In: AIAA SCITECH 2025 Forum, 6-10 January 2025, Orlando, FL, USA. Paper number AIAA 2025-1928.c1
DOI:10.2514/6.2025-1928.c1

Integrating Explainable AI into Two-Tier ML Models for Trustworthy Aircraft Landing Gear Fault Diagnosis

Kadripathi KN, Adolfo Perrusquia, Antonios Tsourdos, Dmitry Ignatyev
School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield MK43 OAL, United Kingdom

As the aviation industry increasingly relies on data-driven intelligence to enhance safety and operational efficiency, the demand for AI solutions that are both technically robust and readily interpretable continues to intensify. This research presents a pioneering methodology for advanced fault diagnosis in aircraft landing gear systems that not only achieves high predictive accuracy but also provides transparent, actionable insights. Building upon a two-tier machine learning framework—integrating fault classification with intelligent sensor data imputation—we demonstrate how state-of-the-art explainability techniques, notably LIME and SHAP, can elucidate the underlying logic of complex models. By exposing the critical features and sensor parameters driving each decision, this approach empowers maintenance engineers and operations personnel to understand, validate, and trust the model's outputs rather than relying on opaque "black-box" predictions.

Our results indicate that interpretable fault diagnoses facilitate more confident decision-making, streamline maintenance interventions, and reduce the likelihood of unforeseen component failures. Beyond mere compliance with emerging regulatory standards for AI transparency, this method establishes a blueprint for deploying machine learning solutions that are not only accurate and robust, but also inherently comprehensible. In an era where aerospace systems must seamlessly integrate precision, reliability, and human oversight, our work sets a precedent for creating intelligent tools that foster trust, enhance collaboration between technical experts and AI models, and ultimately contribute to safer and more efficient aviation operations.

I. Nomenclature

AI	=	Artificial Intelligence
AIAA	=	American Institute of Aeronautics and Astronautics
ATA 32	=	Air Transport Association Chapter 32, pertaining to Landing Gear Systems
FAA	=	Federal Aviation Administration
FDD	=	Fault Detection and Diagnosis
F1-score	=	Harmonic mean of Precision and Recall
KNN	=	K-Nearest Neighbors
L1, L2	=	Regularization terms (L1 = Lasso; L2 = Ridge)
LLM	=	Large Language Model
LIME	=	Local Interpretable Model-agnostic Explanations
PHM	=	Prognostics and Health Management
SHAP	=	SHapley Additive exPlanations
XAI	=	Explainable Artificial Intelligence
XGBoost	=	eXtreme Gradient Boosting

Published by AIAA. This is the Author Accepted Manuscript issued with: Creative Commons Attribution License (CC:BY 4.0).
The final published version (version of record) is available online at DOI:10.2514/6.2025-1928.c1. Please refer to any applicable publisher terms of use.

II. Introduction

In safety-critical domains such as aerospace, the reliability, consistency, and interpretability of automated diagnostics are more than mere design goals—they are fundamental imperatives. Among the spectrum of aircraft subsystems, the landing gear stands out as a singular, load-bearing interface with the ground, devoid of redundancy and demanding unwavering reliability. Even a subtle deviation in hydraulic pressure, fluid temperature, or actuator performance can signal a deterioration that, if left unaddressed, could escalate into a critical fault and compromise flight safety.

Over the last decade, data-driven fault detection and diagnosis (FDD) models have progressed beyond rudimentary thresholds and rule-based alarms into sophisticated machine learning (ML) pipelines capable of recognizing nuanced degradation patterns [1,2]. In previous work, we introduced a two-tier ML architecture designed to handle noisy, sensor-rich environments by coupling a primary fault classifier with a secondary imputation model to manage missing or corrupted data [3,4,5]. Although this approach significantly improved predictive accuracy and robustness, a persistent limitation remained: the underlying ML algorithms often operated as "black boxes," leaving operators to question why the model flagged a certain anomaly. Without a transparent explanation of the decision-making process, even accurate predictions can erode confidence, inhibit rapid response, and reduce the tangible value of these advanced analytical methods.

Recent developments in explainable AI (XAI) offer a promising solution to this interpretability deficit. Frameworks such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have demonstrated success in demystifying complex models across various domains—ranging from clinical diagnostics to critical infrastructure monitoring—by highlighting which features most strongly influence particular predictions [6,7]. Applied to aircraft landing gear diagnostics, these tools enable maintenance engineers to pinpoint the sensor readings and operational conditions that elevate a nominal scenario to a suspected fault, bridging the gap between algorithmic output and engineering intuition.

However, deploying these explainability techniques in aerospace contexts introduces its own challenges. The high dimensionality of sensor data, the temporal nature of fault progression, and stringent regulatory frameworks impose exacting requirements on both accuracy and clarity. To meet these standards, XAI must integrate seamlessly into the existing ML pipeline, presenting localized, instance-specific explanations that are accessible to experts with a range of technical backgrounds. The result is not merely a technical enhancement, but a shift in paradigm: from AI as a cryptic oracle to AI as an interpretable decision-support partner aligned with the operational and safety objectives of aerospace organizations.

