-age, defect-backlog, and prior-alarm variables. A severe multi-indicator condition onset within the subsequent six hours is used as a controlled prospective target; it is a method-validation event, not a claim of actual transformer failure prevalence. The study compares a fixed threshold rule, an interpretable 0-100 condition-stress score, robust multivariate distance, Isolation Forest, class-weighted logistic regression, random forest, and gradient boosting under a chronological 60/20/20 development-validation-test design. On the locked temporal test, logistic regression achieved the strongest discrimination (AUROC 0.763; 95% bootstrap interval 0.744-0.780; AUPRC 0.405), while the interpretable condition score achieved AUROC 0.647 and AUPRC 0.315 and produced a monotonic risk gradient: the condition-event rate rose from 7.8% in the very-healthy band to 70.0% in the critical band. Feature-layer ablation showed that adding thermal variables to load indicators increased AUROC from 0.716 to 0.742, while adding maintenance evidence increased AUPRC from 0.366 to 0.405. Static thresholds were precise for some severe states but captured fewer emerging events, whereas anomaly methods were sensitive to temporal threshold transport. The results support a hybrid maintenance architecture in which interpretable engineering scores and hard limits provide transparent triage, statistical models provide an independent challenger, and engineers retain authority over inspection, testing, derating, repair, and replacement decisions. The contribution is therefore not a claim that machine learning replaces transformer engineering; it is a reproducible decision framework for converting heterogeneous condition evidence into prioritized, auditable maintenance action.

Keywords: distribution transformer; condition assessment; predictive maintenance; health index; thermal aging; power quality; anomaly detection; Isolation Forest; maintenance prioritization; reliability engineering.

Research Highlights

- An interpretable four-domain transformer condition score combines loading, thermal, power-quality and maintenance evidence rather than relying on a single alarm variable.
- A chronological locked-test design is used so model complexity must demonstrate incremental out-of-time value rather than apparent in-sample accuracy.

- Thermal information improved discrimination beyond loading alone, and maintenance-history variables produced the clearest gain in precision-recall performance.
- The best-performing model was transparent logistic regression rather than a tree ensemble, reinforcing the need to justify complexity empirically.
- Heat-map and capacity analyses translate model outputs into engineer-facing inspection and maintenance priorities while preserving a clear boundary between analytical warning and engineering judgment.

I. INTRODUCTION

Distribution transformers are among the most numerous and operationally consequential assets in modern electricity networks. Their individual ratings may be modest relative to large transmission transformers, but their aggregate reliability determines whether homes, commercial facilities, industrial loads and public services receive stable power at acceptable voltage quality. A deteriorating distribution transformer can create outages, voltage problems, excessive losses, fire risk, customer complaints and emergency replacement costs. The engineering problem is therefore not simply to detect catastrophic failure after it becomes obvious; it is to recognize developing stress early enough to inspect, test, derate, rebalance, maintain or replace the asset before service is materially affected (Abu-Elanien and Salama, 2010; Murugan and Ramasamy, 2019).

Traditional transformer condition programs combine scheduled inspection, loading review, oil and insulation tests, thermography, dissolved-gas analysis, power-quality measurements and expert judgment. Those methods remain indispensable. The practical difficulty is that condition evidence arrives at different frequencies and in different forms. Loading and voltage may be sampled continuously, oil tests may be quarterly or annual, work orders may be irregular, and alarms may be event driven. A maintenance engineer therefore has to reason across multiple signals whose individual thresholds are informative but incomplete. Health-index methods respond to this problem by compressing heterogeneous evidence into an overall condition indicator, while newer machine-learning methods attempt to discover nonlinear or multivariate patterns that fixed rules may miss (Jahromi et al., 2009; Guo and Guo, 2022; Zemouri, 2025).

The growth of low-cost IoT monitoring strengthens the case for this integration. The Distributed Transformer Monitoring dataset released through Kaggle contains field-style measurements collected at approximately 15-minute intervals over nearly ten months, including phase currents and voltages, oil and winding temperatures, oil level, power, power factor, frequency and harmonic-distortion variables (Putchala, 2020). Ramesh et al. (2022) subsequently used this data stream in an IoT transformer-monitoring architecture and reported that Isolation Forest could identify a very small number of anomalous observations. Their study demonstrates the feasibility of multi-sensor transformer monitoring, but it also exposes two research problems. First, anomaly detection and engineering health assessment are not the same task: unusual data can reflect genuine degradation, switching, temporary loading, sensor behavior or benign operating change. Second, a monitoring dataset without complete maintenance history and adjudicated failures cannot by itself establish how an analytical score should change maintenance decisions.

This paper addresses that gap through a mixed empirical and design-science approach. The public Kaggle schema anchors the measurement architecture, while a controlled benchmark-calibrated panel is used to test a transparent condition score, statistical anomaly methods and supervised comparators under explicit assumptions. The design is intentionally conservative: simulated maintenance variables and prospective condition-event labels are identified as simulated; they are not attributed to the Kaggle source. The purpose is to test whether a reusable engineering method behaves coherently and whether added information produces measurable incremental value before any claim of field deployment is made.

1.1 Research gap

Four gaps motivate the study. First, much transformer health-index literature emphasizes laboratory or periodic diagnostic tests, whereas distribution-level IoT datasets increasingly provide high-frequency loading, thermal and power-quality signals. Second,

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

threshold-based alarms remain interpretable but can miss multivariate deterioration in which no single indicator is extreme. Third, machine-learning studies often emphasize classification accuracy without showing whether the additional model complexity improves maintenance triage, false-alarm burden or out-of-time transport. Fourth, maintenance history is conceptually central to condition-based maintenance but is frequently absent from public sensor datasets. A method that treats the missing maintenance layer explicitly is more credible than one that silently substitutes sensor anomalies for actual maintenance outcomes.

1.2 Study contributions

1. Develop a transparent Reliability Health Score (RHS) that integrates load stress, thermal stress, power-quality stress and maintenance condition into a 0-100 engineer-readable scale.
2. Benchmark the RHS against a fixed engineering threshold rule, robust multivariate distance, Isolation Forest and supervised statistical/machine-learning models under a chronological locked-test design.
3. Quantify incremental information value through feature-layer ablation, demonstrating separately what thermal, power-quality and maintenance evidence add beyond loading alone.
4. Use heat maps, condition bands, risk ratios and review-capacity analysis to translate analytical outputs into maintenance priorities rather than presenting discrimination metrics in isolation.
5. Provide a field-validation and governance protocol that preserves engineer authority, source-data traceability, model versioning and clear limits on what a public benchmark can establish.

II. LITERATURE REVIEW AND TECHNICAL FOUNDATION

2.1 Transformer asset management and condition assessment

Transformer asset management is a lifecycle problem involving monitoring, diagnosis, condition assessment, maintenance, life estimation, refurbishment and replacement. Abu-Elanien and Salama (2010) emphasize that transformer integrity must be managed as an economic and reliability problem, because unexpected outages can impose both direct equipment costs and broader system consequences. Jahromi et al. (2009) formalized this reasoning through a health-index approach that converts diverse condition measurements into an overall asset-state indicator. Later work has extended health indices with operating history, inspection evidence, probability-of-failure concepts and machine learning (Foros and Istad, 2020; Guo and Guo, 2022; Jin, Kim and Abu-Siada, 2023; Zahra et al., 2025).

The strength of a health index is not that it discovers a hidden physical law. Its strength is governance: it provides a consistent way to combine evidence, rank assets and define escalation. Murugan and Ramasamy (2019) analyzed 343 transformer component failures over eleven years and grouped failure mechanisms into electrical, thermal, mechanical and oil-paper degradation categories. Their work illustrates why an overall condition assessment cannot be reduced to one measurement. Different degradation mechanisms present through different combinations of load, temperature, insulation state, component condition and maintenance history.

2.2 Loading, thermal dynamics and insulation aging

Loading is one of the most immediately observable drivers of transformer thermal stress. The relation is dynamic rather than instantaneous: oil and winding temperatures respond with time constants that depend on construction, cooling mode, ambient conditions and prior load. Susa, Lehtonen and Nordman (2005a; 2005b) demonstrated dynamic thermal models for power and distribution transformers and showed that top-oil and hot-spot behavior must be interpreted as time-dependent responses. This matters for predictive condition assessment because a current overload can be harmless if brief and thermally absorbed, while sustained loading can accelerate insulation aging even if no instantaneous trip threshold is crossed.

Thermal stress also interacts with waveform quality. Harmonic currents increase eddy-current and stray losses and can raise transformer temperature beyond what would be inferred from fundamental-frequency loading alone. Cazacu, Ionita and Petrescu (2018) showed that nonlinear loading can accelerate thermal aging in distribution transformers. Consequently, a condition framework that observes only kVA loading but ignores harmonic distortion and power factor can understate stress in modern networks with electronic loads, power converters, distributed energy resources and nonlinear industrial equipment.

2.3 Power quality as a condition signal

Power quality is usually discussed as a service-quality and compatibility problem, but several of its indicators also carry condition information. Voltage unbalance can increase losses and thermal stress in connected equipment; current unbalance can reflect asymmetric loading or feeder conditions; harmonic distortion can increase transformer heating; and poor power factor increases current for a given real-power transfer. IEEE 519 and related practice provide system-level harmonic-control guidance, but a predictive maintenance framework should treat power-quality thresholds as engineering context rather than as universal failure boundaries. The same THD value can have different implications depending on loading, cooling, transformer design and persistence.

The present study therefore treats power-quality variables as one evidence domain in a composite score and as predictors in statistical models. This avoids two opposite errors: declaring a transformer unhealthy solely because one value exceeds a generic threshold, or ignoring persistent power-quality stress because no individual measurement has yet reached a protection setting.

