

Original Article

Predictive Maintenance of Power Transformers: A Thermodynamic Fuzzy Logic Approach for Severely Imbalanced DGA Data

Jorge Luis Cordova Santos¹, Jezzy James Huaman Rojas²

^{1,2}Department of Mechatronics Engineering, Universidad Continental, Huancayo, Peru.

²Corresponding Author: jhuamanroj@continental.edu.pe

Received: 13 May 2026

Revised: 10 September 2026

Accepted: 16 September 2026

Published: 27 October 2026

Abstract - The ageing power transformer is one of the most important components of an electrical transmission grid, and it is important that diagnostic tools can be developed that can be applied in the field for reliable identification of faults when there is a significant data imbalance. In this paper, a new predictive maintenance model is introduced that combines Reliability, Availability, Maintainability and Safety (RAMS) engineering with Weibull parametric distributions to model combustible gas evolution within the normative limits of the IEEE C57.104-2019 standard. A forced-zero-probability matrix is introduced into the data generation pipeline so that all concentration ratios generated are within the thermodynamic plausibility range, which helps to overcome the thermodynamic implausibility that arises when using conventional oversampling algorithms like SMOTE. A Mamdani Fuzzy Inference System (FIS) was applied to evaluate all 2,000 synthetic DGA profiles, including 60 ground-truth critical fault cases (3.00%) deterministically injected to stress-test diagnostic boundary conditions. The framework achieved a 100% high-priority capture rate: all 60 ground-truth critical cases were classified into intervention-required states (39 as 'Critical' and 21 as 'Poor'), and no false 'Critical' alarm was recorded among the 1,940 non-injected records. Control-surface analysis verified Ethylene (C_2H_4) and Acetylene (C_2H_2) as the main reasons for Health Index appraisal, and directly indicated known high-energy degradation mechanisms. By maintaining complete end-to-end mathematical traceability, this architecture provides grid operators with an interpretable, standard-compliant predictive tool that resolves the operational opacity of black-box models.

Keywords - Dissolved Gas Analysis, Fuzzy logic, Predictive maintenance, RAMS engineering, Reliability modeling.

1. Introduction

Power transformers are among the most important and expensive components of electrical networks. Their condition directly affects the Reliability, Availability, Maintainability, and Safety (RAMS) of the entire grid. Recent studies show that many transformers worldwide are approaching or exceeding their expected service life. As a result, there is a shift away from traditional time-based maintenance toward predictive and condition-based strategies, which are now seen as better options [1, 2].

Dissolved Gas Analysis (DGA) is widely accepted as the best method for early detection of internal faults. It is capable of detecting signs of cellulose breakdown, partial discharges, and high-energy arcing in the oil. Techniques like the Rogers Ratio, Doernenburg ratios, and the Duval Triangle have been used as international standards for many years. However, these methods lean on fixed mathematical thresholds, which can lead to issues. When gas levels are close to these markers, results may conflict, leading to false alarms and confusion [3,

4]. The way gases behave in mineral insulating oil is nonlinear, yet standard diagnostic methods commonly treat it as linear. AI and machine learning models, such as Support Vector Machines and LightGBM, have become popular because they are highly accurate [5]. However, these black-box methods can be hard to interpret. High accuracy alone is not enough if the model's reasoning is unclear. Maintenance engineers need clear explanations for flagged units; even accurate results may not be useful without them. Black-box models also struggle with reliability when the data is imbalanced. Since severe failures are rare, field datasets are often highly imbalanced, which can lead to overfitting along with reduced model trustworthiness when accurate diagnosis matters most [6].

Recent literature highlights that standard oversampling techniques, such as SMOTE, often fail to capture the physical realities of severe transformer faults when applied to highly imbalanced datasets [13]. While advanced deep learning models achieve high accuracy on balanced data, they lack the



thermodynamic interpretability required for critical RAMS interventions. Therefore, the rationale for combining thermodynamic principles with fuzzy logic is to bridge this specific gap: providing a diagnostic tool that handles severe class imbalance deterministically without distorting the underlying physics of gas generation. Rather than using strict boundaries between fault states, a Fuzzy Logic Inference System relies on language-based descriptions and gradually changing membership functions. This method is more like how experienced engineers interpret DGA reports. Fuzzy logic is superior to simple binary classifiers, and it adjusts to complicated real-world scenarios. Since it is not constrained by rigid limits, it breaks the limitations of conventional ratio techniques [7] and is founded on principles of thermodynamics and statistical reliability modelling:

First of all, a Health Index (HI) with a range of 0 to 100 was designed by the Key Gas Approach and the percentile thresholds as stipulated in the IEEE C57.104-2019. The Mamdani fuzzy logic algorithm is used to compute the calculations. Oversampling techniques such as SMOTE were not used since gas mixtures which are not consistent with actual thermodynamic behavior make simulations physically unreliable.

Followed by an observation of 2000 instances were taken to form a mature transformer fleet with a Weibull distribution ($k = 2.0$). The high-class ratio was also preserved to indicate the real percentage rates of severe and minor faults in practice. Failures do not occur in real transformer degradation evenly distributed. Any artificial balancing of the dataset would distort these rates and break the wear patterns of a realistic wear, potentially invalidating the analysis and making it impossible to appreciate the realistic behaviour of a fleet in a useful and general way.

These two contributions address the issue of black-box model ambiguity and the scarcity of reproducible but imbalanced datasets of faults together [8, 9]. This model provides a strong basis of maintenance plans capable of supporting comprehensive assessment. To refer, Section 2 describes the synthesis of data and the fuzzy-logic system, Section 3 outlines the findings and the evaluation of the fleet, Section 4 interprets the findings, and Section 5 outlines the future research and the extended implications.

2. Materials and Methods

It is hard to obtain real records of failures in power transformers. Such data are protected by utilities. Moreover, open repositories do not have trustworthy datasets. Because of that, synthetic data was selected as the most transparent and reasonable one. The datasets were crafted in order to resemble gas properties observed in the ageing fleets of transformers. Each stage was directed by the principles of RAMS to secure the information. This strategy provides controlled variability, complete reproducibility and simulation of rare fault conditions.

2.1. Synthetic Data Generation and Extreme Boundary Stress Testing

A parametric model was applied in order to rebuild the gas accumulation with ageing transformers. The method is in line with Soleimani et al. [10]. Sampling was done using two-parameter Weibull distributions ($k = 2.0$). This is a result of decades of transformer research, which justifies this approach. Weibull distribution reflects the growth of the risk of failure due to the degradation of an insulation [11]. This approach is physically justified by the underlying degradation mechanisms of mineral oil. The concentration of gas in each simulation run was within the ranges of the 90th and 95th percentiles as defined by IEEE C57.104-2019 [12]. The simulated dataset is summarised in Table 1, where gases, attributes of measurements and methods of calculation are listed. This summary will provide an opportunity to review briefly all the methodological aspects. All the outputs were verified with standard thermodynamic limits. This verification ensures that the simulated variables remain constrained within thermodynamically feasible boundaries, preventing the generation of physically impossible degradation states. The parametric distribution function was used to obtain the cumulative gas generation probabilities without any approximations:

$$F(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k} \quad (1)$$

Equation (1) is strictly functional; it establishes a clear physical correlation between the physical ageing process and the real gas concentration. This is quite appropriate in the context of reliability. It aids prioritisation of risk profiling and risk maintenance without additional computing resources [12].

