African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 5, Issue 2, 2026
Page No: 107-121
Website: *https://aaasjournals.com/index.php/ajapas/index*
ISI 2025: 1.126 **SJIFactor 2024: 6.752** **معامل التأثير العربي: 1.62**

Time-Validated Latent Fault Detection in Gas Turbines Using Physics-Based Features and Interpretable Decision Trees

Montaser Ali Saeed [1 *], Abdelgader Agilah Saleh Gheidan [2], Salima Rajab Dakheel [3], Saad Adam [4,] Tarik Hassan Elsonni [5]
[1] Aeronautical Engineering Department, Faculty of Aeronautics, Bright Star University, El Brega, Libya
[2] Department of Mechanical Engineering, Faculty of Technical Engineering, Bright Star University, El Brega, Libya
[3] Department of Engineering Management, Bright Star University, Brega, Libya
[4] Department of Mechanical Engineering, Faculty of Engineering, Tobruk University, Tobruk, Libya
[5] Department of Aeronautical Engineering, Faculty of Engineering, Sudan University of Science and Technology, Khartoum, Sudan

الكشف عن الأعطال الكامنة في التوربينات الغازية باستخدام التحقق الزمني وخصائص مستندة إلى الفيزياء ونماذج شجرة قرار قابلة للتفسير

منتصر علي سعيد محمد [1*]، عبد القادر عقيلة صالح غيضان[2]، سالمة رجب [3]، سعد عباس ادم [4]، طارق حسن السني [5]
[1] قسم هندسة الطيران، كلية علوم الطيران، جامعة النجم الساطع، البريقة، ليبيا
[2] قسم الهندسة الميكانيكية، كليه الهندسة التقنية، جامعة النجم الساطع، البريقة، ليبيا
[3] قسم الإدارة الهندسية، كلية الادارة، جامعة النجم الساطع، البريقة، ليبيا
[4] قسم الهندسة الميكانيكية، كلية الهندسة، جامعة طبرق، طبرق، ليبيا
[5] قسم هندسة الطيران، كلية الهندسة، جامعة السودان للعلوم والتكنولوجيا، الخرطوم، السودان

[*]Corresponding author: montaser.saeed@bsu.edu.ly

Received: February 01, 2026	Accepted: April 26, 2026	Published: May 11, 2026

Abstract:
This study addresses fault detection in gas turbines under sensor-limited conditions, where diagnostically informative measurements such as emissions and internal variables are unavailable. A leakage-aware framework is proposed using only five readily available thermodynamic sensors (AT, AP, TAT, AFDP, TEY). To compensate for missing measurements, physics-based features grounded in Brayton cycle principles are constructed. A proxy fault label is generated offline using high-fidelity variables (TIT, GTEP, CO, NOx), which are assumed to be unavailable during real-time deployment. A Decision Tree classifier is selected to ensure interpretability in safety-critical environments. To reflect realistic industrial conditions, a strict time-based validation strategy is adopted. The results show that the proposed model achieves an F1-score of 0.703, a Recall of 0.881, and an AUC of 0.951. Furthermore, random split validation is found to overestimate performance by approximately 5.7% in F1-score, highlighting the risk of optimistic bias in conventional evaluation practices. The proposed framework provides a practical and interpretable solution for fault detection under constrained sensing conditions, with direct applicability to legacy turbine systems.

Keywords: Gas turbine fault detection; latent fault prediction; physics-informed features; decision tree; industrial diagnostics; time-based validation.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

الملخص
تقترح هذه الدراسة إطارًا واعيًا بتسرب البيانات ومتحققًا زمنيًا لاكتشاف الأعطال الكامنة في أنظمة التوربينات الغازية، وذلك بالاعتماد على خصائص مستندة إلى المبادئ الفيزيائية ونموذج شجرة قرار قابل للتفسير. تُظهر النتائج إمكانية تحقيق أداء تشخيصي مرتفع(AUC = 0.951) ، والاسترجاع = 0.881 (ضمن قيود واقعية، في حين يبرز الفارق الملحوظ في الأداء بين التقسيم العشوائي والتقسيم الزمني (حوالي 5.7%) أهمية اعتماد أساليب تقييم تراعي البنية الزمنية للبيانات.
من الناحية التطبيقية، يوفر النموذج حلاً شفافًا وقابلًا للنشر، حيث تستند قراراته إلى خصائص ذات دلالة فيزيائية واضحة. ومع ذلك، فإن الاعتماد على مؤشرات أعطال بديلة، إلى جانب وجود أعطال غير مكتشفة، يشير إلى مجالات قابلة للتحسين. وتركز الأعمال المستقبلية على التحقق باستخدام بيانات أعطال حقيقية، وتقليل الأخطاء السلبية عبر نماذج متقدمة أو هجينة، بالإضافة إلى تعزيز متانة النموذج من خلال أساليب تعلم تكيفية لمواجهة تغير توزيعات البيانات.

الكلمات المفتاحية: كشف أعطال التوربينات الغازية، التنبؤ بالأعطال الكامنة، تعلم آلي قابل للتفسير، شجرة القرار، التشخيص الصناعي.

Introduction

Gas turbines are critical assets in power generation and industrial applications, where reliable operation is essential for safety and economic performance. Traditionally, condition monitoring relies on comprehensive sensor suites to detect abnormal behavior and prevent failures [3]. However, in many real-world deployments, key diagnostic sensors such as emission monitors (NOx, CO) and internal measurements (e.g., TIT) are often unavailable due to cost, reliability, and accessibility constraints. This limitation is particularly pronounced in legacy systems and remote installations [5,7,10,13]. This creates a fundamental challenge: how can faults be reliably detected when the most informative sensors are unavailable during operation?

Existing studies largely assume full sensor availability and often rely on black-box models or inappropriate validation strategies, limiting their applicability in real-world industrial environments. To address these limitations, this study proposes a leakage-aware and interpretable fault-detection framework based on limited thermodynamic measurements and time-consistent validation [8,12,4,6].