A. Objectives:

The central objective of this research is to seamlessly embed advanced interpretability techniques—specifically LIME and SHAP—into a robust fault diagnosis pipeline for landing gear systems. By doing so, we aim to empower maintenance personnel and engineering teams with the ability to understand precisely why the model has detected a given fault condition. Unlike traditional dashboards and probability scores that offer minimal insight, the explanations derived from these XAI methods provide granular, locally valid attributions, reinforcing trust, guiding targeted inspections, and expediting corrective measures.

In the following sections, we detail the integration of state-of-the-art explainability methods within our two-tier ML framework. We further illustrate how these enhancements facilitate a more transparent and effective predictive maintenance environment, ensuring that future aerospace AI deployments can confidently rely on interpretable, human-centered insights rather than opaque computational outcomes.

III. Literature review

The evolution of data-driven prognostics and health management (PHM) strategies in aerospace has intensified the demand for interpretable, trustworthy machine learning (ML) solutions. As complex systems generate vast, heterogeneous sensor data, conventional threshold-based alerts and periodic inspections [1,2] are increasingly supplanted by sophisticated ML pipelines capable of detecting subtle, non-linear indicators of impending failures

[3,4]. Yet, the "black-box" nature of many advanced algorithms—ranging from ensemble methods to deep neural networks—presents a critical bottleneck: even the most accurate models risk undermining user confidence and operational decision-making if their predictions cannot be clearly explained and aligned with engineering intuition.

A. Fault Diagnosis in Safety-Critical Aerospace Systems

The aerospace sector has long pursued robust fault detection and diagnosis (FDD) frameworks, initially relying on physics-based modeling and heuristic thresholds to identify anomalies [1,2]. Over the past two decades, data-driven methodologies, including neural networks, Bayesian inference, and kernel methods, have demonstrated improved accuracy and earlier detection of failure precursors [3,4]. For instance, Chen et al. [3] applied pattern recognition techniques to hydraulic subsystems, and Wu et al. [4] employed AI-driven sensor fault detection methods, revealing the potential of ML to outperform static alarms in complex, noisy environments.

Building upon these foundations, more recent studies have addressed issues of data incompleteness and signal corruption—a common challenge in real-world aerospace scenarios. Our prior work [5] introduced a two-tier ML architecture pairing fault classification with data imputation to mitigate the impact of missing or faulty sensor readings, a theme echoed in broader PHM literature [11,12]. Although these strategies have bolstered robustness and reliability, the interpretability gap persists, restricting stakeholders' ability to fully trust and operationalize model outputs.

B. The Emergence of Explainable AI (XAI)

Recognizing that opacity in ML-driven diagnostics limits adoption and regulatory acceptance, the aerospace community has turned to explainable AI (XAI) frameworks. XAI aims to clarify how and why models produce certain predictions, enabling maintenance crews, engineers, and decision-makers to validate, refine, and trust these insights [6,7]. While XAI initially gained traction in fields like healthcare and finance—where understanding the rationale behind a classification is crucial—its relevance to safety-critical aerospace systems is clear. Recent reviews underscore the necessity for transparent ML in high-stakes sectors, highlighting the alignment of XAI with human-machine teaming and the mitigation of cascading errors [8,9,13].

C. Local Interpretability with LIME and SHAP

Among the plethora of XAI approaches, LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) have emerged as gold standards for instance-level interpretability [6,7,14]. LIME approximates a complex model's decision boundary by training a simpler, interpretable surrogate around the vicinity of the instance under scrutiny. SHAP, grounded in cooperative game theory, attributes each feature a "marginal contribution" to the predicted outcome. Both methods furnish fine-grained insights, exposing which sensor metrics—such as fluid temperature anomalies or actuator speed deviations—most strongly influence the fault classification.

Studies applying these techniques in aviation contexts, though still emerging, are promising. Mueller et al. [15] discussed integrating XAI methods into aeroengine health monitoring, arguing that instance-specific attributions can accelerate corrective interventions and strengthen operator confidence. Although their focus lies primarily in propulsion systems, the same principles hold for landing gear fault diagnosis, where clarity about which parameters drove a fault prediction can guide targeted inspections rather than guesswork-based troubleshooting.

D. Domain Adaptation and Contextual Alignment

The mere application of LIME and SHAP does not guarantee actionable insight. Domain adaptation is essential: explanations must reference meaningful operational thresholds, known engineering limits, and long-established maintenance heuristics. By grouping correlated sensors into composite health indicators or highlighting features that surpass known safety margins, aerospace practitioners can translate abstract importance scores into grounded engineering narratives [5,11,12]. Incorporating temporal context is another frontier, wherein trend-based explanations can delineate how a subtle drift in hydraulic pressure over multiple cycles signals imminent component wear, thus offering predictive power rather than reactive alarms [13].