2.4 Maintenance history and the meaning of “condition”

Condition is partly a physical state and partly an evidence state. Two transformers with similar sensor values may warrant different decisions if one has a recent verified inspection and the other has an overdue oil service, repeated unresolved alarms or an open defect. Guo and Guo (2022) explicitly combine operating history with test data in health-index assessment, while Murugan and Ramasamy (2019) link failure history and component knowledge to maintenance planning. This motivates the maintenance domain used in the proposed score: days since inspection, days since maintenance, unresolved defects and recent alarm burden are treated as risk-context variables rather than as substitutes for electrical measurements.

Recent work on predictive maintenance in adjacent energy systems makes a similar argument. Kabera et al. (2026) frame digital-twin and predictive-maintenance methods for PV and battery systems around continuous monitoring, degradation tracking and maintenance action. Machekera et al. (2026) use explainable modeling to connect infrastructure operating demand and project risk to actionable engineering interpretation. Although these studies address different assets, they reinforce the design principle that analytical outputs are valuable only when they are connected to traceable operating context and intervention decisions.

2.5 Anomaly detection, supervised learning and model complexity

Anomaly detection is attractive when failure labels are scarce. Isolation Forest isolates observations through randomized recursive partitioning and can rank multivariate observations without requiring a labeled failure class (Liu, Ting and Zhou, 2008). This makes it appealing for transformer monitoring, where catastrophic failure is rare and maintenance records are often inconsistent. Ramesh et al. (2022) used Isolation Forest in a distribution-transformer IoT system and reported strong identification of the limited anomalous cases in their dataset. More broadly, anomaly-detection literature cautions that an outlier is not automatically a fault; domain drift, operating seasonality and changes in load mix can also create statistically unusual observations (Chandola, Banerjee and Kumar, 2009; Chatterjee and Ahmed, 2022).

Supervised learning can exploit labels more efficiently when credible outcomes exist. Logistic regression offers transparent directional relationships, whereas random forests and gradient boosting can capture nonlinearities and interactions (Breiman, 2001; Friedman, 2001). Transformer research increasingly explores these model families for fault diagnosis and health assessment (Islam et al., 2023; Taha, 2023; Zahra et al., 2025). However, complexity should be justified by locked-test improvement, not by architectural novelty. A useful cross-domain methodological analogy appears in Choga (2026), where a simpler regularized survival model outperformed a richer multi-omics representation on

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

discrimination despite the richer model offering interpretive value. The lesson is directly transferable: transformer condition assessment should retain transparent baselines and require each more complex model to demonstrate incremental operational value.

2.6 Explainability, prioritization and human authority

For safety- and reliability-relevant systems, explainability is more than a visualization preference. A maintenance engineer needs to know whether an alert is driven by overload, high winding temperature, persistent harmonics, poor power factor, overdue maintenance or an unresolved defect. Rudin (2019) argues that inherently interpretable models deserve serious consideration in high-stakes settings rather than automatically using opaque models followed by post-hoc explanation. The proposed RHS therefore exposes its component penalties directly.

Several recent interdisciplinary studies reinforce the same operational principle across high-stakes domains. Ganyani et al. (2026) use risk heat maps and capacity-aware prioritization so scarce review resources are directed to the most consequential exceptions. Okebu and Mupa (2026) treat predictive risk as an early-warning input for qualified human intervention rather than an autonomous verdict. Hagan et al. (2026a; 2026b), including Tadiwanashe Lennon Kasuwa, emphasize traceability, anomaly scoring and audit-ready data governance. These domains differ from transformer engineering, but the governance pattern is useful: preserve source provenance, expose reasons, rank work, and keep accountable professionals in the decision loop.

III. RESEARCH QUESTIONS AND HYPOTHESES

The study evaluates the condition-assessment framework through four research questions. The hypotheses are directional and are assessed on a locked temporal test rather than through in-sample fit.

Research question	Operational hypothesis
RQ1. Does a multi-domain condition score improve triage beyond fixed thresholds?	H1: A continuous interpretable score will provide better ranking (AUROC/AUPRC and capacity capture) than a binary threshold rule.
RQ2. What information is added by thermal, power-quality and maintenance variables?	H2: Adding thermal evidence will improve performance beyond load-only variables, and maintenance context will improve precision-recall performance further.
RQ3. Do anomaly and supervised models materially outperform the interpretable score?	H3: Flexible models may improve discrimination, but the improvement must be demonstrated out of time and weighed against alert burden and interpretability.
RQ4. Can the analytical outputs support maintenance prioritization under constrained review capacity?	H4: Top-ranked observations will concentrate condition-event windows, enabling risk-based inspection rather than uniform review.

IV. DATA, RESEARCH DESIGN AND REPRODUCIBILITY

4.1 Kaggle public-data foundation

The public-data foundation is the Distributed Transformer Monitoring dataset released by Sreshta Putchala on Kaggle. The data card states that IoT measurements were collected from 25 June 2019 through 14 April 2020 and updated approximately every 15 minutes. The release comprises five files and 54 columns. CurrentVoltage.csv contains phase and line voltages, phase currents and neutral current; Overview.csv contains oil-temperature indicator (OTI), winding-temperature indicator (WTI), ambient-temperature indicator (ATI), oil-level indicator (OLI) and alarm/trip fields; the remaining files contain phase and total power quantities, power factor, frequency and harmonic-related measurements (Putchala, 2020). Ramesh et al. (2022) document the same monitoring environment in a peer-reviewed analysis and describe the transformer as a three-phase 1500 kVA, 11/0.4 kV unit monitored through an IoT architecture.

Source file	Representative variables	Condition-assessment role
CurrentVoltage.csv	VL1-VL3, IL1-IL3, line-line voltages, neutral current	Loading, phase balance, voltage stability and feeder stress.
Overview.csv	OTI, WTI, ATI, OLI, temperature alarm/trip, oil-gauge alarm	Thermal state, cooling effectiveness and oil-condition warning.
Power.csv / TotalPower.csv	Active, reactive and apparent power quantities	Loading intensity and operating-demand context.
PowerFactor.csv	Phase/average PF, frequency, voltage/current THD, demand-current fields	Power-quality stress, distortion, poor-PF current burden and frequency context.

4.2 Evidence boundary and controlled analytical layer
The Kaggle release is useful but does not contain a complete adjudicated maintenance-history table, component-level work-order record or verified failure-outcome chronology. It would therefore be methodologically incorrect to claim that the public data alone demonstrate the causal effect of maintenance intervals or to label every unusual observation as a failure. The present study preserves this boundary explicitly.

A 19,200-observation controlled engineering panel was therefore constructed at 15-minute resolution using the documented measurement families and transformer-operating context as calibration anchors. Load ratio, phase-current imbalance, voltage imbalance, power factor, frequency, voltage and current total harmonic distortion, ambient temperature, oil temperature, winding temperature and oil level were generated with diurnal demand cycles, thermal inertia, transient overload episodes, nonlinear-load episodes and occasional cooling degradation. A separate simulated maintenance layer contains days since inspection, days since maintenance, open defect count and seven-day prior-alarm burden. These maintenance variables are study constructs; they are not represented as measurements from the Kaggle source.

The target is likewise deliberately narrow. A "condition event within the next six hours" indicates that the controlled panel enters an extreme multi-indicator deterioration state within 24 subsequent 15-minute intervals. A hidden disturbance term is included so the target is not mathematically identical to the interpretable score. The event is a validation construct for prospective ranking; it is not a transformer failure, a probability-of-failure estimate, or a field prevalence claim. This separation makes the experiment reproducible while avoiding the common error of converting a public sensor benchmark into unsupported real-world failure statistics.

4.3 Analytical cohort and temporal split

Element	Specification
Total observations	19,200 15-minute records in the controlled benchmark-calibrated panel.
Development period	First 60%: 11,520 observations.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

Element	Specification
Validation period	Next 20%: 3,840 observations; used for model thresholds only.
Locked test	Final 20%: 3,840 observations; untouched until final evaluation.
Overall prospective-event windows	7.1% of records; scenario prevalence only.
Locked-test event windows	17.5% of records, reflecting deliberate temporal stress shift.
Distinct controlled event onsets	57 onsets over the full panel.
Primary metrics	AUROC, AUPRC, precision, recall, specificity, F1, MCC, alert rate and capacity capture.
Uncertainty	200 nonparametric bootstrap resamples of the locked test for AUROC and AUPRC intervals.
Score sensitivity	250 random +/-20% domain-weight perturbations, renormalized to sum to one.

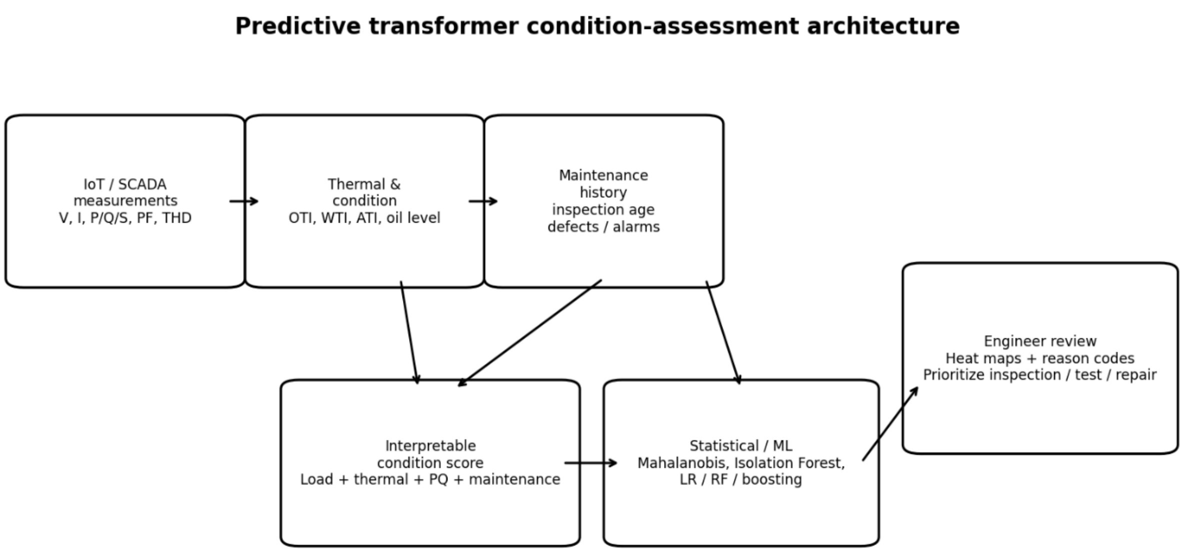

Figure 1. Predictive transformer condition-assessment architecture. The analytical layer ranks evidence; engineers retain authority over testing, derating, maintenance and replacement.