In this study, 'latent faults' refer to abnormal operating conditions that are indirectly inferred from available measurements, as direct fault indicators (e.g., emission sensors) are assumed unavailable [9]. A comprehensive summary of the dataset and experimental setup is provided in **Table 2**, including data partitioning, fault distribution, and statistical shifts between training and testing sets. These variables capture key aspects of turbine performance, such as combustion efficiency, environmental compliance, thermodynamic behavior, and mechanical integrity [1]. The spatial distribution of sensors and their corresponding parameter sources is illustrated in **Figure 1**. The input variables are categorized into two groups: (i) ambient conditions (e.g., temperature, humidity, and pressure), which characterize the external environment, and (ii) process parameters (e.g., turbine energy output and air filter differential pressure), which describe internal system dynamics.

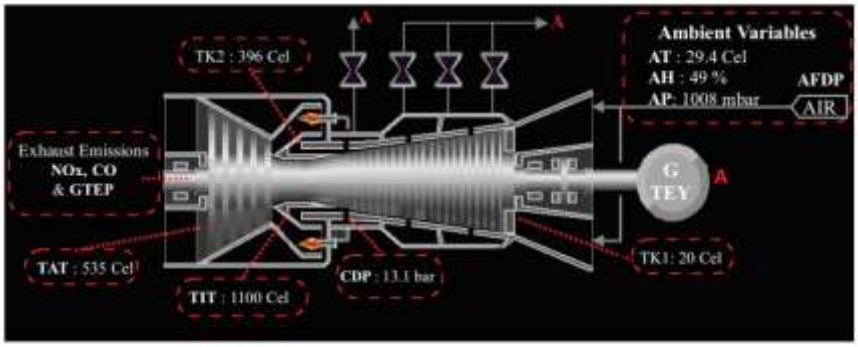

Figure 1: The spatial distribution of sensors and their corresponding parameter sources [1]

Despite their importance, these sensors are not always available in practical applications. Several constraints limit their deployment. First, installing advanced sensor systems in legacy turbines is often economically prohibitive. Second, sensors operating under harsh thermal and mechanical conditions exhibit reduced reliability and require frequent calibration and maintenance [8,14, 18, 22]. Third, accessibility challenges arise in remote environments such as offshore platforms and desert installations, where maintenance operations are logistically complex. In addition, digital twin environments inherently lack physical sensors, and regulatory requirements vary across jurisdictions, leading to inconsistent sensor availability [1, 3, 5]. These limitations raise

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

a fundamental and practically relevant question: how can turbine faults be reliably detected when the most diagnostically informative sensors are unavailable, unreliable, or cost-prohibitive?

Problem Statement and Research Gaps

Given a set of operational measurements limited to five readily available sensors (AT, AP, TAT, AFDP, TEY) and no access to emission or internal measurements, the objective is to predict a proxy fault indicator representing abnormal operating conditions. This problem is inherently challenging due to three factors: (i) the nonlinear and indirect relationship between observable variables and fault conditions, (ii) temporal dependencies and distribution shifts in industrial time-series data, and (iii) the scarcity of reliable ground-truth labels in real-world settings.

Contributions

This study makes three main contributions:

1. Leakage-aware framework: A feature engineering strategy that excludes direct and derived information from target-defining variables.
2. Validation bias quantification: A systematic comparison between random and time-based validation, showing an overestimation of approximately 5.7% in F1-score.
3. Interpretable modeling: A Decision Tree-based approach designed for practical deployment under sensor-limited conditions.

Paper Organization

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the methodology, Section 4 presents the results, Section 5 discusses the findings, and Section 6 concludes the study.

Literature Review

Gas turbine condition monitoring has evolved from traditional model-based diagnostics to advanced data-driven approaches [1,2]. Existing literature can be broadly categorized into three domains: traditional methods, machine learning techniques, and sensor-limited monitoring. Despite this progress, several methodological limitations persist, particularly concerning data leakage, validation strategies, interpretability, and sensor availability. Traditional approaches, including gas path analysis and Kalman filtering [3–6], estimate expected thermodynamic behavior and detect deviations from nominal conditions.

While these methods are inherently interpretable, their effectiveness depends on accurate system modeling, and they remain sensitive to noise and modeling uncertainties. Signal processing techniques [7,8] are primarily suited for mechanical fault detection but are less effective for thermodynamic anomalies without dedicated sensors. Similarly, expert systems [9,10] provide transparent decision-making but require extensive manual knowledge engineering and lack adaptability. Overall, these methods rely on strong assumptions and the availability of comprehensive sensors, limiting their applicability in real-world scenarios. Data-driven approaches, including supervised models such as SVM, ANN, and ensemble techniques [11–16], as well as deep learning architectures (e.g., CNNs, LSTMs, and Transformers) [17,18], have demonstrated strong predictive capabilities by capturing complex nonlinear relationships. Unsupervised methods, such as autoencoders and isolation forests [19,20], are employed when labeled data are scarce but often suffer from high false alarm rates and limited interpretability.

However, these approaches typically assume full sensor availability, rely on black-box models, and frequently adopt random train-test splits that fail to account for temporal dependencies, leading to potential data leakage and overly optimistic performance estimates [21,22]. Sensor-limited fault detection remains relatively underexplored. Existing studies attempt to infer system behavior using reduced thermodynamic measurements [23–26], but often rely on restrictive assumptions and lack rigorous, time-aware validation. Moreover, the impact of validation strategies on performance estimation is rarely quantified. Interpretability remains a critical requirement for industrial deployment, where transparent models are preferred to ensure operator trust and regulatory compliance. While post-hoc explanation techniques such as SHAP and LIME provide insights into complex models, their instability and inconsistency limit their reliability in safety-critical environments. Furthermore, the quantification of validation bias-comparing random versus time-based splits—remains insufficiently addressed in the gas turbine fault detection literature, representing a key gap that this study aims to fill [27-30].