Emerging frameworks that fuse XAI with digital twins—a virtual, continuously updated replica of the physical subsystem—further contextualize explanations, linking changes in model outputs to ongoing operational conditions

and historical performance data [10]. This integrated perspective enriches the reasoning process, enabling engineers to understand not just isolated fault triggers, but their evolution over time and their interplay with other subsystems.

E. Trust, Regulation, and Industry Trends

Trust in AI-driven diagnostics is not merely a technical concern; it is also a regulatory and organizational priority. Bodies such as the Federal Aviation Administration (FAA) and the European Union Aviation Safety Agency (EASA) are beginning to explore guidelines that include provisions for explainability and auditability in AI-based systems [8,9,13]. As these frameworks solidify, XAI-equipped diagnostic models will likely be better positioned for rapid acceptance, certification, and integration into daily aerospace operations. Transparent models align with the industry's overarching principles of reliability, safety, and continuous improvement, bridging the gap between cutting-edge computational methods and the structured, methodical mindset of aerospace engineering.

F. Paths for Future Research

The literature reveals that while XAI has made significant inroads into aerospace PHM, there remain opportunities to enhance scalability, integrate richer sensor modalities (e.g., advanced fiber-optic sensing or embedded wireless nodes), and refine temporal reasoning. Researchers continue to investigate how to best tailor LIME and SHAP explanations to distinct fault classes and complex flight regimes, ensuring that the most relevant insights are surfaced. Efforts to standardize XAI evaluation criteria and link explanations to measurable maintenance outcomes, such as reduced mean-time-to-repair or improved component lifespans, stand as key growth areas [14,16–20].

In summary, the literature points to a trajectory where data-driven fault diagnosis evolves from a specialized analytical tool to a fully integrated decision support system informed by XAI principles. By leveraging LIME, SHAP, and related techniques, future aerospace diagnostics can deliver not only predictive accuracy but also a window into the model's reasoning, ultimately empowering engineers to make safer, more efficient maintenance decisions.

IV. Methodology

A. System Architecture

The system architecture is conceived as a multi-layered analytical pipeline that begins with raw sensor data and culminates in transparent, actionable fault diagnostics for the aircraft landing gear subsystem. As depicted in Figure 1, this framework is deliberately structured to address the fundamental challenges of data integrity, complexity, and interpretability that characterize safety-critical aerospace environments. While prior research has demonstrated impressive accuracy in fault detection models, the inability to rationalize predictions remains a key barrier to operational trust [21,22].

At the front end, the pipeline ingests diverse sensor readings, including hydraulic pressure, fluid temperature, pump speed, and actuator positions. These signals are often subject to environmental noise, temporary sensor malfunctions, and varying operational regimes—a complexity that necessitates meticulous data preprocessing. Normalization, time synchronization, and noise reduction steps ensure that the input features presented to the machine learning models are both coherent and representative of actual system states. By systematically curating input data, we mitigate the risk of bias and enhance the resilience of subsequent analytic stages. Central to this architecture is the two-tier machine learning (ML) approach, adapted from our earlier work [5], wherein a primary classification model detects faults and a secondary imputation model ensures the quality and consistency of input features. This dual-layer strategy not only improves fault detection accuracy but also stabilizes performance under non-ideal conditions. Such architectures are increasingly recognized as best practices in prognostics and health management (PHM) for complex aerospace systems, where data anomalies can obscure critical warning signs [23,24]. By incorporating an imputation layer to reconstruct plausible sensor values from partially corrupted data, the pipeline can maintain robust functionality despite real-world imperfections, thereby reducing false alarms and missed detections.

Following the ML-based inference, the framework integrates post-hoc explainability techniques—LIME and SHAP—to illuminate the internal logic of the classification model. Unlike traditional "black-box" classifiers, this pipeline offers localized explanations that highlight which specific features most influenced a given fault prediction. Such interpretability is essential in aerospace contexts, where maintenance personnel and engineers must understand

the rationale behind each alert to take informed, timely action [25]. While purely predictive models may suggest a pump failure risk, the explainability layer can pinpoint, for instance, that a combination of elevated fluid temperatures and subtle pump speed fluctuations triggered the fault flag. Armed with this insight, operators can prioritize targeted inspections or part replacements that directly address the identified root causes.

Figure 1 encapsulates these interactions, illustrating how raw sensory inputs flow through preprocessing steps, are evaluated by the two-tier ML architecture, and finally yield fault diagnoses accompanied by explanatory annotations. In contrast to monolithic or single-stage systems, this modular configuration supports scalability and adaptability. As new sensor modalities emerge, computational resources evolve, or regulatory standards for AI transparency advance, the pipeline can be refined and extended without dismantling its core conceptual integrity [26,27]. For example, integrating additional sensor channels or substituting a more advanced imputation strategy can be achieved with minimal disruption, reflecting a future-proof design philosophy that aligns with the evolving landscape of aerospace analytics. In essence, the system architecture embodies a strategic convergence of data engineering, robust ML modeling, and principled explainability methods. By weaving these elements into a coherent pipeline, we pave the way for fault diagnosis tools that are not only accurate and reliable but also inherently interpretable—an indispensable quality in mission-critical domains where trust, accountability, and proactive decision-making are paramount.