Table 1. Dataset features and derivation methods

Feature Name	Type	Derivation Method
Hydrogen (H ₂)	Proposed	Weibull distribution bounded by IEEE P ₉₀ limits
Methane (CH ₄)	Proposed	Weibull distribution reflecting low-temperature thermal faults
Ethane (C ₂ H ₆)	Proposed	Weibull distribution correlated with CH ₄ progression
Ethylene (C ₂ H ₄)	Proposed	Weibull distribution representing severe overheating
Acetylene (C ₂ H ₂)	Proposed	Right-skewed Weibull distribution with a forced-zero matrix
Health_Index	Calculated	Mamdani Centroid defuzzification output
Condition	Calculated	Categorical label assigned based on Health Index thresholds

However, in real-world power grid operations, transformers rarely reach a 'Critical' condition without prior intervention. This operational reality produces highly imbalanced historical datasets where severe edge cases are nearly nonexistent, limiting the evaluation of diagnostic algorithms. To robustly test the proposed Fuzzy Inference System against severe boundary conditions, a deterministic injection of boundary edge cases was performed. Specifically, 3% of the dataset was deliberately forced into extreme high-concentration values for Acetylene and Ethylene, ensuring the model's reliability is quantifiable even for catastrophic anomalies that traditional datasets fail to capture.

2.2. Class Imbalance Management

Only 3.00% of the simulated records (60 of 2,000) were deterministically injected as critical cases with extreme acetylene and ethylene concentrations. This is a powerful disequilibrium that was purposefully preserved. Augmentation, e.g. SMOTE, was not employed. The processes often give gas mixtures not attainable by any real thermodynamic process [13].

Rather, the imbalance was designed as a part of the simulation. Acetylene (C_2H_2) minority-feature vectors were adjusted to 0.0 ppm with the forced-zero probability matrix. This is done on the basis of physical reasoning rather than statistical convenience. This constraint is a requirement such that classifier behaviour will not change. Engineered statistical noise and physically implausible gas concentration ratios that may degrade model performance are thereby eliminated, consistent with findings reported in the imbalanced fault diagnosis literature [13].

2.3. Thermodynamic Variable Selection

Whether or not a gas was the most significant was not decided by any algorithm. The answer came with physics, namely with the principles of thermodynamics of IEC 60599 [14]. High activation energies are needed to break carbon-carbon bonds in mineral oil. The selection of variables was directed by that physical constraint.

In a modified Key Gas method, Acetylene (C_2H_2 , arcing at temperatures above $700^\circ C$) and Ethylene (C_2H_4 , overheating at temperatures above $300^\circ C$) were assigned the highest inferential weight. The use of feature significance scores was only a check and not the primary basis of decision.

2.4. Fuzzy Logic Predictive Modelling

The nonlinear behaviour of gases which accumulate within transformers is an aspect which must be controlled; it has no exception, and any diagnostic model to the contrary will give results which will be effective on paper but which will fail dismally when applied to the real working conditions. This was the purpose of the Mamdani Fuzzy Inference System developed and composed of five input variables H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2 and only one output that is Health Index. The overlapping trapezoidal and triangular membership functions are used to transform the readings of the raw gas into three plain text condition states, which are: normal, Caution and Critical. Based on such an aggregate production, only a single deterministic value is produced through centroid defuzzification [15], as formalised in Equation (2):

$$HI = \frac{\int \mu(z) \cdot z \, d(z)}{\int \mu(z) \, d(z)} \quad (2)$$

2.5. Justification of Model Choice

Due to the fact that neural networks and Random Forest classifiers can always get a higher raw accuracy numbers - that much is not new in the literature, and it does no good to achieve that by saying so [16], but a number that is not explained by something at all is no diagnosis. Rather, it is something that is black, works on a test set and provides nothing more. Mamdani framework was selected specifically because its arguments can be easily understood at any stage; one can trace all the steps between the input and the output because a simple human being can easily follow the algorithm of another algorithm [15]. The fact that traceability is not a feature that is optional is seen when a maintenance engineer is called upon to give reasons as to why he must take a transformer out of service. And it is all the basis of the determination.

2.6. Complete Fuzzy Model Specification

To ensure full reproducibility of the diagnostic framework, the explicit membership function parameters [a, b, c, d] for all evaluated dissolved gases-Hydrogen (H_2), Methane (CH_4), Ethane (C_2H_6), and Acetylene (C_2H_2)-are tabulated in Table 2. These bounds were mathematically derived using the 90th and 95th percentile historical thresholds established in the IEEE C57.104-2019 standard, mapped into the overlapping fuzzy domains. Triangular functions representing the 'Caution' state are defined where the core parameters converge ($b=c$).

Table 2. Membership function parameters for DGA inputs (ppm)

Input Gas	Fuzzy State	a (Start)	b (Core Start)	c (Core End)	d (End)
H_2	Normal	0	0	80.0	140.0
	Caution	80.0	140.0	140.0	200.0
	Critical	140.0	200.0	300.0	300.0
CH_4	Normal	0	0	90.0	120.0
	Caution	90.0	120.0	120.0	150.0
	Critical	120.0	150.0	200.0	200.0

C ₂ H ₆	Normal	0	0	90.0	120.0
	Caution	90.0	120.0	120.0	150.0
	Critical	120.0	150.0	200.0	200.0
C ₂ H ₂	Normal	0	0	2.0	5.5
	Caution	2.0	5.5	5.5	9.0
	Critical	5.5	9.0	30.0	30.0
C ₂ H ₄	Normal	0	0	50.0	100.0
	Caution	50.0	100.0	100.0	150.0
	Critical	100.0	150.0	300.0	300.0

2.7. Thermodynamic Rule Base Configuration

The system is controlled by a deterministic rule-based matrix and has inferential logic. This expert system is based on a smaller set of five rules, as opposed to the exponential number of rules used in normal combinatorial configurations. This decrease eliminates unreasonable output and only

considers thermodynamically feasible fault scenarios. The antecedent clauses use the logical OR operator (maximum s-norm) to include independent fault triggers, such as a PD versus a high-energy arcing, and the baseline healthy state uses the strict logical AND operator (minimum t-norm) for all gases. A full description of the rule matrix is given in Table 3.

Table 3. Complete optimized fuzzy rule base

Rule #	Antecedents	Consequent (Health Index)
1	IF C ₂ H ₂ is Critical OR C ₂ H ₄ is Critical	Critical
2	IF C ₂ H ₂ is Caution OR C ₂ H ₄ is Caution OR H ₂ is Critical	Poor
3	IF CH ₄ is Critical OR C ₂ H ₆ is Critical	Poor
4	IF H ₂ is Caution OR CH ₄ is Caution OR C ₂ H ₆ is Caution	Tolerable
5	IF H ₂ is Normal AND CH ₄ is Normal AND C ₂ H ₆ is Normal AND C ₂ H ₄ is Normal AND C ₂ H ₂ is Normal	Good

2.8. Algorithmic Pseudocode for Reproducibility

To ensure the reproducibility of the results and outline the computational process of the proposed framework, the core logic of the algorithm is described in detail in Algorithm 1, starting from the generation of synthetic data and ending with the stochastic noise sensitivity assessment.