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Research Positioning

Based on the identified gaps, current literature reveals three key shortcomings: (i) over-reliance on full sensor availability, (ii) limited interpretability of high-performing models, and (iii) inadequate validation practices that neglect temporal structure and inflate performance estimates. Additionally, sensor-limited scenarios and validation bias remain insufficiently investigated. To address these limitations, this study proposes a leakage-controlled and interpretable framework based on Decision Trees, evaluated using strict time-based validation. The proposed approach is specifically designed for realistic deployment under constrained sensing conditions, ensuring both transparency and robustness in industrial fault detection applications.

Methodology and experimental setup

Dataset Description

The dataset comprises 36,733 hourly observations collected over five years (2011-2015), capturing key ambient, operational, and emission-related variables of the gas turbine system. Table 1 summarizes the statistical properties of all measured variables, including their minimum, maximum, and mean values [1]. The wide variation across features highlights the need for appropriate normalization and physically consistent feature construction.

Experimental Setup

Based on this dataset, the experimental setup is defined using a strict chronological split, as summarized in Table 2. This setup preserves temporal dependencies and reflects real-world deployment conditions, in which models must generalize to unseen future data.

Table 1. Basic statistical information of the data used in the study

Variable	Abbr.	Unit	Min	Max	Mean
Ambient temperature	AT	°C	–6.23	37.10	17.71
Ambient pressure	AP	mbar	985.85	1036.56	1013.07
Ambient humidity	AH	(%)	24.08	100.20	77.87
Air filter difference pressure	AFDP	mbar	2.09	7.61	3.93
Gas turbine exhaust pressure	GTEP	mbar	17.70	40.72	25.56
Turbine inlet temperature	TIT	°C	1000.85	1100.89	1081.43
Turbine after temperature	TAT	°C	511.04	550.61	546.16
Compressor discharge pressure	CDP	mbar	9.85	15.16	12.06
Turbine energy yield	TEY	MWH	100.02	179.50	133.51
Carbon monoxide	CO	mg/m³	0.00	44.10	2.37
Nitrogen oxides	NOx	mg/m³	25.90	119.91	65.29

The measured variables include ambient temperature (AT), ambient pressure (AP), ambient humidity (AH), air filter differential pressure (AFDP), gas turbine exhaust pressure (GTEP), turbine inlet temperature (TIT), turbine after temperature (TAT), compressor discharge pressure (CDP), turbine energy yield (TEY), carbon monoxide (CO), and nitrogen oxides (NOx). These variables collectively capture key aspects of system behavior, including environmental conditions, thermodynamic performance, and emission characteristics. To ensure a realistic evaluation setting and avoid look-ahead bias, the dataset is partitioned chronologically.

Data from January 2011 to December 2013 are used for model development (training and validation), while data from January 2014 to December 2015 are reserved for out-of-sample testing. This time-based splitting strategy preserves the data's temporal structure and reflects real-world deployment scenarios. Within this framework, a subset of thermodynamic variables is selected to simulate sensor-limited conditions, forming the basis for subsequent feature construction and latent target formulation. As shown in Table 2, a time-based split is adopted to preserve temporal consistency and mitigate data leakage. The observed distribution shift, particularly the localized variation in AFDP, highlights the non-stationary nature of the data, reinforcing the need for robust feature engineering and validation strategies.

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Table 2: Dataset and Experimental Setup

Parameter	Value
Total Samples	36,733
Number of Features	11
Physics-Based Features	10
Overall Fault Rate	4.99%
Training Samples	29,386
Test Samples	7,347
Training Fault Rate	4.63%
Test Fault Rate	6.41%
Average Distribution Shift	2.74%
Maximum Shift (AFDP)	10.12%

Sensor Selection and Physics-Based Feature Engineering

To simulate realistic industrial constraints, the sensor set is partitioned into available and unavailable measurements. Let the observable feature vector be defined as:

$$x_t = [AT_t, AP_t, TAT_t, AFDP_t, TEY_t] \quad (1)$$

where t denotes the time index (hourly observation). Sensors are categorized into two groups:

(i) available sensors: AT, AP, TAT, AFDP, TEY, and
(ii) excluded sensors: TIT, GTEP, CO, NOx.

The excluded sensors are not used due to practical constraints, including cost, reliability degradation, and limited accessibility in industrial environments. This setup reflects realistic deployment conditions rather than idealized laboratory settings.

Proxy Fault Definition

A proxy fault label is constructed using high-fidelity variables (TIT, GTEP, CO, NOx), which are assumed to be unavailable during real-time operation but accessible in offline or simulation environments. The label is defined based on threshold conditions applied to these variables, representing abnormal operating states. It is important to note that the term "proxy" reflects that the label is not directly observable during deployment, but is used during model development to approximate fault conditions. Let the unavailable sensor vector be defined as:

$$u_t = [TIT_t, GTEP_t, CO_t, NOx_t] \quad (2)$$

where u_t represents internal and emission-related measurements that are excluded from the model inputs, a latent fault score is then defined as a function of these variables:

$$s_t = f(u_t) \quad (3)$$

where $f(u_t)$ denotes a physically meaningful aggregation function that captures abnormal operating conditions based on high-fidelity measurements, to convert this continuous score into a binary classification label, a threshold-based mapping is applied:

$$y_t = \begin{cases} 1, & \text{if } s_t > \tau \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where τ is a predefined threshold that separates normal and abnormal operating conditions. This formulation enables supervised learning while preserving deployment realism, as the model is trained to predict y_t using only observable variables x_t, without direct access to the unavailable sensor set u_t. Importantly, this strict separation

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

ensures that no target-defining information is included in the input features, thereby eliminating data leakage and maintaining the validity of the proposed framework.