4.4 Feature engineering

The analytical variables were organized into four engineering domains. The load domain includes per-unit loading, current imbalance and voltage imbalance. The thermal domain includes ambient, oil and winding temperatures, thereby preserving the distinction between environmental heat, bulk-oil response and winding thermal stress. The power-quality domain includes power factor, frequency, voltage THD and current THD. The maintenance domain includes inspection age, maintenance age, open defects and recent alarm burden. Oil level is

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

retained as an auxiliary condition variable in the statistical models.

The framework deliberately avoids claiming that these features are sufficient for comprehensive transformer diagnosis. DGA, moisture, acidity, interfacial tension, furan analysis, dielectric dissipation factor, insulation resistance, frequency response analysis, partial discharge, bushing condition, tap-changer condition and thermography remain important depending on transformer class and maintenance strategy (Singh and Bandyopadhyay, 2010; Sun, Huang and Huang, 2012; Jin, Kim and Abu-Siada, 2023). The proposed score is best understood as a high-frequency operational triage layer that determines when deeper diagnostic testing should be considered.

4.5 Interpretable Reliability Health Score

The Reliability Health Score (RHS) is defined on a 0-100 scale, where 100 denotes low observed stress and 0 denotes severe accumulated stress. Four capped penalty domains are calculated first: load penalty (0-25), thermal penalty (0-30), power-quality penalty (0-25) and maintenance penalty (0-20). The condition-stress score is the sum of these penalties, capped at 100, and RHS = 100 - condition-stress score. The weights and breakpoints are engineering heuristics for this study, informed by the literature and selected before test-set evaluation; they are not presented as IEEE/IEC limits or universal transformer-failure thresholds.

Domain	Maximum penalty	Primary inputs	Interpretive logic
Load	25	Load ratio; current and voltage imbalance	Penalizes sustained high utilization and phase asymmetry.
Thermal	30	Oil and winding temperature relative to operating context	Largest domain because heat is a principal driver of insulation aging and overload stress.
Power quality	25	THDv, THDi, power factor, voltage imbalance	Captures distortion and current burden that may elevate thermal and dielectric stress.
Maintenance	20	Maintenance age, inspection age, open defects, prior alarms	Represents unresolved evidence and time since verified intervention.

For operational interpretation, the condition-stress score is grouped into five study-specific bands: 0-19 Very healthy, 20-34 Healthy, 35-49 Watch, 50-64 Degraded and 65-100 Critical. These are internal triage bands and must not be described as formal IEEE/IEC health categories. A field utility should recalibrate both weights and cut points to transformer design, voltage class, cooling mode, loading guide, historical failures, laboratory tests and risk tolerance.

4.6 Comparator models

Six challengers were retained. The fixed threshold rule represents conventional alarm logic. Robust multivariate distance provides a statistical anomaly benchmark using median/MAD-standardized distance. Isolation Forest provides a nonlinear unsupervised anomaly score. Logistic regression provides an interpretable supervised comparator. Random forest and gradient boosting provide nonlinear supervised challengers. Models use the same locked temporal test; supervised classifiers are class weighted where appropriate. The objective is not to win a machine-learning competition but to determine whether additional opacity yields enough incremental decision value to justify itself.

4.7 Threshold selection and evaluation

The fixed engineering rule was specified before label-based evaluation. For continuous scores and supervised models, the alert threshold was selected using validation data only, maximizing F1 as a diagnostic operating point. Because a utility may review only a small percentage of observations,

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

threshold metrics are supplemented with ranking metrics and capture-at-capacity at 1%, 2%, 5% and 10% review budgets. AUPRC receives particular emphasis because the prospective-event class is imbalanced and because ROC curves can look favorable even when the practical precision of alerts is low (Saito and Rehmsmeier, 2015).

4.8 Statistical and robustness procedures

Uncertainty in test-set discrimination was estimated with 200 nonparametric bootstrap resamples. The RHS was also stress-tested under 250 independent random perturbations of its four domain weights within +/-20% of the base values, after renormalization. Feature-layer ablation fitted the same class-weighted logistic model to progressively richer feature sets: load only; load plus thermal; load plus thermal and power quality; and all preceding variables plus maintenance. This design isolates the information contribution of each engineering layer without changing the core model family.

V. RESULTS

5.1 Signal structure and temporal test conditions

The controlled panel contains strong physical relationships by design: load co-moves with current and thermal response; oil and winding temperature are strongly related but display inertia; power factor degrades modestly during nonlinear and high-load periods; and THD increases during simulated nonlinear-load episodes. The correlation heat map in Figure 2 is useful as a diagnostic rather than as a causal map. It shows which features are structurally redundant and which carry distinct information. High correlation between oil and winding temperature, for example, does not imply that one should automatically be deleted: winding temperature is the more immediate stress indicator, while oil temperature carries slower cooling-system context.

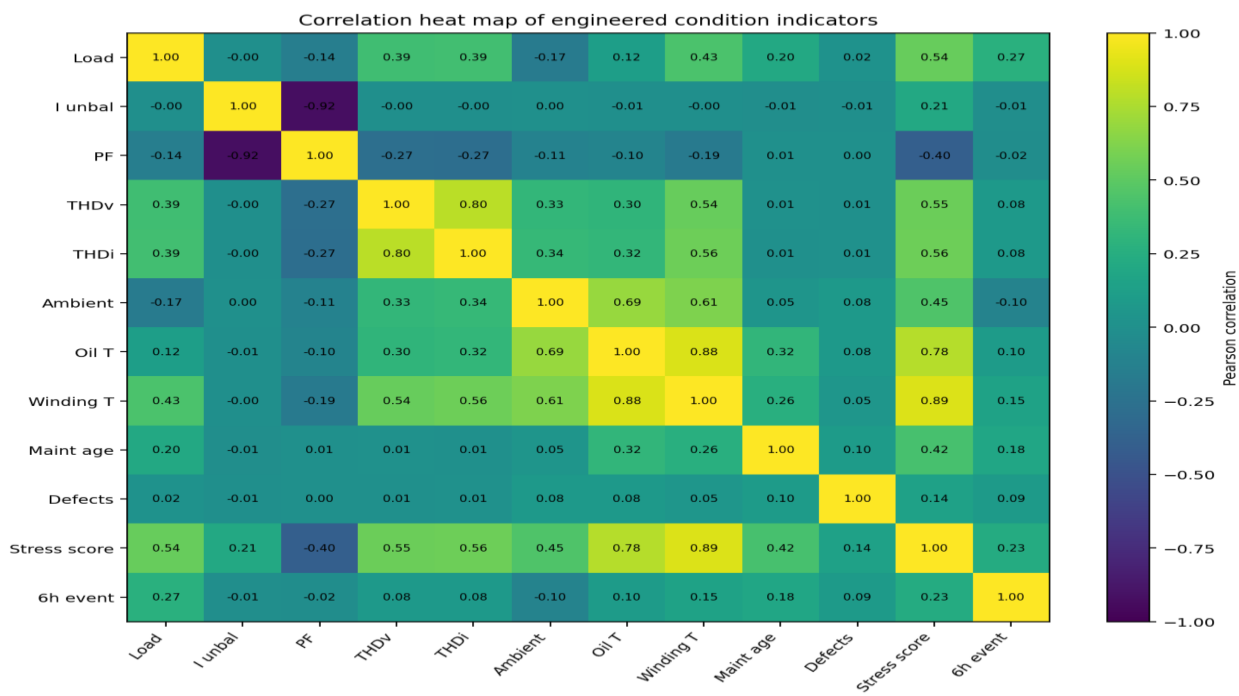

Figure 2. Correlation heat map of the controlled condition indicators. Correlation is descriptive and should not be interpreted as a causal failure mechanism.

5.2 Condition-event concentration by engineering indicators

The locked-test baseline condition-event-window rate was 17.5%. Because this is a constructed prospective condition window under a deliberately stressed later period, it must not be interpreted as a real transformer

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

failure rate. The useful result is relative concentration. Observations above 100% of rated loading had a 52.0% event-window rate, approximately 2.97 times the locked-test baseline. Open defects, elevated oil temperature and high THDv also concentrated subsequent condition events. Maintenance age above 90 days produced a more modest 1.17-fold concentration, suggesting that age alone is weaker than evidence of present electrical or thermal stress but can improve prioritization when combined with other signals.