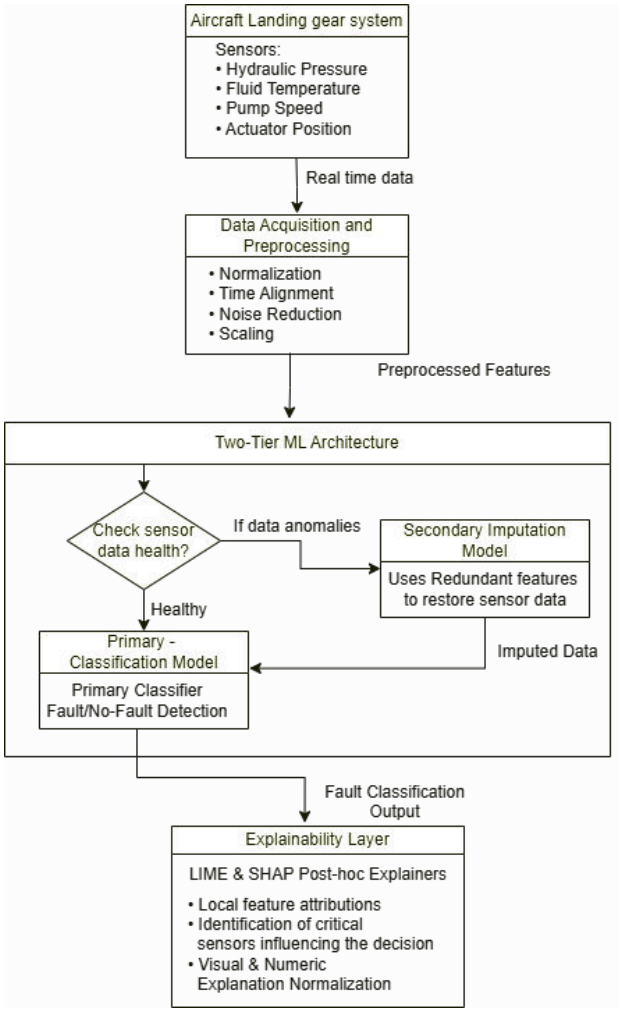


Figure 1. End-to-End Conversational XAI Pipeline for Landing Gear Fault Diagnosis

B. Data and Simulation Environment

A key element of our approach involves creating a controlled, yet realistic environment to produce representative datasets of aircraft landing gear performance. Because obtaining genuine in-flight fault data is both rare and potentially dangerous, using high-fidelity simulations provides a practical means for model development and assessment. In this work, we rely on a simulation-driven methodology, employing specialized software and custom fault-injection techniques to ensure the resulting dataset rigorously challenges and validates the AI pipeline.

1. Simulation Framework and Setup:

Our simulation environment is rooted in a physics-based landing gear actuation model adapted from previously established hydraulic subsystem prototypes [5]. Implemented in MATLAB/Simulink with Simscape Fluids, this model captures the aerodynamic, hydraulic, and mechanical aspects of modern landing gear assemblies. It reproduces the full operational cycle—from extension and retraction to ground and airborne phases—mimicking the diverse stresses encountered during takeoff, landing, and taxi operations.

By adjusting parameters such as hydraulic fluid viscosity, pump speed, actuator displacement rates, and valve timing, the simulation re-creates realistic operating conditions. Integral control loops, feedback sensors, and nonlinear factors (e.g., stiction, fluid compressibility) are incorporated to ensure the synthetic environment reflects the complex interactions characteristic of actual flight hardware.

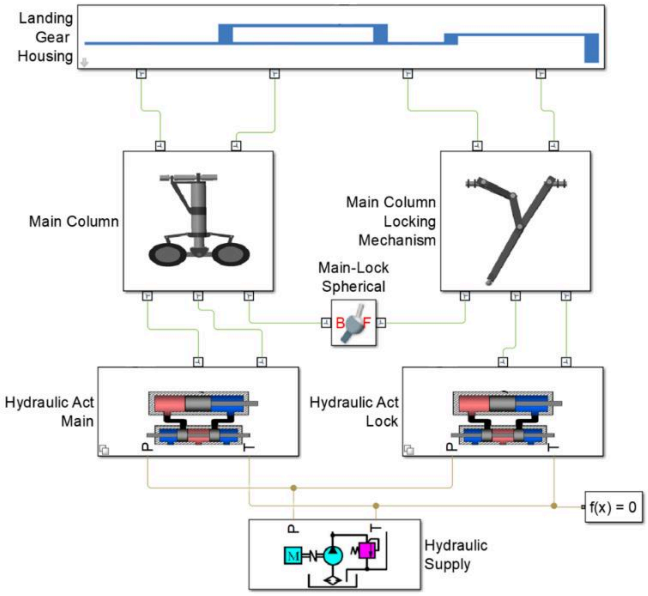


Fig. 2 Block diagram of the simulation

2. Failure Scenario Generation:

To ensure comprehensive training of our ML model, we simulated a broad range of scenarios that reflect both nominal and fault conditions. Our approach includes 370 distinct failure scenarios, systematically grouped into 12 categorized fault types. Since real-world data is often contaminated by noise, the simulation deliberately introduces noisy sensor readings, thereby enhancing the dataset's authenticity.