Physics-Based Feature Engineering

To validate the physical consistency of the dataset and its derived features, the relationships among key thermodynamic variables are examined. In particular, the correlation between turbine inlet temperature (TIT) and NOx emissions is analyzed, as higher combustion temperatures are expected to increase NOx formation. The Pearson correlation coefficient between TIT and NOx is computed, and the coefficient of determination (R^2) is used to quantify the strength of this relationship. A strong positive correlation confirms that the dataset preserves expected physical behavior, supporting the validity of the proposed physics-based feature construction. Although direct measurement of all cycle variables is unavailable, this relationship motivates the design of proxy features that capture efficiency-related behavior, given the heterogeneous physical units and ranges (see Table 1). To address this, normalization is applied. where μi and σi represent the mean and standard deviation of feature i, respectively.

$$\text{Normalized Power} = \frac{TEY}{AP} \tag{5}$$

$$x_t^{(i,\text{norm})} = \frac{x_t^{(i)} - \mu_i}{\sigma_i} \tag{6}$$

A temperature ratio is introduced to capture relative thermal behavior:

$$\phi_1(t) = \frac{TAT_t}{AT_t} \tag{7}$$

This feature reflects thermal amplification throughout the turbine system, serving as a proxy for heat-transfer efficiency under varying ambient conditions. To normalize energy output with respect to environmental pressure, the following feature is defined:

$$\phi_2(t) = \frac{TEY_t}{AP_t} \tag{8}$$

This formulation accounts for the influence of ambient pressure on turbine performance and provides a pressure-adjusted energy indicator. A simplified efficiency proxy is constructed as:

$$\phi_3(t) = \frac{TEY_t}{TAT_t} \tag{9}$$

This ratio approximates thermal efficiency by relating useful energy output to exhaust temperature, which reflects residual thermal losses. To quantify flow resistance and potential degradation effects, a pressure loss indicator is defined:

$$\phi_4(t) = \frac{AFDP_t}{AP_t} \tag{10}$$

This feature captures the relative pressure drop across the air filtration system, which may indicate fouling or blockage. A combined thermodynamic efficiency index is formulated as:

$$\phi_5(t) = \frac{TEY_t}{TAT_t \cdot AFDP_t} \tag{11}$$

This composite feature integrates thermal and flow effects, providing a more comprehensive proxy for system efficiency under constrained sensing. To compensate for the absence of emission and internal measurements, physics-based features were constructed using the available sensors. These features are grounded in thermodynamic principles of gas turbine operation, particularly the Brayton cycle:

$$\eta_{th} = 1 - \frac{T_1}{T_2} \tag{12}$$

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

These physics-based features provide an indirect yet informative representation of the underlying thermodynamic behavior. In the next step, these features are used to predict a latent fault indicator defined using unavailable high-fidelity measurements, as formalized in the following section.

Data Splitting Strategy

Due to the dataset's temporal nature, a strict time-based splitting strategy is adopted. Random splitting is avoided as it violates temporal dependencies and introduces look-ahead bias. The dataset is divided chronologically into 80% training and 20% testing samples, ensuring that all training data precede testing data. A Decision Tree classifier is employed due to its interpretability and suitability for low-dimensional feature spaces. The model partitions the feature space using recursive binary splits based on impurity reduction. The Gini impurity is used as the split criterion, enabling the model to identify decision boundaries that separate normal and faulty conditions. The resulting model provides transparent decision rules that operators can directly interpret. To quantify the impact of improper validation, model performance is evaluated under both random and time-based splits. The difference in F1-score (ΔF1) is used as a measure of overestimation bias [31,32].

Given a temporally ordered dataset:

$$\{(x_t, y_t)\}_{t=1}^{T} \tag{13}$$

The training and testing sets are defined as:

$$\mathcal{D}_{train} = \{x_t\}_{t=1}^{\lfloor 0.8T \rfloor}, \quad \mathcal{D}_{test} = \{x_t\}_{t=\lfloor 0.8T \rfloor + 1}^{T} \tag{14}$$

This formulation guarantees that all training observations precede testing observations, eliminating look-ahead bias and aligning evaluation with real-world deployment conditions. To quantify the impact of improper validation strategies, model performance is evaluated under both random and time-based splits. The resulting overestimation bias is defined as:

$$\Delta F1 = F1_{\text{random}} - F1_{\text{time}} \tag{15}$$

A positive ΔF1 indicates that random splitting inflates performance due to temporal information leakage. Hyperparameter tuning is performed using time-series cross-validation, ensuring that validation data always follow the training data in time. The model learns a mapping from the input feature space to the latent fault label:

$$\hat{y}_t = g(x_t) \tag{16}$$

where $g(x_t)$ represents a hierarchical structure of recursive binary splits. At each node, the model partitions the feature space based on a threshold applied to a selected feature:

$$x_t^{(j)} \leq \theta \tag{17}$$

where $x_t^{(j)}$ is the j-th feature, and θ is the split threshold. To determine the optimal split, the model minimizes node impurity using the Gini index.

$$Gini(S) = 1 - \sum_{k=1}^{K} p_k^2 \tag{18}$$

where p_k^2 is the proportion of samples belonging to class k in node S. For a candidate split, the optimal threshold is selected by minimizing the weighted impurity of the child nodes:

$$\theta^* = \text{argmin}_\theta \left(\frac{|S_L|}{|S|} Gini(S_L) + \frac{|S_R|}{|S|} Gini(S_R) \right) \tag{19}$$

The recursive partitioning continues until a stopping criterion is met, yielding a transparent set of decision rules that can be directly interpreted in terms of the physical system's behavior. The Decision Tree learns a mapping:

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

$$\hat{y} = f(x; \theta) \tag{20}$$

where decisions are made through recursive binary splits of the form:

$$x_j < t \tag{21}$$

To select optimal splits, the model minimizes node impurity using the Gini criterion:

$$Gini(m) = 1 - \sum_{k=1}^{2} p_{mk}^2 \tag{22}$$

where p_{mk}^2 represents the proportion of class k samples in node m. For a given input x, the model outputs a probabilistic estimate:

$$P(y = 1 \mid x) = \frac{N_{\text{fault}}}{N_{\text{leaf}}} \tag{23}$$

Model hyperparameters are optimized using grid search combined with time-aware cross-validation. The objective is to maximize the average validation F1 score:

$$\overline{F1}_{CV}(\lambda) = \frac{1}{k} \sum_{i=1}^{k} F1_{val}^{(i)}(\lambda) \tag{24}$$

The optimal configuration is selected as:

$$\lambda^* = \text{argmax}_{\lambda \in \Lambda} \overline{F1}_{CV}(\lambda) \tag{25}$$