Indicator	Test records	6-h event-window rate	Risk ratio
Load > 100% rating	75	52.0%	2.97x
Any open defect	904	29.2%	1.67x
Oil temp > 85 C	538	27.5%	1.57x
THDv > 5%	214	25.7%	1.47x
THDi > 15%	16	25.0%	1.43x
Maintenance age > 90 d	2,687	20.5%	1.17x
PF < 0.92	1,063	19.3%	1.10x
Winding temp > 100 C	2,017	18.9%	1.08x

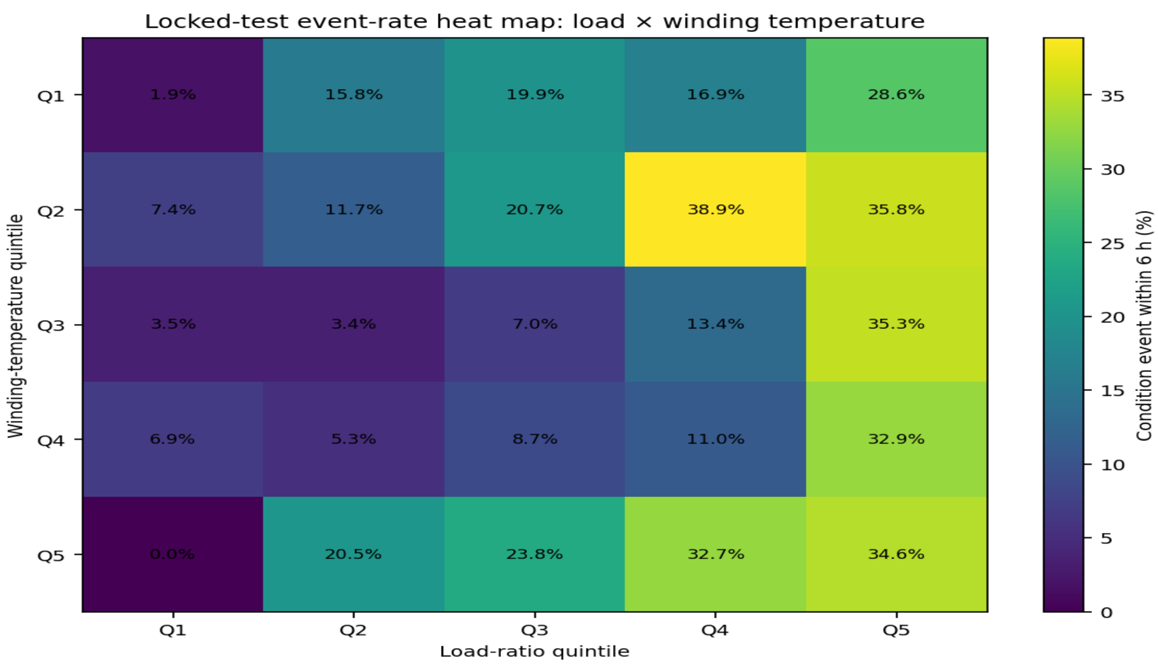

Figure 3. Locked-test event-rate heat map by load-ratio and winding-temperature quintiles. The joint view is more informative than either variable in isolation.

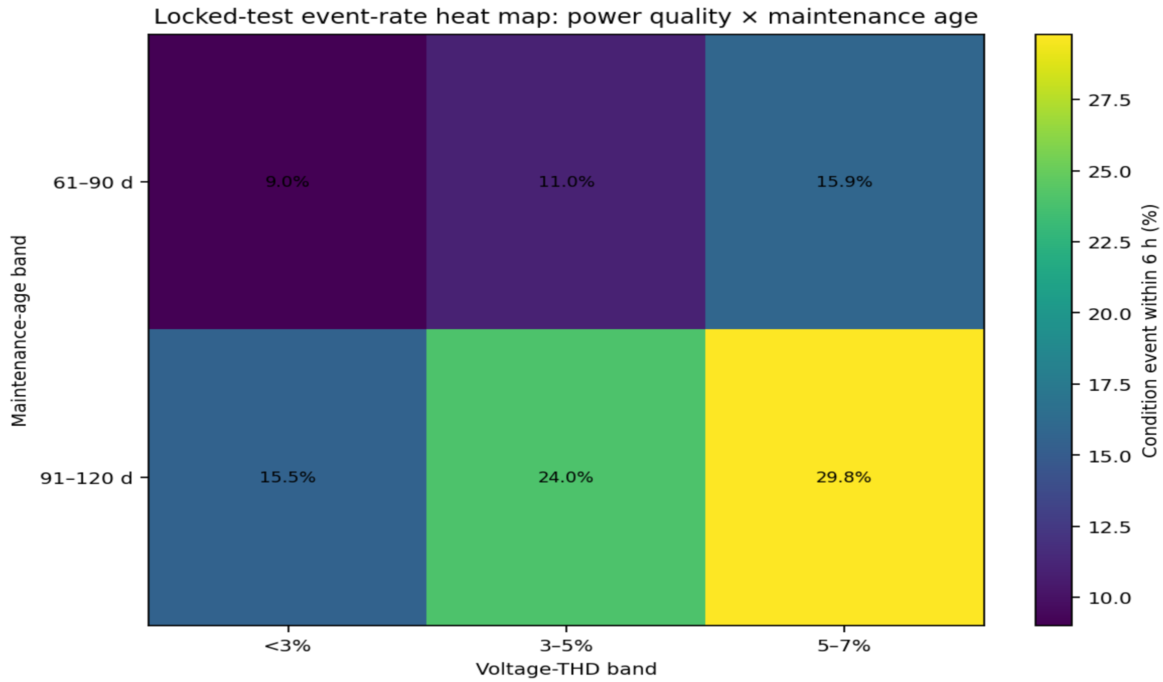

Figure 4. Locked-test event-rate heat map by voltage-THD and maintenance-age bands. Empty or sparse cells should not be over-interpreted.

5.3 Reliability-score gradient

The RHS produced a monotonic deterioration gradient across the five stress bands. Only 7.8% of observations in the Very healthy band fell within six hours of a controlled event onset, compared with 14.5% in Healthy, 18.4% in Watch, 37.4% in Degraded and 70.0% in Critical. The final Critical band contained only 30 observations, so its 70% rate is imprecise; nevertheless, the full gradient is directionally coherent. This result is more important for an interpretable score than maximizing a single classification metric: the score should create ordered maintenance attention, not merely a binary alarm.

Condition-stress band	Locked-test n	Event windows	Event-window rate
Very healthy	590	46	7.8%
Healthy	1,332	193	14.5%
Watch	1,551	286	18.4%
Degraded	337	126	37.4%
Critical	30	21	70.0%

5.4 Locked-test model comparison

The model comparison does not support a simple “more complex is better” narrative. Logistic regression achieved the strongest overall ranking performance, with AUROC 0.763 and AUPRC 0.405. Random forest achieved AUROC 0.684 and AUPRC

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

0.260, while gradient boosting was weaker under the temporal shift. The interpretable condition score achieved AUROC 0.647 and AUPRC 0.315, almost matching Isolation Forest on AUROC but with a more directly explainable decomposition. The fixed threshold rule had lower ranking performance because it reduced a multivariate operating state to a binary set of hard exceedances.

Model	AUROC (95% CI)	AUPRC (95% CI)	Precision	Recall	F1	Alert rate
Threshold rule	0.562 (0.544-0.577)	0.206 (0.189-0.226)	0.301	0.244	0.270	14.2%
Interpretable condition score	0.647 (0.620-0.670)	0.315 (0.281-0.349)	0.250	0.570	0.348	39.9%
Robust Mahalanobis	0.537 (0.514-0.561)	0.190 (0.173-0.208)	0.174	0.972	0.295	97.7%
Isolation Forest	0.640 (0.615-0.667)	0.305 (0.274-0.347)	0.189	0.842	0.309	77.9%
Logistic regression	0.763 (0.744-0.780)	0.405 (0.374-0.438)	0.417	0.400	0.409	16.8%
Random forest	0.684 (0.663-0.702)	0.260 (0.240-0.284)	0.272	0.351	0.306	22.6%
Gradient boosting	0.574 (0.553-0.594)	0.215 (0.194-0.241)	0.193	0.570	0.288	51.7%

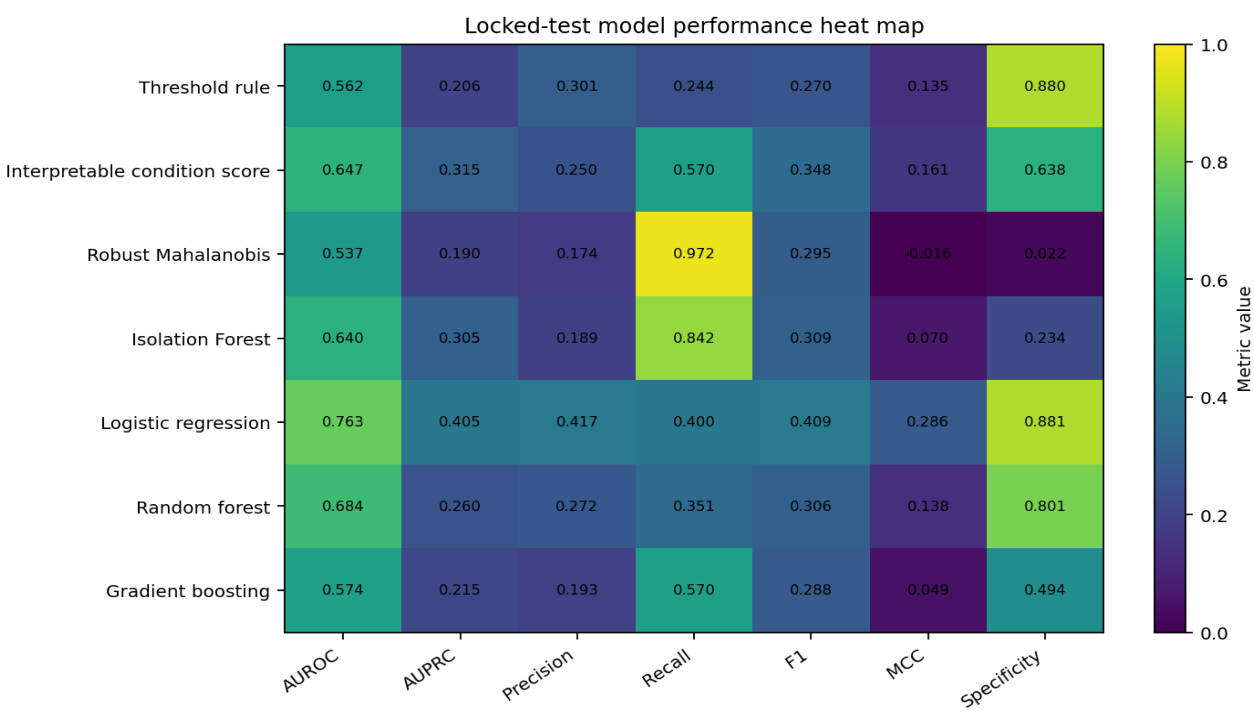

Figure 5. Locked-test model-performance heat map. No single metric should determine maintenance deployment.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