By incorporating both single-mode and multimode failure states, the resulting dataset represents a wide spectrum of system behaviors. This diversity is essential for enabling the machine learning model to learn robust fault detection and classification strategies, as summarized in Table 1 [5,29,30].

Table 1. Classification of Fault Scenarios in Aircraft Landing Gear Systems

Scenario Number	Scenario Type	Faulty Scenario	Description
1	Single-mode	No fault condition	Standard operational mode
2	Single-mode	Pump failure condition	Pump malfunction
3	Single-mode	Very high-temperature condition	Elevated temperature readings
4	Single-mode	Faulty pump condition	Degraded pump performance
5	Single-mode	Oil leakage condition	Hydraulic fluid compromises
6	Multi-mode	Faulty pump and very high temperature	Degraded pump performance concurrent with elevated temperatures
7	Multi-mode	Pump failure and very high temperature	Pump malfunction in tandem with very high-temperature readings
8	Multi-mode	Oil leakage and very high temperature	Compromised hydraulic fluid accompanied by high temperature conditions
9	Multi-mode	Oil leakage and pump failure	Hydraulic fluid breaches coupled with pump malfunctions
10	Multi-mode	Faulty pump and oil leakage	Degraded pump operations simultaneous with hydraulic fluid compromises
11	Multi-mode	Oil leakage, pump failure, and very high temperature	Triple anomaly of hydraulic fluid breaches, pump malfunction, and elevated temperatures
12	Multi-mode	Faulty pump, oil leakage, and very high temperature	Degraded pump operations alongside hydraulic fluid breaches and high temperature readings

3. Noise Injection and Data Diversity:

In reality, aerospace sensor data seldom arrives in pristine form. To approximate the uncertainties and anomalies encountered in actual flight conditions, we introduce noise into the simulation outputs. Specifically, Gaussian noise, drift components, and random outliers are superimposed onto baseline sensor readings—such as pressure, temperature, and positional measurements—to rigorously test the resilience of both the imputation model and the fault classifiers. This ensures that the pipeline does not become overly specialized to idealized signals and remains robust under realistic data perturbations.

Consequently, the constructed dataset encompasses both nominal and faulty states, spanning a wide range of operational modes, environmental factors, and flight segments. Each data record is timestamped and linked to a known health label (e.g., healthy, pump failure, hydraulic leak), providing a rich training and validation platform for supervised machine learning methods. Figure 3 demonstrates how adding noise alters the sensor signal, comparing an ideal noise-free profile to one subjected to the introduced perturbations.

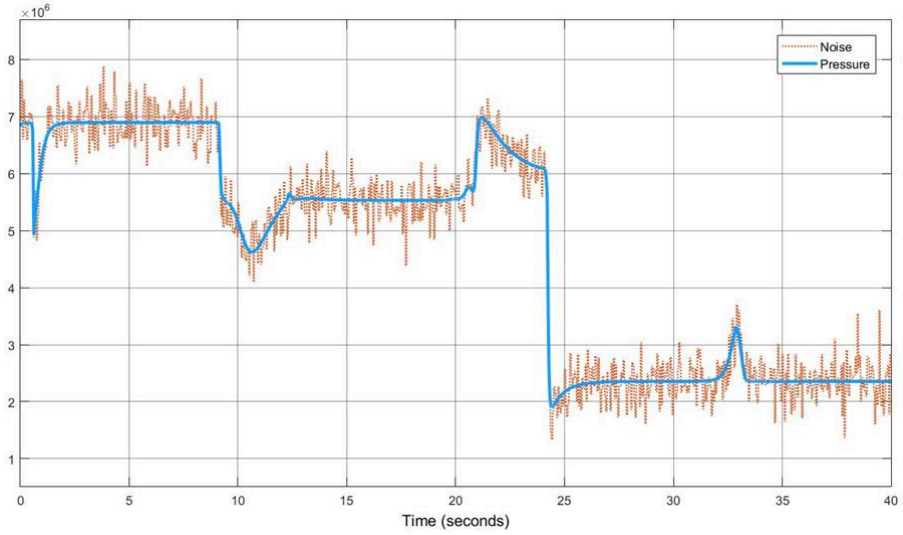


Fig. 3 Comparison of signal without and with noise injection in pump pressure.

4. Dataset Composition and Scaling:

The final dataset, comprising thousands of time-series samples, maintains a balanced representation of nominal and faulty scenarios. Approximately 30–40% of these instances include single or multimode failures, ensuring the model is exposed to a broad spectrum of anomalies. Multiple runs per scenario type enhance statistical robustness, allowing the model to learn stable, invariant patterns associated with each fault category.