This ensures a balance between model complexity and generalization performance. Instead of using the default classification threshold, the final prediction is defined as:

$$\hat{y} = \begin{cases} 1 & \text{if } P(y = 1 \mid x) \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

The optimal threshold is determined by maximizing the F1 score:

$$\tau^* = \text{argmax}_{\tau} F1(\tau) \tag{27}$$

Given the safety-critical nature of gas turbine operation, the threshold is further adjusted to prioritize recall over precision:

$$\tau < 0.5 \tag{28}$$

In this study, the selected threshold is:

$$\tau = 0.56$$

This choice reflects a deliberate trade-off that reduces missed faults at the expense of increased false alarms, aligning with industrial risk management priorities. All stages of model development, including feature engineering, hyperparameter tuning, and threshold selection, are performed exclusively on the D_{train}. The test set D_{test} remains completely unseen until the final evaluation:

$$D_{train} \cap D_{test} = \emptyset \tag{29}$$

Results and discussion

Although ensemble and deep learning models (e.g., Random Forest, XGBoost, neural networks) may achieve higher predictive accuracy, they are intentionally excluded to prioritize interpretability and deployability in safety-

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

critical environments. Decision Trees provide transparent, rule-based decisions that support operator trust and regulatory requirements. A Mann–Whitney U test confirms statistically significant differences ($p < 0.01$) between normal and anomalous conditions for key features (e.g., TAT, AFDP), indicating that detected anomalies reflect meaningful system deviations.

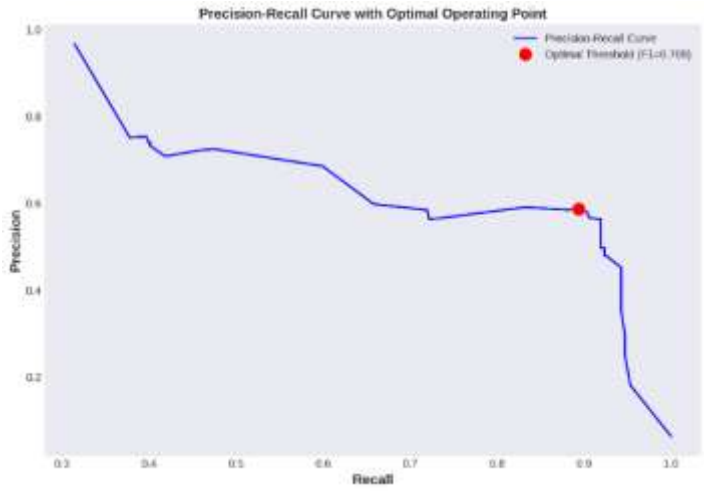

Figure 2: Precision–Recall Trade-off and Safety-Oriented Operating Point for Latent Fault Detection

The selected operating point ($\tau = 0.56$) achieves an F1-score of 0.703, with Recall 0.881 and Precision 0.585. This high-recall region reflects a safety-oriented strategy that prioritizes fault detection over minimizing false alarms. Such a trade-off is appropriate in industrial settings where missed faults carry a higher risk.

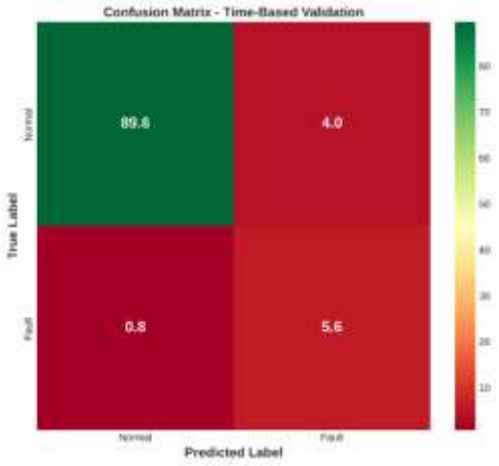

Figure 3: Confusion Matrix under Safety-Oriented Threshold Selection

At $\tau = 0.56$, the model detects 331 faults with 35 missed cases (Recall = 0.881) and generates 224 false positives (false alarm rate ≈ 4.28%). This behavior aligns with the selected threshold, minimizing false negatives while maintaining acceptable false alarm levels.

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

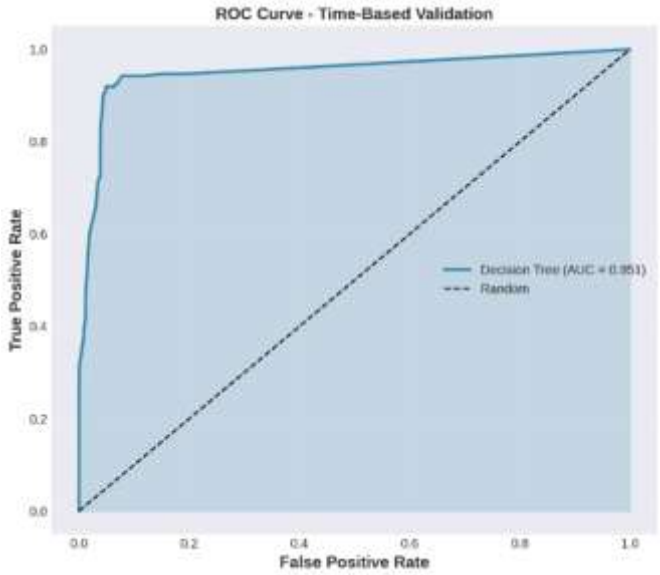

Figure 4: Receiver Operating Characteristic (ROC) Curve for Time-Based Validation of the Proposed Decision Tree Model

The ROC curve demonstrates the discriminative capability of the Decision Tree under strict temporal validation, achieving an AUC of 0.951, which indicates excellent separability between normal and fault conditions. The curve consistently exceeds the random baseline, confirming that the model captures meaningful structure rather than noise. The steep initial slope reflects high true-positive rates at low false-positive rates—an essential property for safety-critical systems. Unlike random splitting, time-based evaluation ensures that performance reflects generalization to unseen future data. Overall, the result validates the robustness of the proposed leakage-aware and temporally consistent framework under realistic deployment conditions.