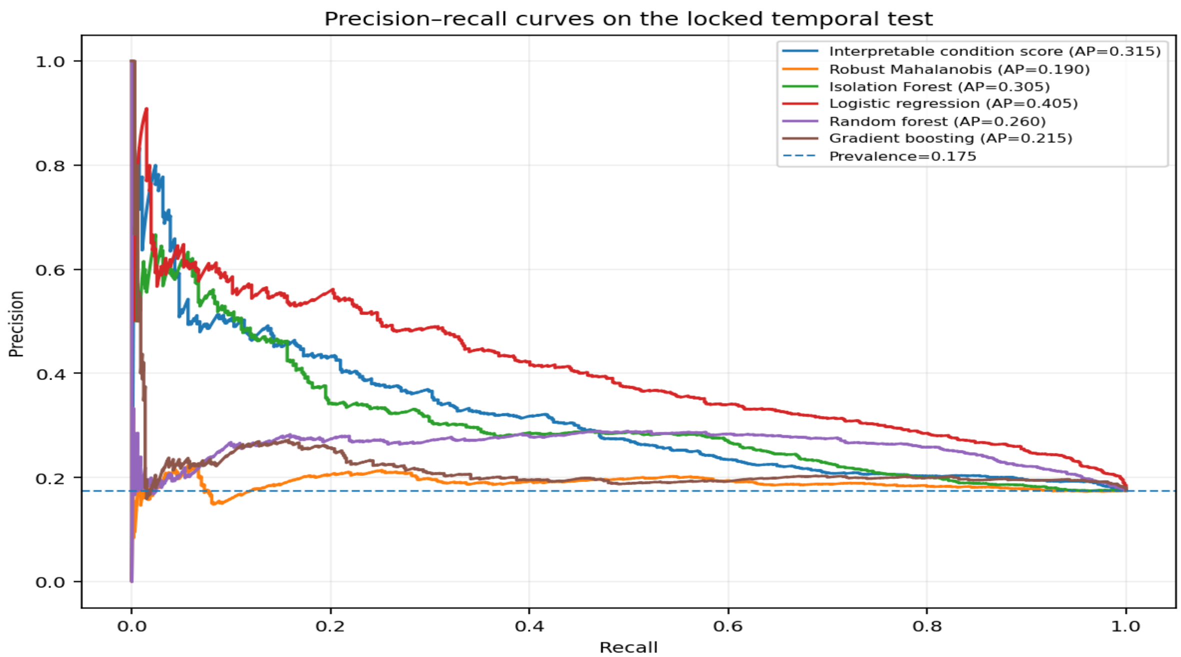

Figure 6. Precision-recall curves on the locked temporal test. AUPRC is emphasized because the positive class is imbalanced.

5.5 Incremental value of feature domains

Feature-layer ablation provides the clearest evidence for the study's central proposition. A load-only logistic model achieved AUROC 0.716 and AUPRC 0.346. Adding thermal variables increased AUROC to 0.742 and AUPRC to 0.366. Power-quality variables added little additional aggregate discrimination in this particular test period, but maintenance variables increased AUPRC to 0.405 and precision at the selected threshold to 0.417 while reducing the alert rate to 16.8%. This is operationally important: maintenance context did not simply increase sensitivity; it helped concentrate positive cases in a smaller review queue.

Feature layer	AUROC	AUPRC	Precision	Recall	F1	Alert rate
Load only	0.716	0.346	0.337	0.476	0.395	24.7%
Load + thermal	0.742	0.366	0.371	0.409	0.389	19.3%
Load + thermal + PQ	0.742	0.366	0.369	0.387	0.378	18.3%
All + maintenance	0.763	0.405	0.417	0.400	0.409	16.8%

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

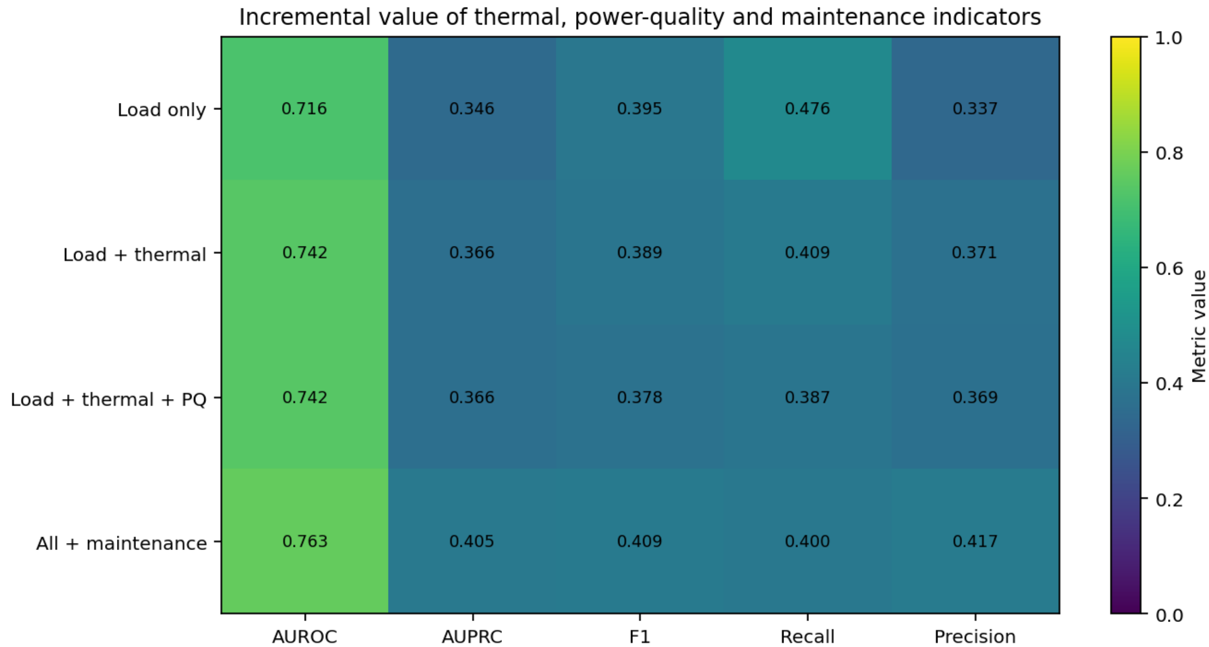

Figure 7. Feature-ablation heat map. Thermal variables add information beyond load, while maintenance evidence produces the largest precision-recall gain.

5.6 Review-capacity analysis

A utility rarely has capacity to inspect every high-scoring 15-minute interval. Ranking performance therefore needs to be translated into a finite review budget. At the top 1% of observations, the interpretable condition score achieved 68.4% precision, compared with 60.5% for logistic regression and 57.9% for Isolation Forest. Logistic regression captured more total event windows as the budget expanded: at the top 10% it captured 27.7% of all positive windows with 48.4% precision. The RHS captured 22.8% with 39.8% precision. These results support a hybrid workflow: a transparent score can generate high-confidence engineer-readable alerts at very small budgets, while a supervised risk model can expand the queue when maintenance capacity permits.

Ranking method	Review budget	Event-window capture	Precision in reviewed set
Interpretable condition score	1%	3.9%	68.4%
Interpretable condition score	5%	13.8%	48.4%
Interpretable condition score	10%	22.8%	39.8%
Isolation Forest	1%	3.3%	57.9%
Isolation Forest	5%	13.4%	46.9%
Isolation Forest	10%	19.8%	34.6%
Logistic regression	1%	3.4%	60.5%
Logistic regression	5%	15.6%	54.7%
Logistic regression	10%	27.7%	48.4%

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

5.7 Model importance and a caution on predictive explanation

Permutation analysis of the random forest ranked load ratio as the most influential variable for average-precision performance, followed by ambient and winding temperature. Several variables showed near-zero or negative permutation importance under the locked test, indicating that they did not transport reliably through the temporal shift in that model. This finding should not be misread as proof that those engineering variables are unimportant. A feature can be physically important yet provide little incremental predictive information once correlated variables and a particular time period are considered. Predictive importance is not causal importance.