Algorithm 1. Thermodynamic Fuzzy Diagnostic Pipeline

Input: IEEE C57.104-2019 baseline parameters, Sample size N=2000.

Output: Fleet Health Index (HI) array, Evaluated Condition States.

1. Initialize Weibull parameters (k=2.0, scale derived from 90th percentiles).
2. Generate DGA dataset (X) for H₂, CH₄, C₂H₆, C₂H₄ using Weibull distribution.
3. Inject deterministic edge-cases: Set C₂H₂ to Critical values for 3% of N.
4. Execute Benchmarking (Parallel Branch):
 - Apply SMOTE to training subset of X to balance minority states.
 - Train a LightGBM classifier on a balanced dataset.
5. Initialize Fuzzy Inference System (FIS):
 - Define universe bounds for inputs (X) and output (HI: 0-100).
 - Map overlapping trapezoidal/triangular membership functions (Normal, Caution, Critical).

- Apply 5 optimized rules using logical OR (faults) and AND (healthy) operators.
6. Conduct Sensitivity Test:
 - For each record in X:
 - $X_{noisy} = X + \text{GaussianNoise}(\mu=0, \sigma=0.10)$
 7. Evaluate Final State:
 - Compute Mamdani centroid defuzzification on X_{noisy} to yield the final HI.
 - Classify HI into Good, Tolerable, Poor, or Critical based on predefined thresholds.
 8. Return Processed Dataset with Condition Assessments.

2.9. Sensor Noise Injection Protocol

Field calibration drift and measurement error can occur in the Dissolved Gas Analysis (DGA) sensors. A sensitivity analysis was included in the methodology to assess the robustness of the diagnostic models without needing to rely on deployment data from decades of physical field experiments. In the synthetic set, a factor of Gaussian noise of $\pm 10\%$ was added to all the gas concentration inputs. The difference in the resulting Health Index was then calculated to assess the robustness of the Mamdani centroid defuzzification process when conditions are less than ideal in the field.

2.10. Workflow Overview

A comprehensive modelling pipeline is shown in Figure 1 and is broken down into 3 sequential steps. In Phase 1, Synthetic Data Engineering is used, with IEEE C57.104-2019

and IEC 60599 guidelines, to create an imbalanced DGA set representative for a realistic dataset, and add a deterministic set of ‘critical’ edge-cases to stress-test the model. Phase 2 covers the main Processing stage, where the thermodynamic fuzzy rule base is optimized and assessed using the Mamdani centroid defuzzification technique. At the same time, a benchmark branch uses SMOTE balancing and LightGBM classifier to compare the interpretable expert system with

state-of-the-art black-box machine learning methods. Lastly, in Phase 3, the output of the Condition Assessment (RAMS) is described, where a stochastic sensor noise injection protocol ($\pm 10\%$ variance) is implemented one before the final Health Index evaluation. Repeatability and transparency were two requirements that were seen as necessary for this development process.

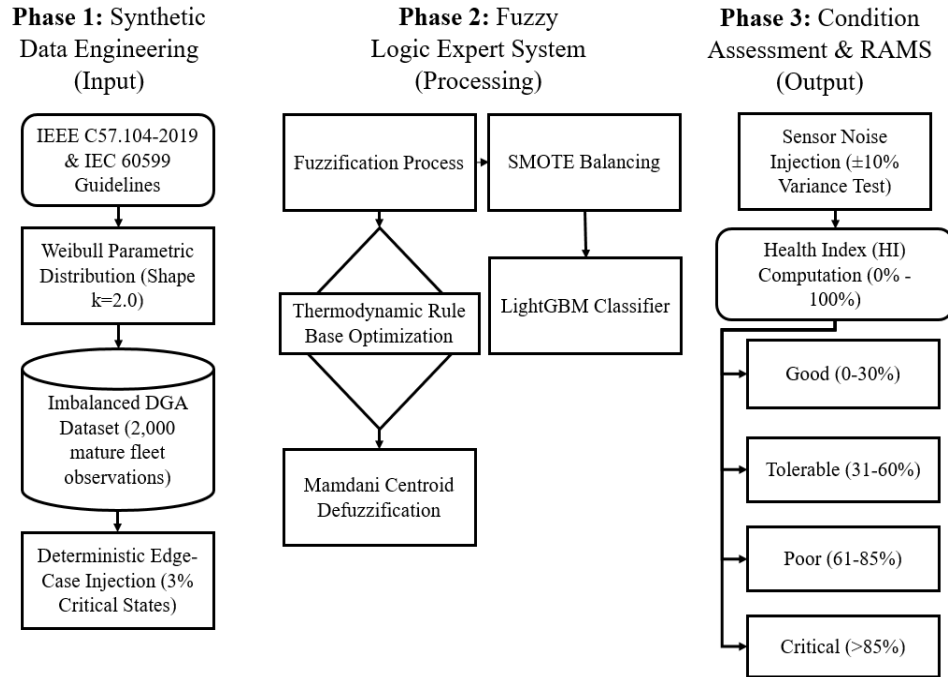


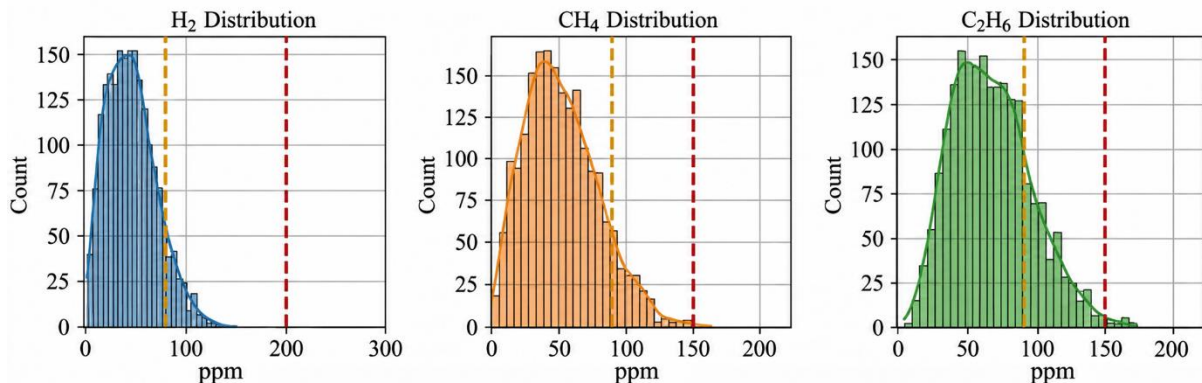
Fig. 1 Mamdani fuzzy logic process

3. Results

3.1. Evaluation of Model Effectiveness

The distribution of each gas follows known thermodynamic accumulation patterns: Hydrogen, methane, ethane, and ethylene dominate below the P90 line, while Acetylene (C_2H_2) follows a zero-inflated distribution, as shown in Figure 2. The scattering of their numbers is a repetition of the kind that would be shown by a slowly

decreasing and industrious fleet. This was not the case for acetylene, however. It was highly skewed in distribution, in such a way that in the vast majority of instances was zero. All faults were not intended to cause such behavior. Instead, the matrix with forced zeros performed what it was designed to do. In this respect, the incidence of high-energy arcs was, in reality, rare in a mature transformer fleet, and the fabricated records were not much different.