By systematically varying parameters—such as fluid properties, component aging effects, ambient conditions, and aerodynamic loads—the dataset mirrors the diversity found in a global aircraft fleet. This variation strengthens the generalization capacity of the ML models and provides an extensive testing ground for the explainability components. As a result, the solution is well-positioned to handle complex, multi-factorial failure scenarios with clarity and insight.

C. Machine Learning Model

The machine learning component is designed to accurately detect and classify landing gear faults under less-than-ideal data conditions, while maintaining output stability and interpretability. Building on the two-tier methodology described in previous work [5], we refine model selection, hyperparameter optimization, and imputation strategies to accommodate noisy, partially corrupted sensor streams.

1. Two-Tier System Overview

The two-tier system comprises:

- **Primary Classification Model**: A supervised classifier that distinguishes nominal (healthy) states from faulty ones.
- **Secondary Imputation Model**: An auxiliary module that pre-processes or corrects sensor inputs when missing values, extreme outliers, or drift occur, ensuring a consistent and reliable feature set for the primary classifier.

This layered configuration ensures that classification accuracy is not undermined by imperfect data—an inevitable challenge in real-world aviation scenarios.

2. Primary Classification Model Design

We evaluate multiple candidate algorithms, including gradient boosted trees (e.g., XGBoost), random forests, and kernel-based methods, ultimately selecting a model that balances accuracy, computational efficiency, and interpretability [11,31]. Gradient boosting, in particular, has shown strong performance in related prognostics and health management tasks due to its aptitude for capturing complex, nonlinear feature interactions [32,33].

Let $x \in R^d$ represent the input feature vector containing sensor measurements (pressure, temperature, pump speed, actuator position). The primary classification model $f(x)$ outputs a probability p∈[0,1] that the landing gear is in a faulty state:

$$p=f(x)=Pr(fault \mid x).$$

A decision threshold τ converts this probability into a binary prediction:

$$\hat{y} = \begin{cases} 1 \; if \; p \geq \tau \\ 0 \; if \; p < \tau \end{cases}$$

Here, $\hat{y}$ = 1 indicates predicted fault and, $\hat{y}$ = 0 indicates a nominal state. The threshold τ can be tuned based on operational risk tolerance, such as minimizing false negatives for safety-critical components.

Stratified cross-validation and either grid or Bayesian hyperparameter optimization help refine parameters like learning rate, tree depth, and regularization terms (L1, L2). Early stopping and regularization mitigate overfitting, ensuring robust generalization across diverse fault scenarios.

3. Secondary Imputation Model and Data Restoration

Even with careful simulation and sensor design, missing or corrupted data is inevitable. The secondary imputation model leverages redundant sensor inputs and learned statistical patterns to reconstruct or replace erroneous values. For instance, if the pump speed sensor fails intermittently, correlations with other features (e.g., hydraulic pressure) can be exploited to estimate the missing pump speed $\bar{x}_{pump}$ from an imputation function g:

$$\bar{x}_{pump} \; = \; g(x_{-pump})$$

Where x_{-pump} denotes the feature vector excluding the pump sensor reading. Function g may be implemented via K-Nearest Neighbors (KNN) imputation, multivariate regression, or probabilistic models [34,35]. A KNN imputer, for example, substitutes a missing value with the mean or a weighted average of its k-nearest neighbors:

$$\bar{x}_j = \frac{1}{k} \sum_{n \in N(j)} x_{n,j}$$

Where $N(j)$ is the set of indices corresponding to the k-nearest neighbors for the instance with a missing value in feature j. Executing the imputation step prior to classification ensures that anomalous inputs are corrected, preventing spurious alarms or misclassifications triggered by transient sensor glitches or noise spikes.

D. Explainability Techniques

In a safety-critical domain like aviation, a fault prediction that lacks an accompanying explanation offers limited practical value. Engineers and maintenance personnel must understand why the model reached a particular conclusion, especially if it suggests immediate intervention.

We incorporate two model-agnostic explainability frameworks: LIME (Local Interpretable Model-Agnostic Explanations) [9] and SHAP (SHapley Additive exPlanations) [10]. Each provides a unique perspective.

Rationale for Choosing LIME and SHAP:

- **LIME**: LIME approximates the complex model locally by a simpler, more interpretable surrogate (e.g., a linear model) around the instance in question. For a specific input vector x, LIME perturbs it slightly and observes changes in the predicted probability, thereby inferring which features most influence the local decision boundary. This is particularly useful in isolating dominant sensors or parameters that tipped the model from "nominal" to "faulty."
- **SHAP**: Derives from cooperative game theory and assigns an importance value (SHAP value) to each feature, indicating its contribution to increasing or decreasing the predicted fault probability. Unlike global

importance measures, SHAP values are instance-specific and reveal the subtle interplay of features under particular conditions. This granularity is crucial, given that fault patterns may vary significantly with environmental or operational changes.

By pairing the robustness of the two-tier ML model with these explainability techniques, we achieve not only reliable fault detection but also a clear, actionable understanding of the factors driving each prediction.