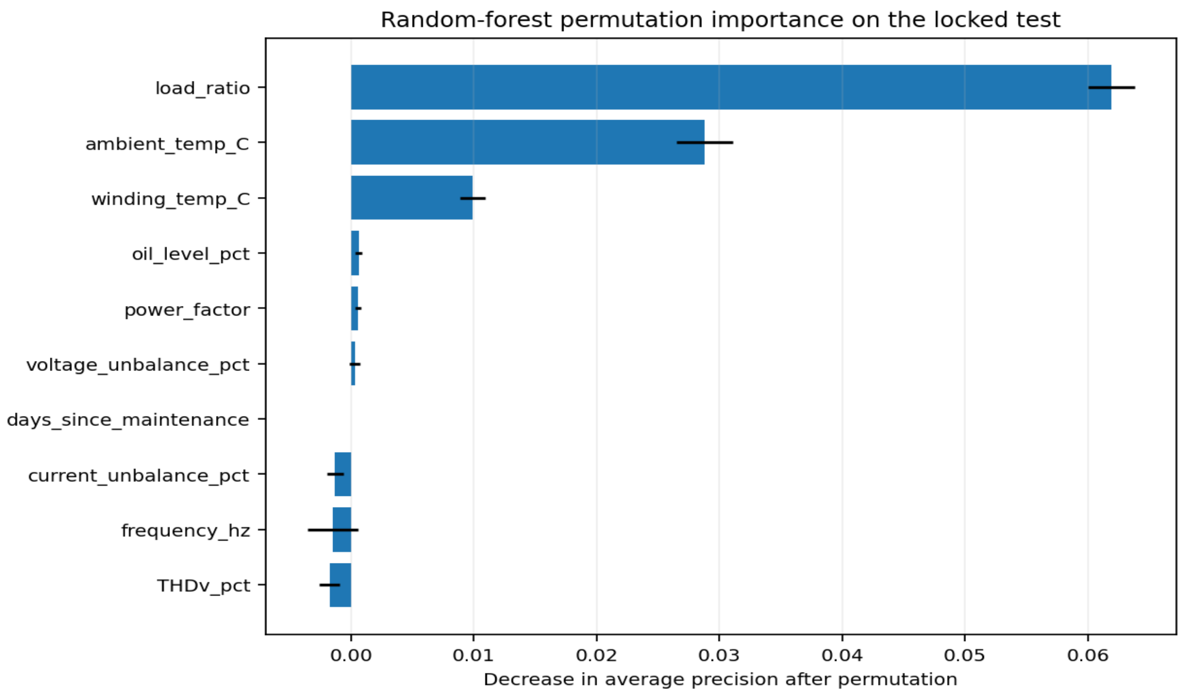

Figure 8. Random-forest permutation importance on the locked test. Negative values indicate that permuting the variable did not degrade test average precision.

5.8 Score-weight sensitivity

Across 250 random +/-20% perturbations of the four RHS domain weights, median locked-test AUROC was 0.646, with the 2.5th-97.5th percentile range 0.634-0.658. The score is therefore not completely dependent on one precise set of weights, but the spread is material enough to justify utility-specific calibration. The correct interpretation is not that the study has discovered universal load/thermal/power-quality/maintenance weights; it has demonstrated a stable scoring architecture whose weights can be re-estimated or governed locally.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

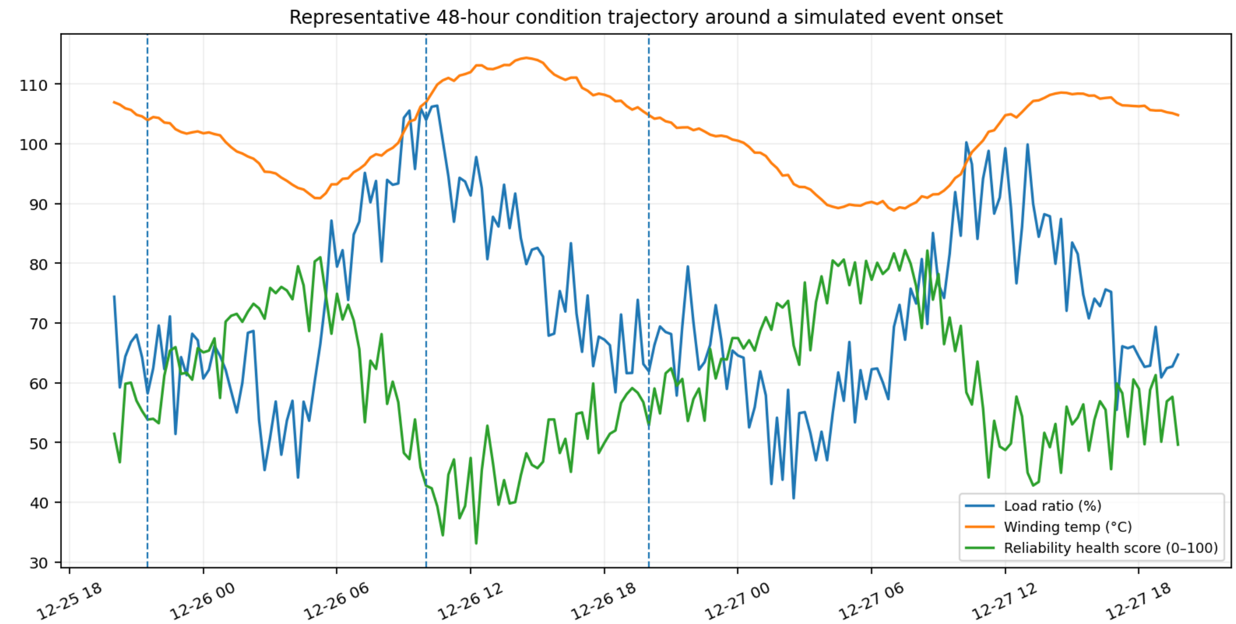

Figure 9. Representative 48-hour trajectory around a controlled event onset. The health score deteriorates as multiple operating stresses accumulate; the vertical marker denotes a simulated event onset.

VI. DISCUSSION

6.1 What the results say about transformer condition assessment

The results support a layered view of condition assessment. Fixed thresholds remain necessary because certain states demand immediate action regardless of statistical rarity. A winding-temperature trip, severe oil-temperature excursion or protection alarm should not be suppressed merely because a machine-learning model assigns a low score. Yet fixed thresholds alone are a weak ranking mechanism when deterioration develops through moderate stress across several domains. The continuous RHS improves that ranking by accumulating smaller penalties and exposing the reason for the score.

The ablation results also show why condition monitoring should extend beyond loading. Load-only analytics had meaningful predictive value, but thermal response added information, and maintenance context improved precision-recall performance further. This is physically and operationally plausible. Load is the forcing variable, temperature is part of the transformer's response, power quality modifies losses and stress, and maintenance history describes whether observed problems have been inspected or remain unresolved. The four domains therefore represent different parts of the same reliability pathway rather than interchangeable predictors.

6.2 Why power-quality variables may add little aggregate discrimination yet remain useful

The power-quality ablation result is deliberately nuanced. Adding power-quality variables after load and thermal features produced almost no further AUROC/AUPRC improvement in the locked period. It would be incorrect to conclude from this that harmonics or poor power factor do not matter. First, their thermal consequences may already be partly reflected in winding and oil temperature. Second, the controlled test contains relatively few severe high-THD observations. Third, aggregate discrimination can hide specific failure mechanisms. Cazacu, Ionita and Petrescu (2018) provide a physical basis for harmonic-related thermal aging, so power-quality measures remain useful as reason codes, root-cause clues and maintenance diagnostics even when their incremental predictive contribution is small in one test window.

6.3 Static alarms versus anomaly detection

Isolation Forest achieved similar ranking performance to the RHS but its validation-selected classification threshold transported poorly, generating a large review queue. This distinction between score quality and threshold stability is operationally important. An anomaly detector can order observations reasonably well while its absolute decision boundary becomes unreliable when load patterns, seasons or maintenance state change. Utilities should therefore monitor score distributions, threshold drift and alert volumes over time, and should avoid interpreting an unsupervised anomaly as a confirmed fault. The anomaly score is best used as an independent challenger to engineer rules and supervised models.

6.4 Why the transparent logistic model won

Logistic regression outperformed the nonlinear tree models on the locked test. That result is scientifically useful because it contradicts a common assumption that higher model capacity automatically produces better condition prediction. A linear model can generalize well when the dominant risk structure is monotonic and when nonlinear models overfit development-period interactions that do not persist. The finding mirrors the methodological discipline highlighted by Choga (2026): richer representations can have value, but they should not displace simpler baselines unless the held-out evidence supports the change. For a utility, the additional benefit of logistic regression is that coefficient direction, calibration and threshold behavior are easier to audit than a large ensemble.

6.5 The value of a hybrid architecture

The study therefore does not recommend choosing one winner and discarding the rest. A robust architecture has distinct layers. Hard limits protect against known dangerous states. The RHS provides interpretable continuous triage and an engineering narrative. A supervised statistical model provides a data-driven second opinion. Isolation Forest watches for unusual multivariate states that were not specified in advance. Where the layers disagree, the disagreement itself is valuable evidence: it can reveal sensor drift, new operating regimes, inadequate thresholds or model decay.

VII. ENGINEERING IMPLEMENTATION FRAMEWORK

7.1 From measurements to maintenance work

A production implementation should operate as a closed reliability loop rather than a dashboard-only project. The loop begins with timestamped data acquisition from approved transformer monitors, SCADA/AMI sources and maintenance systems. Data-quality controls verify sensor freshness, units, phase mapping, missingness and timestamp alignment. Feature logic then calculates load ratio, phase imbalance, thermal rise, harmonic bands, alarm history and maintenance status. The RHS and challenger models rank the observation. Finally, the alert is routed to an engineer with the component reason codes and supporting trend window, and the disposition is captured for learning.

Stage	Minimum control	Evidence retained
1. Acquire	Time-synchronized electrical, thermal and maintenance inputs with asset ID and engineering units.	Raw source reference, timestamp, sensor quality flag and data lineage.
2. Validate	Range, missingness, sensor-staleness, phase mapping and impossible-value checks.	Quality report and rejected/filled-value log.
3. Score	RHS plus at least one independent statistical/model challenger.	Feature version, model version, score, component penalties and threshold.
4. Review	Engineer examines trends, alarms, work orders and asset criticality before action.	Reviewer, reason code, disposition, requested diagnostic test.
5. Act	Inspect, thermograph, sample oil, rebalance load, repair cooling, derate, refurbish or replace as justified.	Work order, test results, intervention date and asset status.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723216

Stage	Minimum control	Evidence retained
6. Learn	Compare subsequent evidence with alert and intervention.	Outcome label, false-alert reason, missed-event analysis and recalibration decision.