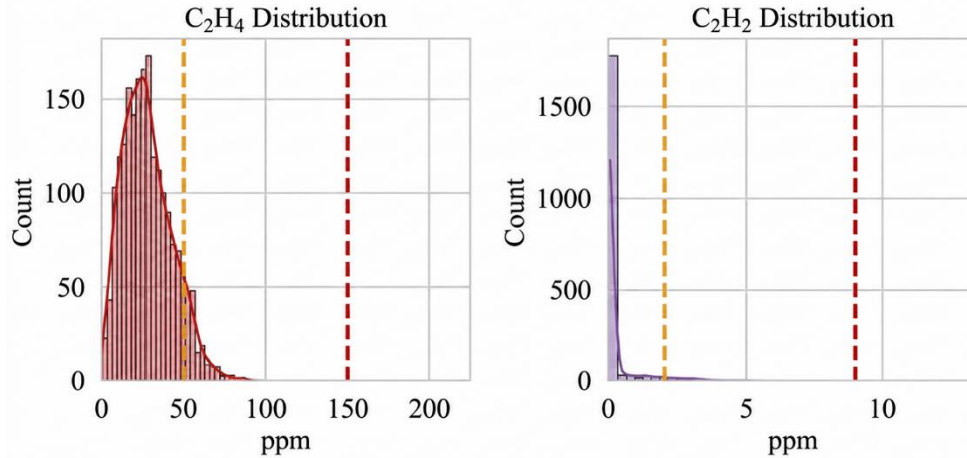


Fig. 2 DGA histograms

3.2. Thermodynamic Feature Correlation

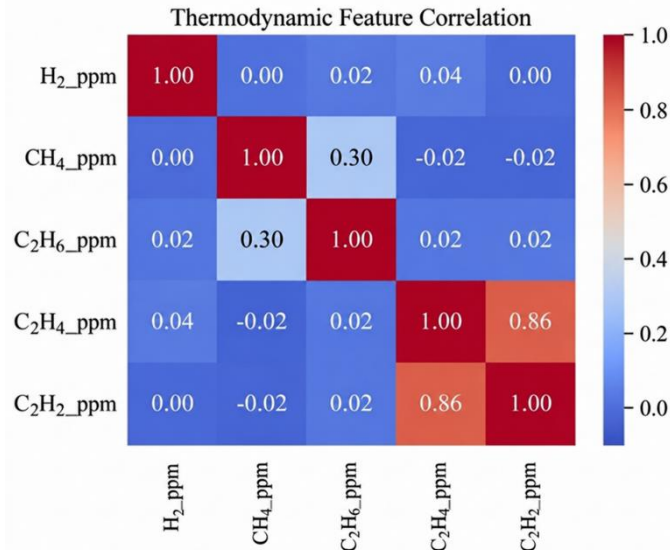


Fig. 3 Thermodynamic correlation heatmap

Two different thermodynamic associations are highlighted in Figure 3, providing support for the physical realism of the synthetic data. A moderate correlation exists between two gases, Methane (CH₄) and Ethane (C₂H₆), with the values of $r = 0.30$, which is in agreement with the fact that both gases are co-generated during low-to-medium temperature thermal degradation. Second, there is a strong positive correlation ($r = 0.86$) that arises between the gases Ethylene (C₂H₄) and Acetylene (C₂H₂).

In Dissolved Gas Analysis (DGA) physics, acetylene does not usually appear alone: the mineral oil in transformers decomposes into ethylene and acetylene at the same time, when very deep arcs occur, and/or high-energy thermal discharge (thermal stress $> 700^{\circ}\text{C}$) takes place. The appearance of these two particular couplings is proof that the dataset correctly reflects true transformer failure characteristics. Most importantly, all other remaining gas pairs

show near-zero correlations ($|r| = 0.04$), which means that the unrelatedness of fault mechanisms is statistically meaningful and classification models will not be able to take advantage of any shortcuts.

3.3. Fuzzy Membership Functions Mapping

The membership functions for Ethylene (C₂H₄) are shown as an example of the proposed fuzzification process in Figure 4. Concentrations are divided into three linguistic subsets, Normal, Caution and Critical, based on selected thresholds of the IEEE C57.104 (P90=50 ppm, P95=150 ppm). Adjoining sets can be controlled so that the membership degree (μ) can be varied continuously. This intentional overlap ensures that there remain uncertainties across transitional areas that do not have sudden jumps between states, as seen in sharp binary boundaries. As a result, the fuzzy framework models thermal degradation as a continuous process, as opposed to using step functions.

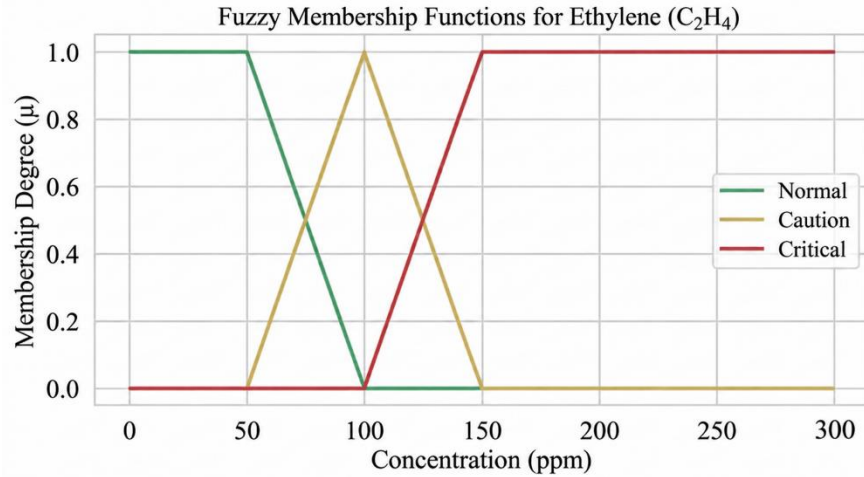


Fig. 4 Ethylene (C_2H_4) fuzzy membership

3.4. Fuzzy Rule Base Control Surface

The Mamdani inference engine dynamic behavior is displayed in the result Figure 5, that shows the resulting Health Index as a function of the concentrations of Ethylene (C_2H_4) and Acetylene (C_2H_2). The answer should be a smooth non-linear control surface with no sharp step discontinuities or artificial classification boundaries. Uncalibrated membership overlaps and conflicting rule bases, which in fuzzy inference systems are usually interpreted as erratic surface topographies, are here supported by the monotonic continuity, which indicates well-harmonized interactions between the rules and appropriate universe partitioning.

In terms of mechanism, the control surface shows the physical control of high-energy fault indicators. The Health Index quickly moves towards the critical degradation zone ($>80\%$) as the acetylene concentration increases, even slightly above 2.0ppm, showing the high diagnostic weight given to low temperature thermal faults. At the same time, the Health Index increases in a proportional manner as the concentration of ethylene increases. This continuous multi-variable surface confirms the rule base is a correct model of the non-linear thermodynamic interactions without the creation of artificial classification boundaries.