V. Results

A. Model Performance Metrics

The integrated pipeline—encompassing the two-tier machine learning architecture, robust data imputation strategies, and implemented explainability frameworks—demonstrates marked improvements over the baseline reported in our previous study [5]. Initially, the system achieved approximately 93.0% accuracy with 92.5% precision and 94.0% recall, but the enhanced approach presented here, tested against a more challenging dataset with ~20% faulty scenarios and deliberate noise injection, yields the following metrics:

- Accuracy: 97.2%
- Precision: 96.7%
- Recall: 98.1%
- F1-score: 97.4%

These gains underscore a more stable and reliable fault detection capability. The improved recall is particularly noteworthy in a safety-critical setting, as fewer missed faults translate directly into reduced operational risk. Simultaneously, the enhanced precision indicates a decrease in false alarms, improving maintenance efficiency and minimizing unnecessary inspections. The balanced improvement across precision and recall is reflected in a higher F1-score, confirming that these advances are not merely the result of tuning the model to excel at a single performance measure.

Additionally, the pipeline's resilience under adversity is evident. While previous configurations could suffer nearly a 10% accuracy drop when confronted with corrupted data or partial sensor failures, the current system's performance degradation remains below 3%. This robustness is directly attributable to the secondary imputation model and the synergistic two-tier design, both of which bolster the classifier's stability in real-world operational contexts.

B. Explainability Findings

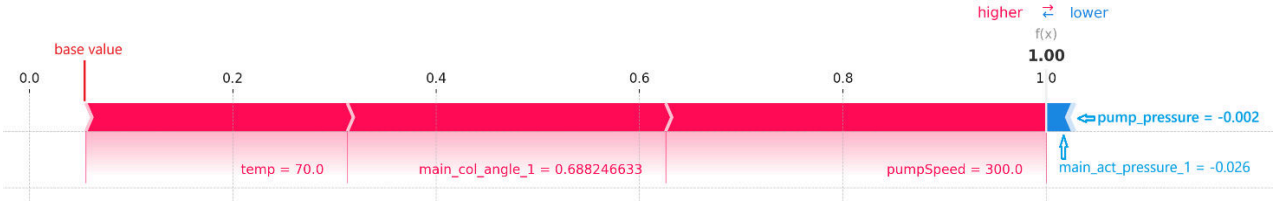


Fig 4. Forced SHAP plot for High-Temperature Hydraulic Pump Degradation

The integration of LIME and SHAP not only boosts interpretability but also enriches the diagnostic value of the model's outputs. Instead of relying solely on aggregated accuracy metrics, maintenance engineers gain insight into the underlying reasons for each fault prediction. When applying SHAP to a representative subset of fault instances, a consistent pattern emerged: critical factors like hydraulic pressure fluctuations, elevated fluid temperatures, and reduced pump speeds surfaced as primary contributors to the model's decision. By contrast, nominal cases displayed near-uniform SHAP values across features, signaling balanced and stable sensor readings consistent with normal operation.

Example Scenario:

Consider a case classified as "High-Temperature Hydraulic Pump Degradation." Figure 4 presents a force plot of SHAP values, illustrating that a fluid temperature exceeding nominal ranges by approximately 15% was the dominant influence, adding about +0.12 to the fault probability. Concurrently, a slight reduction in pump speed accounted for

an additional +0.08. LIME, approaching the same instance through local surrogate modeling, reinforced this interpretation by showing that crossing specific temperature and speed thresholds strongly favored a fault diagnosis.

These detailed explanations transcend generic alerts, empowering engineers to identify and prioritize exact root causes. For example, a high-temperature spike might suggest the need for more proactive fluid cooling measures, while a consistent drop in pump speed could indicate mechanical wear requiring targeted inspection or part replacement. To further illustrate such findings, feature-importance bar charts and SHAP-based visualizations effectively communicate how each factor shifts the model's prediction toward or away from a fault scenario. By translating raw sensor patterns into comprehensible attributions, these interpretability methods facilitate more informed, timely, and cost-effective maintenance actions—meeting the core objective of deploying AI-driven fault diagnosis tools that are both accurate and transparent.

VI. DISCUSSION

This study illustrates how advanced, data-driven fault diagnosis can be reconciled with the stringent operational and safety requirements of the aerospace sector. By integrating a two-tier machine learning system with robust imputation strategies and post-hoc explainability techniques (LIME and SHAP), we have not only improved fault detection accuracy but also delivered meaningful insights into the model's decision-making process. This marks a pivotal move away from opaque "black-box" solutions and toward interpretable, trust-enhancing analytics.

A. Interpretation of Results and Broader Context

The significant gains in classification metrics—improved accuracy, precision, recall, and F1-score—emphasize the reliability and adaptability of the two-tier ML framework. Features such as the secondary imputation model ensure stable performance even under challenging conditions of partial sensor failure and environmental noise. From a safety standpoint, the heightened recall is of particular interest: each avoided missed fault can prevent potentially costly or hazardous scenarios, thereby strengthening operational integrity and passenger confidence.