7.2 Condition bands and recommended action logic

RHS range	Study interpretation	Default action	Override examples
81-100	Very healthy	Continue routine monitoring and scheduled maintenance.	Protection alarm, severe DGA/PD finding or known critical defect overrides score.
66-80	Healthy	Trend review; no acceleration unless deterioration is persistent.	Rapid temperature rise, recurring PQ stress, known cooling issue.
51-65	Watch	Engineer review; confirm loading, thermal trend, maintenance status and sensor validity.	Asset criticality may justify inspection or thermography.
36-50	Degraded	Prioritized inspection and targeted diagnostics; review loading/derating and pending work.	High-consequence loads may require accelerated outage planning.
0-35	Critical	Immediate engineering assessment and appropriate protection/maintenance decision.	Do not rely on score alone; confirm against protection settings, tests and safe-work procedures.

7.3 Maintenance evidence should be structured, not free text only

A practical implementation should connect condition analytics to the computerized maintenance management system (CMMS). At minimum, work orders should capture asset identifier, date, inspection type, diagnostic test, defect category, severity, corrective action, parts used, closure evidence and next due date. Free-text notes remain useful but should not be the sole source of maintenance history. Structured maintenance evidence allows the score to distinguish an old unresolved alarm from a recently inspected and dispositioned condition. It also creates the outcome labels needed for future field validation.

7.4 Reason codes and engineer-facing explanation

Every high-priority alert should contain a short reason package, for example: “sustained >100% loading + winding temperature >100 C,” “elevated THDv + deteriorating power factor,” “oil-temperature rise inconsistent with ambient/load,” or “moderate thermal stress + maintenance overdue + unresolved defect.” These are not automated diagnoses. They are traceable prompts for engineering review. The distinction is important because causal diagnosis may require oil testing, thermography, partial-discharge testing, protection records, inspection or outage-based examination.

7.5 Validation gates before field use

6. Freeze the intended use: early condition triage for defined transformer classes, not autonomous trip control or replacement authorization.
7. Collect at least one full seasonal cycle of synchronized sensor and CMMS data, preserving both routine and abnormal operating periods.
8. Define event labels through documented engineering review: alarm excursion, confirmed defect, diagnostic-test abnormality, maintenance intervention and actual failure should remain distinct outcome classes.

9. Create development, calibration and later out-of-time test periods; avoid random splitting of adjacent time-series records.
10. Benchmark the RHS, hard thresholds, a transparent statistical model and at least one anomaly detector before considering more complex models.
11. Report PR metrics, alert burden, false negatives by failure mechanism, calibration and threshold stability by season and load regime.
12. Perform asset-level and transformer-design subgroup analysis; do not assume a score calibrated to one cooling mode or rating transports unchanged to another.
13. Pilot in advisory/silent mode before using the score to change inspection or maintenance frequency.
14. Require engineering sign-off for threshold changes, model releases, risk acceptance and asset disposition.
15. Archive data lineage, feature logic, model version, alert rationale, reviewer decision and subsequent outcome for audit and learning.

VIII. RELEVANCE TO DISTRIBUTION-SYSTEM RELIABILITY AND ELECTRIFIED FACILITIES

Distribution networks are changing in ways that make high-frequency condition intelligence more valuable. Electrification increases peak and coincident loads; inverter-based resources and power electronics alter waveform characteristics; distributed generation can change load direction and voltage profiles; and severe weather can create unusual loading and restoration patterns. A transformer condition framework that combines load, temperature and power quality can therefore support both ordinary maintenance and adaptation to new operating regimes. The objective is not to predict every failure mechanism from feeder-side sensors, but to identify assets and periods where deeper engineering attention has the highest expected value.

The same logic extends to electrified industrial facilities. Transformers feeding drives, data centers, charging infrastructure, HVAC plants and automated manufacturing may experience high utilization, nonlinear loads and stringent uptime requirements. An interpretable score can support maintenance scheduling, capital planning and spares prioritization when it is linked to asset criticality and failure consequence. Machekera et al. (2026) similarly emphasize interpretable analytics in energy-intensive infrastructure, while Kabera et al. (2026) connect predictive maintenance to renewable and storage system reliability. These related applications reinforce the wider importance of turning sensor data into auditable engineering decisions rather than isolated model outputs.

IX. LIMITATIONS

The most important limitation is the evidence boundary. The numerical model comparison is based on a controlled benchmark-calibrated panel, not on an adjudicated utility fleet. The maintenance variables are simulated, and the prospective condition-event label represents a severe multi-indicator state rather than an actual failure. Consequently, the reported AUROC, AUPRC, risk ratios and band event rates should not be generalized to U.S. transformer populations, vendor designs, climate zones or failure rates.

Second, the public Kaggle source is primarily an online monitoring dataset and does not include the full diagnostic breadth needed for comprehensive transformer health assessment. DGA, moisture, furan, dielectric tests, partial discharge, FRA, bushing diagnostics, tap-changer condition and protection-event histories are absent. A field version of the RHS should incorporate available high-value diagnostics rather than treating the four present domains as exhaustive.

Third, the study uses a six-hour condition-event horizon to test prospective ranking. Maintenance decisions often operate on multiple horizons: immediate protection, next-shift inspection, seven-day diagnostic testing, annual maintenance planning and multi-year replacement. Field research should estimate horizon-specific models and determine whether the same features and thresholds remain appropriate.

Fourth, alert thresholds were selected on one validation period. Isolation Forest in particular showed sensitivity to threshold transport. A deployed system therefore requires drift monitoring, seasonal calibration and a policy for suspending automated ranking when sensor distributions change materially. Fifth, no causal claim is made that an intervention triggered by the score would reduce failures. That question requires prospective or quasi-experimental field evaluation with documented maintenance actions and comparable controls.

X. FUTURE RESEARCH

The next research phase should replace the controlled maintenance layer with authorized CMMS/work-order history and link high-frequency monitoring to diagnostic-test outcomes. A multi-utility study should include multiple transformer ratings, cooling modes, manufacturers, climates and load compositions. Event definitions should separate thermal overload, cooling failure, insulation deterioration, bushing problems, tap-changer problems, oil leakage, protection operation and external feeder faults. This would allow mechanism-specific recall and false-negative analysis rather than a single composite event label.

A second direction is time-series modeling. Recurrent networks, temporal convolution, transformers and state-space models may capture trajectory information that pointwise models miss. However, more sophisticated architectures should be evaluated against rolling thermal features and regularized statistical baselines. Zemouri (2025) notes both the opportunity and the current limitations of advanced machine learning in transformer prognostics and health management. The strongest research design would therefore preregister baselines, use external temporal validation, report calibration and measure maintenance utility, not merely classification accuracy.

A third direction is uncertainty-aware asset decision support. Health scores could be paired with confidence intervals, probability-of-failure models, consequence of failure, replacement lead time and customer criticality to produce a risk-based maintenance priority. Foros and Istad (2020) illustrate the progression from health index toward risk and remaining lifetime estimation. Cost-sensitive learning is also relevant because false reassurance on a truly degraded high-consequence transformer can be more expensive than a false positive inspection (Taha et al., 2024).

Finally, future work should test whether condition scoring improves real maintenance outcomes: fewer emergency failures, reduced truck rolls, better targeting of oil/thermal diagnostics, lower forced-outage hours, reduced overload exposure and improved asset-life planning. Those outcome measures will determine whether predictive condition assessment creates engineering value beyond attractive dashboards.

XI. CONCLUSION

This study presents a transparent and reproducible framework for predictive condition assessment of distribution transformers using load, thermal, power-quality and maintenance indicators. The evidence supports three main conclusions. First, multi-domain condition assessment is more informative than hard alarms alone because moderate stress can accumulate across domains before a single protective threshold is reached. Second, thermal and maintenance information add measurable value beyond loading alone; in the controlled locked test, maintenance variables produced the strongest precision-recall improvement in the ablation study. Third, model complexity is not inherently valuable. Transparent logistic regression outperformed the nonlinear tree models, while the engineer-readable RHS remained competitive with unsupervised anomaly detection and offered clearer reason codes.

The practical implication is a hybrid maintenance system rather than an autonomous predictor. Hard protection limits remain authoritative. The RHS provides a consistent health narrative. Statistical and anomaly models serve as challengers. Heat maps and capacity metrics prioritize scarce engineering attention. Maintenance decisions remain grounded in professional review, asset criticality, diagnostic testing and applicable standards. With field calibration and longitudinal validation, this architecture can become a reusable bridge between increasingly rich transformer telemetry and reliability-centered maintenance practice.

DOI: https://doi.org/10.64388/IREV10I3-1723216

Declarations

Data and reproducibility. The public data source is the Kaggle Distributed Transformer Monitoring dataset. The numerical experiment reported here uses a separate controlled benchmark-calibrated panel because the public release does not contain a complete maintenance/outcome record. Maintenance variables and prospective condition-event labels are simulated and are not attributed to the Kaggle source. A field replication should preserve the variable definitions, split logic and claims boundary while replacing simulated maintenance variables with authorized utility evidence.

Ethics and safety. No human subjects, personal data or confidential utility records are used in the controlled analysis. This manuscript is a research and decision-support framework, not operating instructions for energized equipment. Inspection, testing, derating, protection changes and maintenance must be performed by qualified personnel under applicable electrical-safety procedures, equipment manuals, utility standards and regulatory requirements.

Conflict of interest. No conflict of interest is declared for this manuscript draft.

REFERENCES

[1] Abu-Elanien, A.E.B. and Salama, M.M.A. (2010) 'Asset management techniques for transformers', *Electric Power Systems Research*, 80(4), pp. 456-464. doi: 10.1016/j.epsr.2009.10.008. ScienceDirect

[2] Arias Velasquez, R.M. and Mejia, J.M. (2018) 'Health Index for Transformer Condition Assessment', *IEEE Latin America Transactions*, 16(12), pp. 2843-2849.