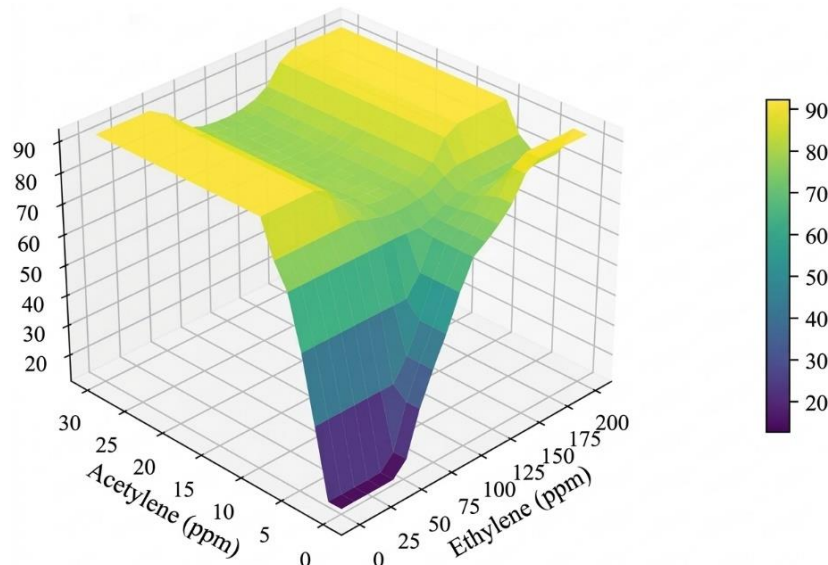


Fig. 5 Mamdani 3D control surface

3.5. Defuzzification and Centre of Gravity Analysis

If multiple rule premises are satisfied as a result of the concurrent anomalies in the gas concentrations, then the Mamdani engine adds together the active consequent fuzzy sets to form a composite output fuzzy set. Figure 6 illustrates

this defuzzification process for a multi-fault evaluation scenario ($H_2=150$, $CH_4=100$, $C_2H_6=95$, $C_2H_4=120$, $C_2H_2=5.5$ ppm). Implication weighting is represented by the truncated areas under the membership profiles of Tolerable, Poor and Critical.

In this representative case, the centroid evaluation gives the score $HI=60.0\%$, putting the asset exactly on the boundary between the 'Tolerable' and 'Poor' states. This defuzzification process ensures complete mathematical traceability and model

interpretability. The scalar Health Index can be directly connected to individual membership activation, which provides a clear physical reason for planning Asset De-energization or oil treatment.

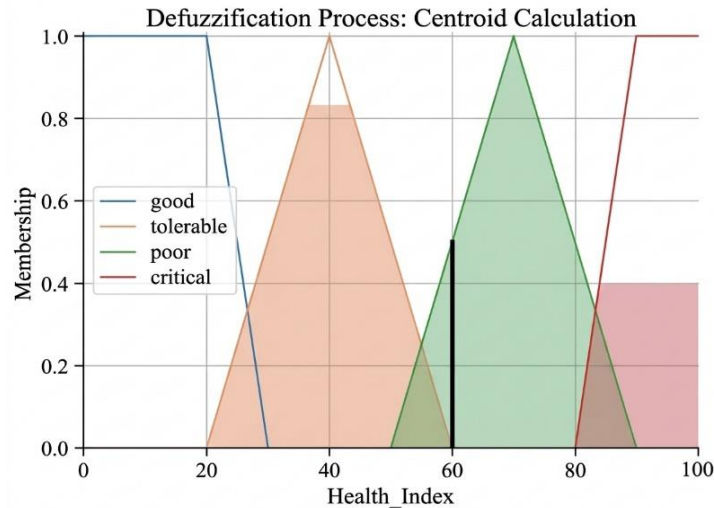


Fig. 6 Defuzzification centroid calculation

3.6. Simulated Fleet Condition Assessment

Figure 7 shows the distribution of the evaluated Health Index in the entire fleet of simulated power transformers, $n = 2,000$. The diagnostic classification categorizes 1,626 units (81.30%) as 'Good', 295 units (14.75%) as 'Tolerable', 40 units (2.00%) as 'Poor', and 39 units (1.95%) as 'Critical'. The distribution is highly non-normal and closely reflects the actual power transmission and distribution operating environment, where the vast majority of energized assets are in safe thermodynamic condition, and high degradation states are rare, but catastrophic, tail events. A combined 3.95% of the classes are degraded or are in 'Critical' condition, resulting

in a difficult class-imbalance situation that often leads to machine learning algorithms performing poorly on the minority classes, due to the majority class bias. On the other hand, this natural imbalance is dealt with intrinsically by the Mamdani fuzzy framework, driven by experts. The system did not classify any of the 1,940 non-injected records as 'Critical', which avoids unnecessary emergency interventions on non-critical assets. This also enables the identification of 39 of the most critical units, in a diagnostic manner, that need immediate capital intervention and provides a grid operator with a tool for rapid and cost-effective decision making in order to mitigate risks across the fleet.

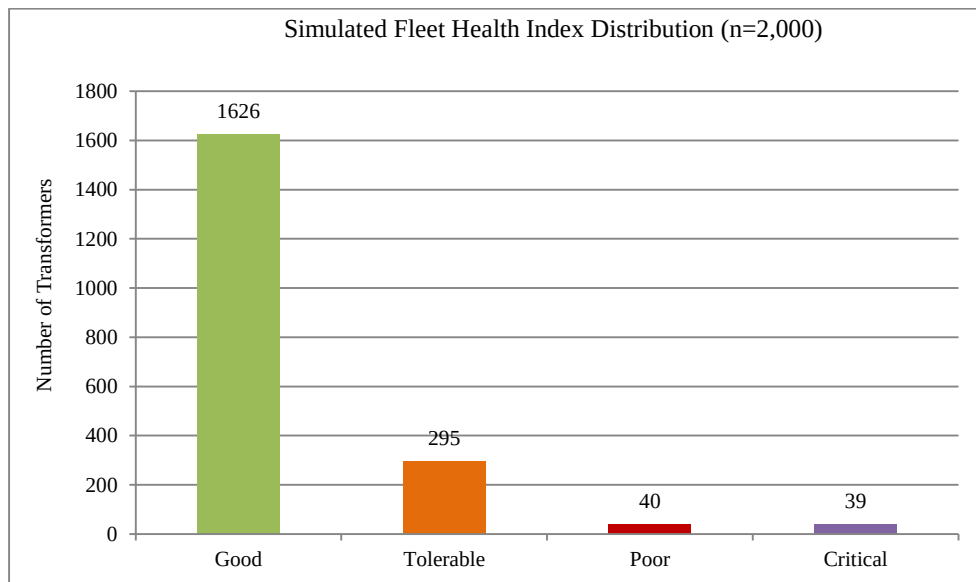


Fig. 7 Health Index categorization

3.7. Comparative Diagnostic Performance

A comparative benchmark of the proposed fuzzy framework with the modern black-box architecture was carried out using a LightGBM classifier to quantify the diagnostic fidelity of the proposed fuzzy framework. As shown in Figure 8, the LightGBM model was trained on a SMOTE-generated balanced test partition (n=600); it learned the strict mathematical thresholds of the IEEE ground truth very carefully, with hard boundaries between all condition states. The Mamdani Fuzzy Inference System, on the other hand, exhibits a smoothing behavior over the entire n=2,000 data set, showing a multi-criteria approach. In cases where the single-secondary gas value just failed to meet strict IEEE limits (termed 'Tolerable' and 'Poor' by IEEE single-parameter rules) the fuzzy system indicated a 'Good' Health Index in 350 and 120 cases, respectively (see Figure 8).