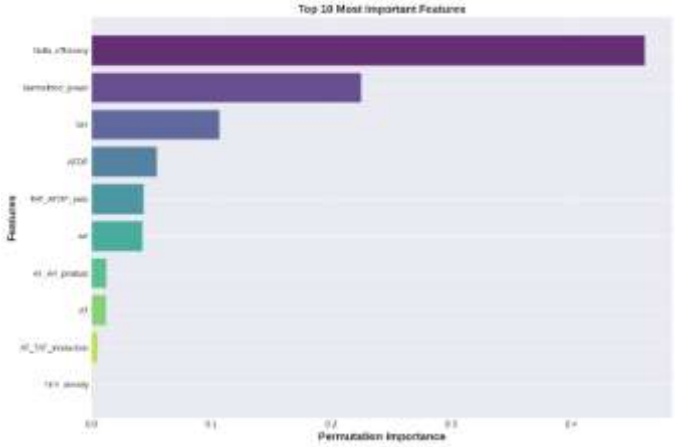

Figure 5: Permutation-Based Feature Importance Ranking for the Proposed Model

The figure presents the top 10 most important features ranked using permutation importance on the test set. The results indicate that Delta_efficiency is the dominant predictor, followed by Normalized_power and TAT, suggesting that derived thermodynamic relationships contribute more significantly to fault detection than raw sensor measurements. The dominance of Delta_efficiency underscores the effectiveness of physics-based feature engineering in capturing latent system behavior not directly observable from raw measurements. Similarly, Normalized_power and TAT reflect key thermodynamic interactions associated with performance degradation and abnormal operating conditions. Notably, features derived from combinations of accessible sensors (e.g., TAT_AFDP_ratio) contribute more meaningfully than individual low-level variables such as AT or TEY_density. This confirms that informative representations can be constructed without relying on unavailable emission or internal sensors. The use of permutation importance ensures that feature contributions are evaluated based on their actual impact on model performance, avoiding bias associated with model-specific metrics. Overall, the results

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

validate the proposed feature engineering strategy and demonstrate that interpretable, physics-informed features can effectively support accurate fault detection in sensor-limited environments.

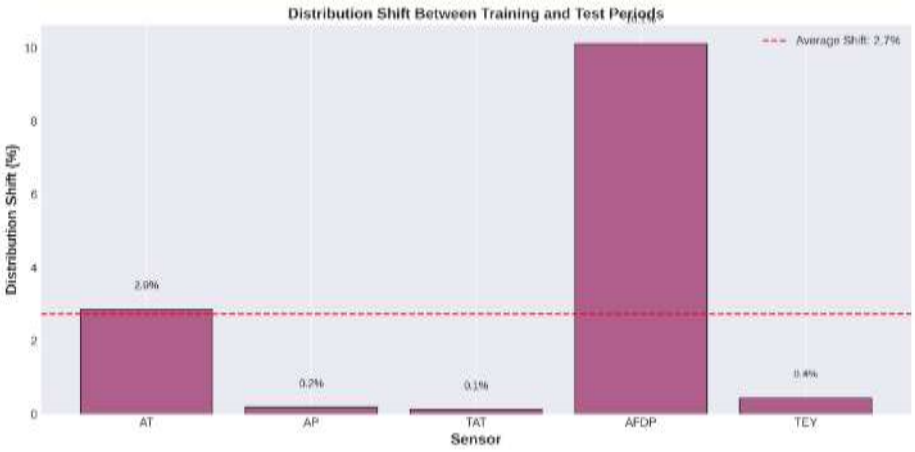

Figure 6: Distribution Shift Between Training and Test Periods for Key Sensors

The figure illustrates that the distribution shift between training and testing periods is limited (≈2.7% on average), although AFDP exhibits a higher deviation (~10%), suggesting localized operational changes. Despite this shift, model performance remains stable, indicating robustness to moderate temporal variability. The presence of a measurable shift further justifies the use of time-based validation, as it reflects realistic deployment conditions where data distributions evolve.

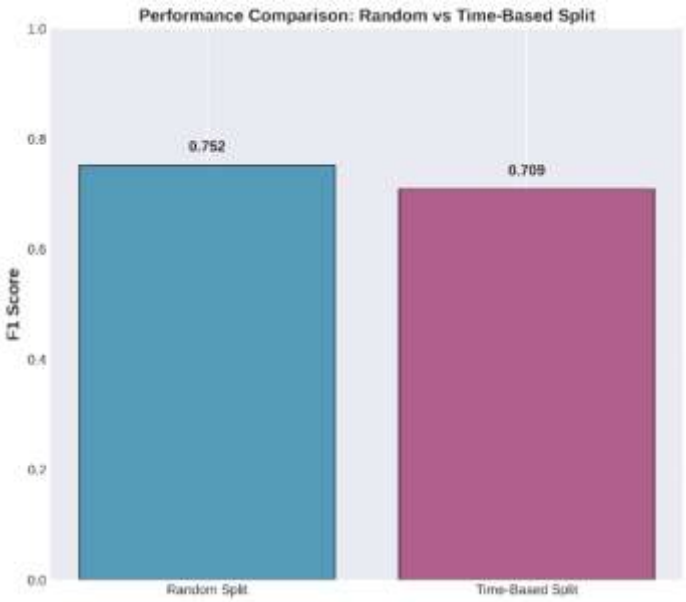

Figure 7: Impact of Data Splitting Strategy on Model Generalization Performance

The figure shows a clear performance gap between the random split (F1 = 0.752) and the time-based split (F1 = 0.709), corresponding to an overestimation bias of ≈5.7%. This confirms that random splitting introduces temporal leakage and yields overly optimistic results. In contrast, time-based evaluation provides a more reliable estimate of real-world performance. This finding reinforces the necessity of temporally consistent validation in industrial time-series applications.