[3] Breiman, L. (2001) 'Random forests', *Machine Learning*, 45, pp. 5-32. doi: 10.1023/A:1010933404324. Springer

[4] Cazacu, E., Ionita, V. and Petrescu, L. (2018) 'Thermal Aging of Power Distribution Transformers Operating under Nonlinear and Balanced Load Conditions', *Advances in Electrical and Electronic Engineering*, 16(1). doi: 10.15598/aeee.v16i1.2701. AEEE

[5] Chandola, V., Banerjee, A. and Kumar, V. (2009) 'Anomaly detection: A survey', *ACM Computing Surveys*, 41(3), Article 15. doi: 10.1145/1541880.1541882. ACM

[6] Chatterjee, A. and Ahmed, B.S. (2022) 'IoT anomaly detection methods and applications: A survey', *Internet of Things*, 19, 100568. ScienceDirect

[7] Choga, C.R. (2026) *Integrating DNA and RNA Multi-Omics for Improved Survival Prediction in Lung Cancer*. Master's thesis. Yeshiva University, New York.

[8] Foros, J. and Istad, M. (2020) 'Health Index, Risk and Remaining Lifetime Estimation of Power Transformers', *IEEE Transactions on Power Delivery*, 35. doi: 10.1109/TPWRD.2020.2972976. IEEE

[9] Friedman, J.H. (2001) 'Greedy function approximation: A gradient boosting machine', *Annals of Statistics*, 29(5), pp. 1189-1232. doi: 10.1214/aos/1013203451. Project Euclid

[10] Ganyani, L., Nare, S., Danda, C., Chingezi, L. and Mupa, M.N. (2026) 'From Tax Mismatches to Revenue Protection: A Data-Driven Framework for Audit Prioritization, Voluntary Compliance, and SME Compliance Education', *Iconic Research and Engineering Journals*, 10(1), pp. 517-530. doi: 10.64388/IREV10I1-1719533. IRE Journals

[11] Ghoneim, S.S.M. and Taha, I.B.M. (2016) 'A new approach of DGA interpretation technique for transformer fault diagnosis', *International Journal of Electrical Power & Energy Systems*, 81, pp. 265-274. doi: 10.1016/j.ijepes.2016.02.018. ScienceDirect

[12] Guo, H. and Guo, L. (2022) 'Health index for power transformer condition assessment based on operation history and test data', *Energy Reports*, 8, pp. 9038-9045. doi: 10.1016/j.egyr.2022.07.041. ScienceDirect

[13] Hagan, E., Akakpo, L., Kasuwa, T.L., Nhemachena, T.L., Taanisa, T., Osman, E.W., Tsveta, T. and Mupa, M.N. (2026a) 'Climate-centric data governance for timber traceability: A GRC framework for decarbonized and audit-ready supply chains', *World Journal of Advanced*

Research and Reviews, 30(2), pp. 353-361. doi: 10.30574/wjarr.2026.30.2.1196. WJARR

[14] Hagan, E., Akakpo, L., Kasuwa, T.L., Nhemachena, T.L., Taanisa, T., Osman, E.W., Tsveta, T. and Mupa, M.N. (2026b) 'Predictive analytics for climate-resilient timber supply chains: Integrating anomaly detection with carbon sequestration and risk scoring', *World Journal of Advanced Research and Reviews*, 30(2). doi: 10.30574/wjarr.2026.30.2.1197. WJARR

[15] Huang, Y.-C. (2012) 'Data mining for oil-insulated power transformers: an advanced literature survey', *WIREs Data Mining and Knowledge Discovery*, 2. doi: 10.1002/widm.1043. Wiley

[16] IEEE (2011) *IEEE Std C57.91-2011: IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step-Voltage Regulators*. New York: IEEE. IEEE

[17] IEEE (2022) *IEEE Std 519-2022: IEEE Standard for Harmonic Control in Electric Power Systems*. New York: IEEE. IEEE

[18] International Electrotechnical Commission (IEC) (2018) *IEC 60076-7: Power transformers - Part 7: Loading guide for mineral-oil-immersed power transformers*. Geneva: IEC. IEC

[19] Islam, N., Khan, R., Das, S.K., Sarker, S.K., Islam, M.M., Akter, M. and Muyeen, S.M. (2023) 'Power transformer health condition evaluation: A deep generative model aided intelligent framework', *Electric Power Systems Research*, 218, 109201. doi: 10.1016/j.epsr.2023.109201. ScienceDirect

[20] Jahromi, A.N., Piercy, R., Cress, S., Service, J.R.R. and Fan, W. (2009) 'An approach to power transformer asset management using health index', *IEEE Electrical Insulation Magazine*, 25(2), pp. 20-34. doi: 10.1109/MEI.2009.4802595. IEEE

[21] Jin, L., Kim, D. and Abu-Siada, A. (2023) 'State-of-the-art review on asset management methodologies for oil-immersed power transformers', *Electric Power Systems Research*, 218, 109194. doi: 10.1016/j.epsr.2023.109194. ScienceDirect

[22] Kabera, S.T., Mupa, M.N., Gezah, F.Y., Nhevera, P.N., Mtemeli, P.S. and Hove, T.L. (2026) 'Leveraging Digital Twin Technology and Predictive Maintenance for Optimizing Solar PV and Battery Storage Systems', *World Journal of Advanced Research and Reviews*, 29(2), pp. 1051-1059. doi: 10.30574/wjarr.2026.29.2.0335. WJARR

[23] Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2008) 'Isolation Forest', *Proceedings of the 8th IEEE International Conference on Data Mining*, pp. 413-422. doi: 10.1109/ICDM.2008.17. IEEE

[24] Machekera, G.N., Mtemeli, P.S., Dongo, M.M., Sajanga, G.A. and Mupa, M.N. (2026) 'An Explainable Deep Learning Framework for Energy-Efficient Water and Sewer Reticulation Infrastructure: Predicting Project Risk, Pumping Demand, and Sustainable Construction Performance', *Iconic Research and Engineering Journals*, 9(12), pp. 3197-3209. doi: 10.64388/IREV9I12-1719256. IRE Journals

[25] Mirowski, P. and LeCun, Y. (2012) 'Statistical machine learning and dissolved gas analysis: a review', *IEEE Transactions on Power Delivery*, 27(4), pp. 1791-1799. doi: 10.1109/TPWRD.2012.2197868. IEEE

[26] Mitiche, I., McGrail, T., Boreham, P., Nesbitt, A. and Morison, G. (2021) 'Data-driven anomaly detection in high-voltage transformer bushings with LSTM auto-encoder', *Sensors*, 21, 7426. MDPI

[27] Murugan, R. and Ramasamy, R. (2019) 'Understanding the power transformer component failures for health index-based maintenance planning in electric utilities', *Engineering Failure Analysis*, 96, pp. 274-288. doi: 10.1016/j.engfailanal.2018.10.011. ScienceDirect

[28] Okebu, D.T.C. and Mupa, M.N. (2026) 'Predictive training analytics for competency-based pilot instruction: Early Detection of Checkride Risk, Procedural Noncompliance and Safety-Relevant Learning Gaps', *World Journal of Advanced Research and Reviews*, 30(3), pp.

1928-1942. doi: 10.30574/wjarr.2026.30.3.1774. WJARR

[29] Putchala, S. (2020) *Distributed Transformer Monitoring* [Dataset]. Kaggle. Available at: https://www.kaggle.com/datasets/sreshta140/ai-transformer-monitoring (Accessed: 26 August 2026). Kaggle

[30] Ramesh, J., Shahriar, S., Al-Ali, A.R., Osman, A. and Shaaban, M.F. (2022) 'Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System', *Energies*, 15(21), 7981. doi: 10.3390/en15217981. MDPI

[31] Rudin, C. (2019) 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nature Machine Intelligence*, 1, pp. 206-215. doi: 10.1038/s42256-019-0048-x. Nature

[32] Saito, T. and Rehmsmeier, M. (2015) 'The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets', *PLOS ONE*, 10(3), e0118432. doi: 10.1371/journal.pone.0118432. PLOS

[33] Singh, S. and Bandyopadhyay, M.N. (2010) 'Dissolved gas analysis technique for incipient fault diagnosis in power transformers: a bibliographic survey', *IEEE Electrical Insulation Magazine*, 26(6), pp. 41-46. doi: 10.1109/MEI.2010.5599978. IEEE

[34] Sun, H.-C., Huang, Y.-C. and Huang, C.-M. (2012) 'A Review of Dissolved Gas Analysis in Power Transformers', *Energy Procedia*, 14, pp. 1220-1225. doi: 10.1016/j.egypro.2011.12.1079. ScienceDirect

[35] Susa, D., Lehtonen, M. and Nordman, H. (2005a) 'Dynamic Thermal Modelling of Power Transformers', *IEEE Transactions on Power Delivery*, 20(1), pp. 197-204. doi: 10.1109/TPWRD.2004.835255. IEEE

[36] Susa, D., Lehtonen, M. and Nordman, H. (2005b) 'Dynamic thermal modeling of distribution transformers', *IEEE Transactions on Power Delivery*, 20(3), pp. 1919-1929. doi: 10.1109/TPWRD.2005.848675. IEEE

[37] Taha, I.B.M. (2023) 'Power Transformers Health Index Enhancement Based on Convolutional Neural Network after Applying Imbalanced-Data Oversampling', *Electronics*, 12(11), 2405. doi: 10.3390/electronics12112405. MDPI

[38] Taha, I.B.M. et al. (2024) 'Power Transformer Health Index Using Cost-Sensitive Learning to Consider the Impact of Misclassification', *IEEE Access*, 12, pp. 191790-191807. doi: 10.1109/ACCESS.2024.3516785.

[39] Zahra, S.T., Imdad, S.K., Khan, S., Khalid, S. and Baig, N.A. (2025) 'Power transformer health index and life span assessment: A comprehensive review of conventional and machine learning based approaches', *Engineering Applications of Artificial Intelligence*, 139, 109474. doi: 10.1016/j.engappai.2024.109474. ScienceDirect

[40] Zemouri, R. (2025) 'Power Transformer Prognostics and Health Management Using Machine Learning: A Review and Future Directions', *Machines*, 13(2), 125. doi: 10.3390/machines13020125. MDPI