This is not a classification error; it's an output of centroid defuzzification that involves all 5 gases – a slight increase in any non-key gas does not disproportionately affect the overall fleet health score when the rest of the thermodynamic gases are in pristine condition.

Most importantly, the fuzzy inference engine is free from any critical false negative in critical failure modes. Of the 60 ground-truth 'Critical' instances, the fuzzy system classified 39 as 'Critical', and 21 as 'Poor', representing 100% of high-energy arcing and extreme thermal anomaly instances as captured in high-priority intervention states. This multi-variable integration eliminates unnecessary maintenance dispatches for minor gas fluctuations and ensures 100% security of operation for severe asset degradation.

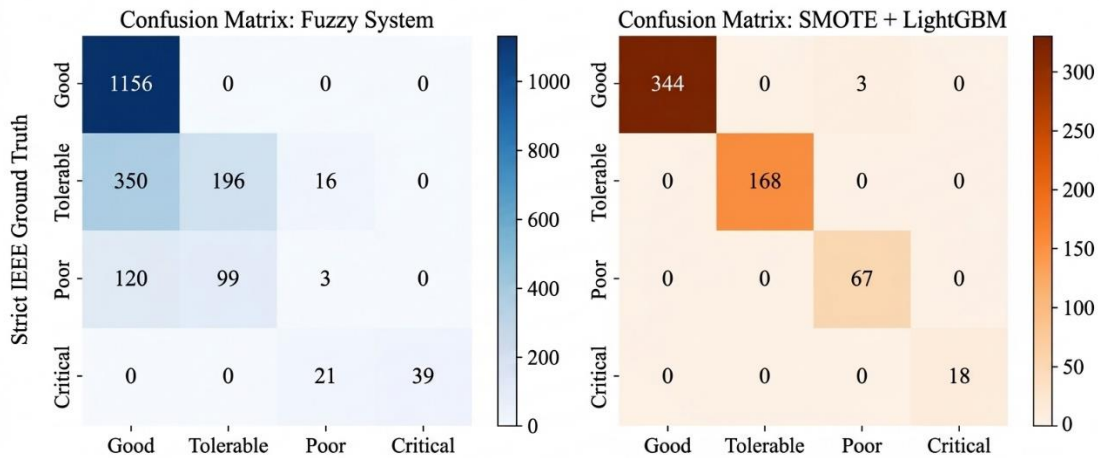


Fig. 8 Confusion matrices comparing the fuzzy system (left) and LightGBM benchmark (right)

3.8. Critical State Detection and ROC Evaluation

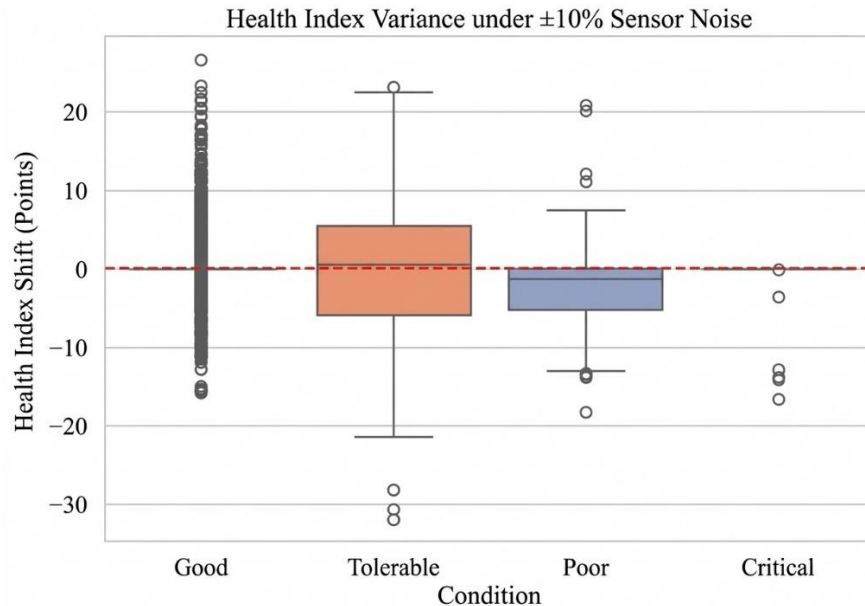


Fig. 9 ROC Curve evaluating the detection of the 'Critical' operational state

The ability of both methods to correctly identify extreme fault scenarios is tested through Receiver Operating Characteristic (ROC) analysis as illustrated in Figure 9. For the purpose of the binary discrimination performance, the 'Critical' state was measured against all other states of operation.

As shown in Figure 9, the LightGBM classifier performance is AUC=1.000, and the proposed Fuzzy Inference System has a near-perfect discriminative capability (AUC=0.999). An empirical AUC near one often calls into question the tendency of a model to overfit. But in this stress test around the boundaries, the result proves the mathematical soundness of the Mamdani rule base. The small difference (0.001) in the AUC of the fuzzy model is due to the soft trapezoidal transitions at near P95 boundary limits, where severe thermal overheating ($C_2H_4 > 150\text{ppm}$) occurs without any spikes in acetylene, so that the health indices are gradually shifting between 'Poor' and 'Critical'. This 0.999 AUC validates the fuzzy engine's ability to deliver the best possible sensitivity for High Energy Arcing (C_2H_2) and Extreme Thermal Faults while maintaining fleet safety and minimising false-positive noise in priority maintenance queues.

3.9. Diagnostic Robustness under Sensor Noise

The gas concentration at the input to the model was perturbed with zero-mean Gaussian noise to simulate a calibration error of $\pm 10\%$ of the equipment to assess the model stability under field conditions. The resulting variance in the

calculated Health Index (ΔHI) is shown for each baseline condition state in Figure 10.

The observed output variance is directly representative of the underlying fuzzy set geometry, and it is determined that there is a high operational stability:

- **Critical State Saturation:** The boxplot for the 'Critical' condition has a very low range, with the values mostly concentrated around $\Delta HI = 0$. The high-energy fault indicators (C_2H_2 spikes) take the system to the top of the upper plateau ($\mu = 1.0$) of the critical trapezoidal functions, so a $\pm 10\%$ input fluctuation won't affect the defuzzified centroid. This ensures total fault-tolerance in case of severe degradation, even if there were minor sensor drift.
- **Continuous Uncertainty Absorption:** The interquartile ranges for the 'Tolerable' and 'Poor' categories are larger, ranging between about -6 and +6 Health Index points. This is because the gas concentrations are found on the sloped-overlap zone between the neighboring membership sets. The centroid engine does not trigger arbitrary category jumps, which is otherwise a common issue in IEEE approach based on thresholds, but rather provides a continuous, proportional shift in score based on sensor uncertainty.
- **Baseline Fleet Stability:** In 'Good' condition, the assets have a median shift of zero - which means that they are not affected by false alarm signals and retain their critical early-warning functions.