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

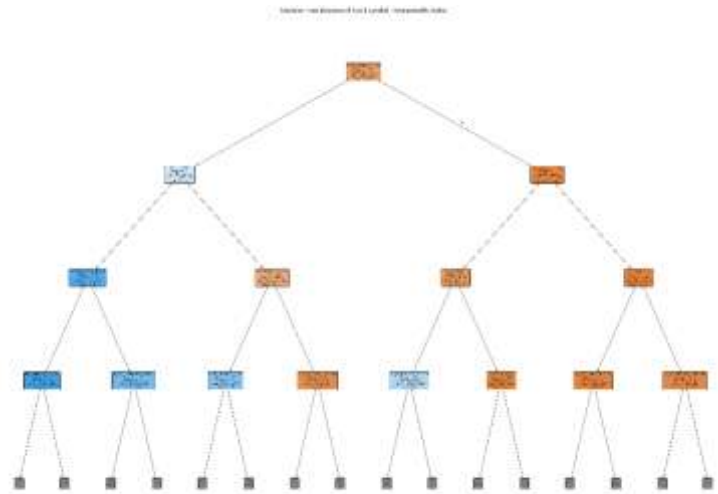

Figure 8: Interpretable Surrogate Decision Tree for Weakly-Supervised Fault Classification in Gas Turbine Diagnostics

The figure presents a decision tree that provides a transparent representation of the learned decision boundaries, where each node corresponds to a physically interpretable rule. The limited depth and balanced structure indicate that the classification problem is well-structured and does not require excessive complexity. This interpretability is critical in safety-critical systems, enabling traceability, validation, and operator trust. The result demonstrates that high diagnostic performance can be achieved without sacrificing explainability, supporting practical deployment. Including this figure is critical as it bridges the gap between model accuracy and interpretability—an essential requirement in safety-critical systems such as gas turbines. It demonstrates that the proposed framework not only achieves high performance but also yields explainable decision pathways, thereby supporting trust, validation, and potential deployment in real-world prognostics and health management (PHM) systems.

Model Performance and Operating Point

Table 3 summarizes the model performance under time-based validation. The model achieves strong discriminative capability (AUC = 0.951) and high accuracy (0.952), indicating reliable separation between normal and fault conditions. More importantly, the model achieves a high recall (0.881), which is essential in fault-detection scenarios. As shown in Figure 2 (Precision–Recall curve), threshold tuning shifts the operating point toward higher sensitivity.

Table 3: Model Performance and Optimal Operating Point

Metric	Default ($\tau = 0.50$)	Optimal ($\tau = 0.56$*)
Accuracy	0.9524	—
Precision	0.5853	0.5872
Recall	0.8811	0.8938
F1 Score	0.7034	0.7088
AUC-ROC	0.9510	—
MCC	0.6954	—

The optimal threshold ($\tau = 0.56$) improves recall to 0.894 while marginally increasing precision, resulting in a slightly higher F1 score (0.709). This confirms that the model can be tuned toward safety-oriented operation without compromising robustness.

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Validation and Generalization

Table 4 compares random and time-based validation. A clear performance gap is observed (F1: 0.752 vs. 0.709), corresponding to an overestimation bias of approximately 5.7%. As illustrated in Figure 7, this gap arises from temporal leakage introduced by random splitting. In contrast, time-based validation enforces chronological consistency and provides a more realistic estimate of model performance. Furthermore, the distribution shift analysis (Figure 6) shows moderate variation (~2.7% on average), with higher deviation in AFDP. This explains part of the performance degradation and highlights the importance of evaluating models under evolving operational conditions.

Table 4: Validation Strategy Comparison

Method	F1 Score	Precision	Recall
Random Split	0.7519	0.6886	0.8279
Time-Based Split	0.7088	0.5872	0.8938

Safety Analysis

Table 5 presents the confusion matrix-derived safety metrics. The model achieves a low false alarm rate (4.28%) while maintaining high recall, detecting 415 out of 471 fault instances. However, 56 faults remain undetected (missed fault rate = 11.89%). This trade-off reflects the selected operating point ($\tau = 0.56$), which prioritizes fault detection over minimizing false alarms. As observed in Figure 3 (Confusion Matrix), the model prioritizes sensitivity, consistent with safety-critical requirements in which missed faults are more costly than false positives.

Table 5: Error Analysis and Safety Metrics

Metric	Value
True Negatives	6582
False Positives	294
False Negatives	56
True Positives	415
False Alarm Rate	4.28%
Missed Fault Rate	11.89%

Feature Importance and Interpretability

Table 6 shows that physics-informed features dominate the model, particularly Delta_efficiency (0.46) and Normalized_power (0.23). As supported by Figure 5, these features capture key thermodynamic relationships that are not directly observable from raw measurements. The Decision Tree structure (Figure 8) further enhances interpretability by providing explicit decision rules. This transparency enables direct mapping between input features and predictions, supporting trust and deployment in industrial environments.

Table 6: Top 10 Feature Importance

Feature	Importance
Delta_efficiency	0.4614
Normalized_power	0.2250
TAT	0.1066
AFDP	0.0543
TAT_AFDP_ratio	0.0437
AP	0.0425
AT_AP_product	0.0126
AT	0.0121
AT_TAT_interaction	0.0052
TEY_density	0.0010

Conclusion and Future Work

This study proposed a leakage-aware, time-validated framework for latent fault detection in gas turbine systems, using physics-informed features and an interpretable Decision Tree model. The results demonstrate that high diagnostic performance (AUC = 0.951, recall = 0.881) can be achieved under realistic constraints. In contrast, the observed performance gap between random and time-based validation (≈5.7%) highlights the importance of

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

temporally consistent evaluation. From a practical perspective, the model provides a transparent and deployable solution, with decisions grounded in physically meaningful features.

However, reliance on proxy fault labels and missed faults indicates areas for improvement. Future work will focus on validation using real fault data, reducing false negatives through advanced or hybrid models, and enhancing robustness via adaptive learning under distribution shift.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Kaya, Heysem & Tufekci, Pinar & Uzun, Erdinç. (2019). Predicting CO and NOx emissions from gas turbines: novel data and benchmark PEMS. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES. 27. 4783-4796. 10.3906/elk-1807-87.