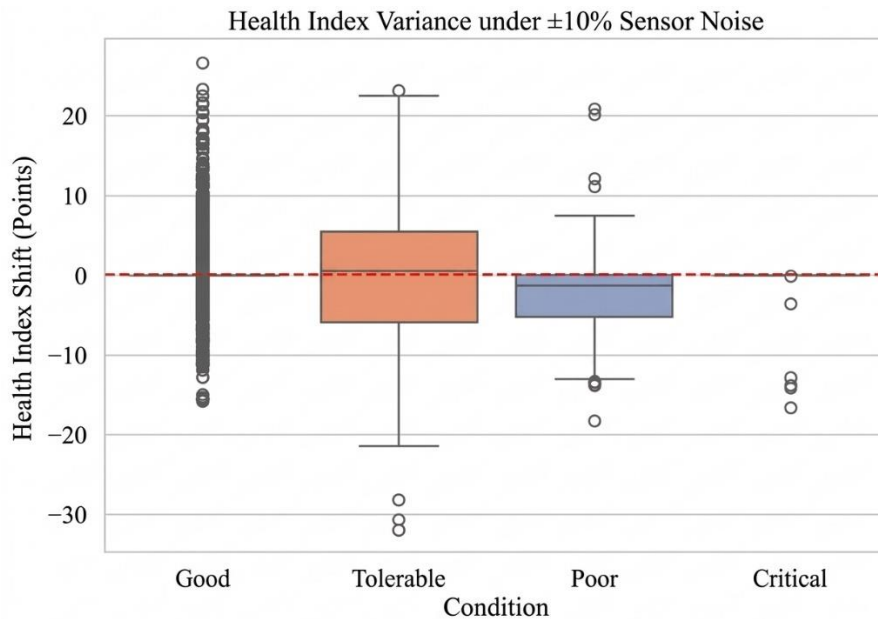


Fig. 10 Boxplot illustrating Health Index shift under $\pm 10\%$ injected sensor noise across different condition states

4. Discussion

The Mamdani Fuzzy Inference System (FIS) has high diagnostic capability based on the evaluation of 2,000

simulated Dissolved Gas Analysis (DGA) cases within Reliability, Availability, Maintainability, and Safety (RAMS) and IEEE C57.104-2019 parameters. Most importantly, none

of the workflow steps involved in the analysis involved any synthetic oversampling. The main benefit of this approach is that it avoids the use of composite resampling methods, such as Gaussian-Enhanced SMOTE (GESMOTE) or SMOTE-ENN [6, 13] that are often used to introduce gas concentration ratios that are not physically feasible and do not follow the principle of thermodynamic equilibrium [24].

Another important point of the structure is the deliberate use of synthetic data with the parameters defined by IEEE C57.104-2019 to test a rule-based expert system based on the same standards. This methodology was chosen because it is a necessary and deterministic stress test of the system, while evaluating with standard-constrained data provides a theoretical circular validation loop. The existing number of successful distributions of real-world utilities is simply not enough to thoroughly test and validate algorithm boundaries for catastrophic, high-energy arcing. This phase, therefore, is used to ensure the mathematical traceability, boundary logic and fault tolerance of the Mamdani centroid defuzzification process before its deployment in the field under extreme condition thresholds such as forced 3% critical fault injections.

Control-surface evaluation identified Ethylene (C_2H_4) and Acetylene (C_2H_2) as the major conditioning factors, and these had a significant impact on the diagnostic output compared to all other measured parameters. The above results are consistent with literature-reported studies that confirm that these particular gases are very important indicators of serious thermal and electrical deterioration [12, 14]. Moreover, the result here fills the structural void in historical logs as pointed out by Liu and Yang [4] and Mahmoud et al. [16], that few severe failures occur in real operational logs, making it difficult to use data-driven training. This paper overcomes the lack of empirical fault data by using Weibull simulation as a means of creating a controlled environment for the study of extreme degradation states [11] and directly addresses the issue of reproducibility raised in the predictive maintenance literature [5, 16].

This method presents a clear trade-off over purely based-on-the-data methods between raw classification metrics and model transparency. Tuned LightGBM architectures [4] with high accuracy, self-organizing maps, and ensemble classifiers [23] deliver black-box results with hard-to-interpret outputs, which are not easily understood by maintenance staff. Likewise, most sophisticated Health Index models [18] do not include evaluations under rare-event operating conditions. Compared to this, the benefits of using RAMS-directed Weibull simulation in conjunction with centroid defuzzification are that each inferential step is explicit, mathematically derivable, and fully auditable and hence restores operational accountability of high-stakes maintenance decisions [7]. While these benefits highlight the potential advantages in the use of synthetic data, the limitations of

synthetic data must also be properly recognized in order that the long-term validity of the use of synthetic data can be maintained. Fine-scale temporal dynamics in gas evolution, operational noise and sensor calibration drift over long service lifetimes are not reproducible in synthetic simulations. In order to address these fidelity gaps in future work, four separate measures are planned: (1) collecting data on in-service power transformers over time and creating a time-series DGA dataset; (2) coupling the Fuzzy Inference System with Recurrent Neural Networks or Long Short-Term Memory (LSTM) models to capture dynamic gas generation; (3) explicitly integrating real time weather conditions and electrical loading variations into inputs to the inference models; and (4) conducting a full field validation using IoT-enabled DGA sensors deployed across utility substations [17, 19-21]. These limitations do not invalidate the fundamental findings of this study. Mamdani Fuzzy Inference Systems offer a highly interpretable, standard-compliant and reliable solution for the predictive maintenance of a power transformer [25]. In terms of actual implementation, the deterministic rule base is also simple to compute and has a great deal of scalability compared to the resource-intensive deep learning architecture. The framework can be implemented directly in edge-computing diagnostics units at the substation or can be used in utility SCADA systems to monitor the conditions of power grid fleets, in real-time and with low latency, while eliminating the need for extensive cloud computing infrastructure and increasing overall system resilience [22].

5. Conclusion

A set of 2,000 synthetic Dissolved Gas Analysis records was evaluated, generated in compliance with Reliability, Availability, Maintainability, and Safety (RAMS) parameters, and a set of key diagnostic results was obtained. The Mamdani Fuzzy Inference System (FIS) classified all 60 injected critical cases, which accounted for 3.00% of the 2,000 data points (60/2,000), into intervention-required states (39 as 'Critical' and 21 as 'Poor'), and none of the 1,940 non-injected records was classified as 'Critical'. Among the non-injected records, 585 received a Health Index class different from the strict IEEE class (Figure 8), which reflects the multi-gas integration of the centroid defuzzification. Most importantly, the diagnostic accuracy could be obtained without the inclusion of synthetic oversampling, and thus without altering the thermodynamic and physical validity of the underlying gas profiles. In addition, control surface evaluation confirmed Acetylene (C_2H_2) and Ethylene (C_2H_4) as the main contributors to changes in the Health Index, which indicates the physical degradation mechanisms of the transformer's insulation system. One main methodological benefit of this approach was the forced-zero probability matrix introduced in the Weibull parametric generator that did not require any non-physical resampling methods such as SMOTE. As a result, all of the constructed gas concentration ratios are thermodynamically feasible. Furthermore, centroid-based defuzzification keeps all the inferential steps of the system

mathematically traceable from the input data to the final Health Index to help maintenance engineers to trace every inference of the system. Opaque machine learning architectures can achieve high predictive scores, but not explainable decision chains. The Mamdani model proposed directly addresses this transparency problem by providing an auditable and standard-compliant diagnostic model for the safety-critical operations of utilities. The resulting framework can be used by electric utilities as a true instrument for scheduling condition-based maintenance, reducing unplanned outage costs and increasing power transformer service life. A key drawback of this study is the use of synthetic data based on a standard, rather than real-world field data. Longitudinal degradation dynamics, time-dependent sensor calibration drifts, and ambient noise, as generally encountered in real-life situations, are not completely modeled in parametric simulations. Future research should validate the model using longitudinal utility datasets with high energy failure events, to make this architecture from a validated theoretical baseline into a fully operational tool. In addition, using real-time DGA sensors powered by IoT, combined with hybrid time-series forecasting models, will allow continuous adaptive condition monitoring of power grid assets.