[2] Zaccaria, Valentina & Rahman, Moksadur & Aslanidou, Ioanna & Kyprianidis, Konstantinos. (2019). A Review of Information Fusion Methods for Gas Turbine Diagnostics. Sustainability. 11. 6202. 10.3390/su11226202.

[3] Zeng, Detang, Dengji Zhou, Chunqing Tan, and Baoyang Jiang. 2018. "Research on Model-Based Fault Diagnosis for a Gas Turbine Based on Transient Performance" Applied Sciences 8, no. 1: 148. https://doi.org/10.3390/app8010148

[4] Tsui, Kwok-Leung & Chen, Nan & Zhou, Qiang & Hai, Yizhen & Wang, Wenbin. (2015). Prognostics and Health Management: A Review on Data-Driven Approaches. Mathematical Problems in Engineering. 2015. 1-17. 10.1155/2015/793161.

[5] Haub GL, Hauhe WE. Field evaluation of online compressor cleaning in heavy duty industrial gas turbines. ASME 90-GT-107; 1990.

[6] Meher-Homji CB. Gas turbine axial compressor fouling – a unified treatment of Its effects detection and control. ASME cogeneration-turbo conference, vol. 5; 1990. p. 179–90.

[7] Peltier RV, Swanekamp RC. LM2500 recoverable and non-recoverable power loss. ASME cogeneration-turbo power conference, Vienna, Austria; August 1995.

[8] Sasahara O. JT9D engine/module Performance Deterioration results from back to back testing; 1986.

[9] Flashbery LS, Haub GL. Measurement of combustion turbine non-recoverable degradation. ASME 92-GT-264; 1992.

[10] Crosby JK. Factors relating to deterioration based on Rolls-Royce RB211 in service performance. Turbomachin Perform Deteriorat 1986;37:41–7.

[11] Saravaramuttoo HIH, Maclsaac BD. Thermodynamic models for pipeline gas turbine diagnostics. ASME J Eng Power 1983(October):105.

[12] Baojun N, Mei Y, Xingjian S, Peng W. Random FEA and reliability analysis for combustor case., 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin 2017: 1-5, https://doi.org/10.1109/PHM.2017.8079268.

[13] Echarda B, Gaytona N, Bignonnet A. A reliability analysis method for fatigue design. International Journal of Fatigue 2014: 59: 292-300, https://doi.org/10.1016/j.ijfatigue.2013.08.004.

[14] Le Clainche, Soledad & Ferrer, Esteban & Gibson, Sam & Cross, Elisabeth & Parente, Alessandro & Vinuesa, Ricardo. (2022). Improving aircraft performance using machine learning: a review. 10.48550/arXiv.2210.11481.

[15] Batayev, Nurlan & Onbayev, Ayan. (2018). Prediction of Gas Turbine Parameters based on Machine Learning Regression Methods. 217-221. 10.18638/scieconf.2018.6.1.495.

[16] Y.G. Li, P. Nilkitsaranont, Gas turbine performance prognostic for condition-based maintenance, Applied Energy, Volume 86, Issue 10, 2009, Pages 2152-2161, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2009.02.011.

[17] de Castro-Cros, M., Velasco, M., & Angulo, C. (2021). Machine-Learning-Based Condition Assessment of Gas Turbines—A Review. Energies, 14(24), 8468. https://doi.org/10.3390/en14248468.

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

[18] Carvalho, T.P.; Soares, F.A.; Vita, R.; Francisco, R.d.P.; Basto, J.P.; Alcalá, S.G. A systematic literature review of machine-learning methods applied to predictive maintenance. Comput. Ind. Eng. 2019, 137, 106024.

[19] Qiu, J.; Wu, Q.; Ding, G.; Xu, Y.; Feng, S. A survey of machine learning for big data processing. Eurasip J. Adv. Signal Process. 2016, 2016, 67.

[20] M. Tahan, E. Tsoutsanis, M. Muhammad, and Z.A. Abdul Karim, "Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: a review," Appl. Energy, 198 122–144 (2017). doi:10.1016/j.apenergy.2017.04.048.

[21] Y.G. Li, and P. Nilkitsaranont, "Gas turbine performance prognostic for condition-based maintenance," Appl. Energy, 86 (10) 2152–2161 (2009). doi:10.1016/j.apenergy.2009.02.011.

[22] J. Lee, H. Park, and S. Kim, "Gas turbine performance prediction using multiple linear regression," IEEE Trans. Ind. Electron., vol. 66, no. 12, pp. 9874 - 9883, Dec. 2019.

[23] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed., Melbourne, Australia: OTexts, 2018.

[24] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[25] X. Zhang, Y. Liu, and Y. Wang, "Random forest-based aircraft engine failure prediction," Aerosp. Sci. Technol., vol. 84, pp. 790-797, Aug. 2018.

[26] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees, Belmont, CA: Wadsworth, 1984.

[27] B. Scholkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, Cambridge, MA: MIT Press, 2002.

[28] D. Casanova, F. J. Garzón, and J. V. Busquets-Mataix, "Support vector machine-based gas turbine efficiency prediction," J. Eng. Gas Turbines Power, vol. 142, no. 6, pp. 061006, June 2020.

[29] J. Xu, M. Du, and J. Li, "Comparative analysis of SVM and neural networks for fault diagnosis in rotating machinery," Mech. Syst. Signal Process., vol. 92, pp. 226-239, July 2017.

[30] Said, B., Mazouz, L., NAAS, T.T., Yildirim, Ö. and Mohammedi, R.D., 2024. Broken magnets fault detection in PMSM using a convolutional neural network and SVM. ITEGAM-JETIA, 10(48), pp.55-62.

[31] Jiangjiao Li, Jin Han, Dapeng Niu, Xi Zhuo Jiang, Fast and accurate gas turbine emission prediction based on a light and enhanced Transformer model, Fuel, Volume 376, 2024, 132750, ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2024.132750.

[32] James, G., Witten, D., Hastie, T., and Tibshirani, R., 2013. An Introduction to Statistical Learning, Vol. 103. Springer New York, New York, NY.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **AJAPAS** and/or the editor(s). **AJAPAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

 Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).