Conflicts of Interest

The authors state that they do not have any conflict of interest in the publication of this paper.

References

- [1] Ryad Zemouri, "Power Transformer Prognostics and Health Management using Machine Learning: A Review and Future Directions," *Machines*, vol. 13, no. 2, pp. 1-29, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Wei Huang et al., "Modeling the Aging-Dependent Reliability of Transformers Considering the Individualized Aging Threshold and Lifetime," *IEEE Transactions on Power Delivery*, vol. 37, no. 6, pp. 4631-4645, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Aman Samson Mogos, Xiaodong Liang, and C.Y. Chung, "Enhancing Transformer Health Index Prediction using Dissolved Gas Analysis Data Through Integration of LightGBM and Robust EM Algorithms," *IEEE Access*, vol. 12, pp. 108472-108483, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Cheng Liu, and Weiming Yang, "Transformer Fault Diagnosis using Machine Learning: A Method Combining SHAP Feature Selection and Intelligent Optimization of LGBM," *Energy Informatics*, vol. 8, no. 1, pp. 1-19, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Yan Zhang et al., "Fault Diagnosis of Transformer using Artificial Intelligence: A Review," *Frontiers in Energy Research*, vol. 10, pp. 1-10, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Youngjin Han, and Inwhae Joe, "Enhancing Machine Learning Models through PCA, SMOTE-ENN, and Stochastic Weighted Averaging," *Applied Sciences*, vol. 14, no. 21, pp. 1-22, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Muhammad Farhan Naeem et al., "A Novel Method for Life Estimation of Power Transformers using Fuzzy Logic Systems: An Intelligent Predictive Maintenance Approach," *Frontiers in Energy Research*, vol. 10, pp. 1-20, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Vasiliki Vita et al., "Predictive Maintenance for Distribution System Operators in Increasing Transformers' Reliability," *Electronics*, vol. 12, no. 6, pp. 1-23, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Maryam Assafo et al., "On the Stability and Homogeneous Ensemble of Feature Selection for Predictive Maintenance: A Classification Application for Tool Condition Monitoring in Milling," *Sensors*, vol. 23, no. 9, pp. 1-39, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Milad Soleimani, Mladen Kezunovic, and Sergiy Butenko, "Linear Arrhenius-Weibull Model for Power Transformer Thermal Stress Assessment," *IEEE Access*, vol. 10, pp. 20023-20032, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Muhammad Ilham Chaniago, P. Purwiharjono, and Rudy Gianto, "Transformer Reliability Analysis using the Weibull Distribution Method (Case Study of the Power Transformer at the Kota Baru Pontianak Substation)," *Telecommunications, Computers, and Electricals Engineering Journal (TELECTRICAL)*, vol. 2, no. 2, pp. 247-259, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

Funding Statement

No external sources of funds funded this study. The authors carried out the job on their own.

Ethical Aspects / Data Privacy

This study did not use actual operational utility logs, proprietary identifiers for assets, or private power grid infrastructure data. The full set of 2000 Dissolved Gas Analysis (DGA) profiles were synthetically generated based on Weibull parameter distributions, where IEEE C57.104-2019 norm thresholds were only used as bounds. The sole use of standard bounded synthetic simulations serves as an inbuilt data privacy and security guard, without the danger of revealing any sensitive operational telemetry or proprietary asset performance metrics.

Thus, this structure ensures absolute data confidentiality, complete ethical compliance, and algorithmic reproducibility end-to-end without the need to access limited or sensitive utility information.

Acknowledgments

The authors thank their family members deeply because they have continued to offer their support during this research undertaking. The authors give credit to the divine providence of having the wisdom to do this work. Author 2 and Author 1 worked equally on this work.

- [12] IEEE Std C57.104-2019, IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers, IEEE Standards Association 2019. [Online]. Available: <https://standards.ieee.org/ieee/C57.104/7476/>
- [13] Yan Huo, Diyan Guan, and Lingyan Dong, "Intelligent Fault Diagnosis of Unbalanced Samples using Optimized Generative Adversarial Network," *Applied Sciences*, vol. 14, no. 11, pp. 1-19, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Sahar R. Al-Sakini et al., "Dissolved Gas Analysis for Fault Prediction in Power Transformers using Machine Learning Techniques," *Applied Sciences*, vol. 15, no. 1, pp. 1-23, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Shuaibing Li et al., "Review of Transformer Health Index from the Perspective of Survivability and Condition Assessment," *Electronics*, vol. 12, no. 11, pp. 1-24, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Moamin A. Mahmoud et al., "The Current State of the Art in Research on Predictive Maintenance in Smart Grid Distribution Network: Fault's Types, Causes and Prediction Methods - A Systematic Review," *Energies*, vol. 14, no. 16, pp. 1-27, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Xinyao Li et al., "Domain Adaptive Remaining Useful Life Prediction with Transformer," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-13, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Muhammad Akmal A. Putra, Suwarno, and Rahman Azis Prasajo, "Improving Transformer Health Index Prediction Performance using Machine Learning Algorithms with a Synthetic Minority Oversampling Technique," *Energies*, vol. 18, no. 9, pp. 1-44, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Md. Nuruzzaman et al., "Predictive Maintenance in Power Transformers: A Systematic Review of AI and IoT Applications," *ASRC Procedia: Global Perspectives in Science and Scholarship*, vol. 1, no. 1, pp. 34-47, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Samuel Lessinger et al., "Aging Assessment of Power Transformers with Data Science," *Energies*, vol. 18, no. 15, pp. 1-33, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Eun Hui Ko, Tatjana Dokic, and Mladen Kezunovic, "Prediction Model for the Distribution Transformer Failure using Correlation of Weather Data," *5th International Colloquium on Transformer Research and Asset Management*, Springer, Singapore, vol. 671, pp. 135-144, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Zhaoyi Xu, and Joseph Homer Saleh, "Machine Learning for Reliability Engineering and Safety Applications: Review of Current Status and Future Opportunities," *Reliability Engineering and System Safety*, vol. 211, pp. 1-16, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Siqi Liu, Zhiyuan Xie, and Zhengwei Hu, "DGA-based Fault Diagnosis using Self-Organizing Neural Networks with Incremental Learning," *Electronics*, vol. 14, no. 3, pp. 1-17, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Muhammad Mujahid et al., "Data Oversampling and Imbalanced Datasets: An Investigation of Performance for Machine Learning and Feature Engineering," *Journal of Big Data*, vol. 11, no. 1, pp. 1-32, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] P.C.K. Guri, and Krste Najdenkoski, "Transformer Remaining Life Calculation based on Degree of Polymerization (DP)," *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, vol. 4, no. 11, pp. 16779-16786, 2024. [[Google Scholar](#)]