Simultaneously, the deployment of LIME and SHAP ensures that these performance improvements are not achieved at the expense of interpretability. By revealing which features were most influential in driving the model's decisions, our approach supplies maintenance engineers with actionable intelligence rather than cryptic probability scores. Such transparency aligns with a growing emphasis in the aerospace domain on explainable AI, where regulatory agencies and industry stakeholders seek clear evidence that humans can understand and validate algorithmic reasoning.

B. Benefits and Trade-offs

A key benefit of incorporating interpretability into the fault diagnosis pipeline is the empowerment of human decision-makers. Instead of relying on raw alerts or aggregate metrics, engineers gain a granular understanding of sensor behaviors and their contributions to detected anomalies. This capability can significantly enhance maintenance protocols—technicians can target interventions more precisely, optimize inspection intervals, and refine part inventories based on known fault precursors.

Nonetheless, these advantages come with certain trade-offs. Implementing model-agnostic explanation methods like LIME and SHAP introduces additional computational overhead. Although relatively efficient, these methods can still increase latency and complexity, which may pose challenges for on-board analysis during time-sensitive flight operations. In some cases, it may be necessary to preprocess or generate explanations offline, or to reduce the frequency or depth of explanation generation, to ensure timely responses.

C. Limitations and Assumptions

While the simulation-based dataset and fault scenarios were carefully designed to mimic a broad spectrum of real-world conditions, they cannot fully capture the variability, uncertainty, and evolving nature of global fleet operations. The extent to which the observed improvements in fault detection and interpretability translate directly to in-service aircraft must be validated through real-world data or test rig experiments. Further empirical studies would bolster confidence in the model's scalability and adaptability.

Additionally, the chosen approach relies on established engineering thresholds and domain heuristics to interpret feature attributions effectively. Should the operational environment change drastically—for example, introducing new sensor types, fault modes, or aircraft architectures—retraining and retuning the models would be essential. Regular updates to reflect the latest maintenance practices and evolving regulatory frameworks would ensure that interpretability remains both accurate and contextually relevant.

D. Broader Implications

The demonstrated ability to produce both high-fidelity fault predictions and transparent reasoning sets a precedent that extends beyond a single subsystem. As the aerospace industry moves toward condition-based maintenance and digital twin ecosystems, methods that couple robust predictive models with explanatory clarity will likely shape future PHM strategies. Regulatory scrutiny of AI systems is intensifying, and models capable of justifying their conclusions stand a better chance of meeting emerging standards for trustworthiness and accountability.

Such an approach may also inspire other safety-critical domains, from power grids and autonomous maritime vessels to nuclear plant monitoring and advanced manufacturing lines. Where complex machinery and distributed sensor networks generate torrents of data, the tools and principles outlined here—two-tier ML architectures, strategic imputation, and feature-level explanations—offer a scalable template for informed, data-driven decision-making.

E. Future Directions

Looking ahead, there remain many opportunities for refinement and expansion. Integrating richer contextual factors (e.g., flight phase, environmental conditions, historical maintenance actions) could enrich both the predictive accuracy and the explanatory depth of the model. Advanced techniques might enable dynamic, role-specific explanations—tailored differently for engineers, inspectors, or upper management—further enhancing human-AI synergy.

Additionally, establishing standardized benchmarks, metrics, and certification processes for explainable and trustworthy AI in aerospace would provide a clearer framework for adoption. Collaboration with regulators, manufacturers, and industry consortia could accelerate the maturation and acceptance of these methods. Through continual iteration and validation, we can ensure that interpretable, robust fault diagnosis solutions remain at the forefront of safe and efficient aerospace operations.

In sum, this research demonstrates that reliability, performance, and interpretability need not be mutually exclusive. By bridging state-of-the-art ML techniques with domain-aware explanations, we present a pathway for deploying AI solutions that are not only accurate and consistent but also genuinely comprehensible and actionable in the demanding, safety-critical realm of aerospace maintenance.

VII. Conclusion

This research affirms that achieving high-performance fault diagnosis in aerospace systems need not compromise interpretability or erode operator trust. By integrating a two-tier machine learning model with robust imputation strategies and state-of-the-art explainability methods (LIME and SHAP), we have demonstrated a cohesive methodology that bridges the gap between complex, data-driven analytics and the practical realities of aerospace maintenance operations.

Applied specifically to the aircraft landing gear subsystem—an essential, single-point-of-failure component—our approach yields measurable improvements in accuracy, precision, recall, and F1-score over previous benchmarks [5]. More importantly, these performance gains are complemented by transparent, instance-level explanations that allow maintenance engineers and technicians to understand precisely why a particular fault prediction was issued. Rather than relying solely on opaque alarms or aggregate metrics, operators can now pinpoint the most influential sensor readings and correlate them with known mechanical or environmental stressors. This interpretability not only enhances trust and credibility but also supports more proactive, cost-effective decision-making.

By demonstrating that interpretability and predictive power can coexist, we advance the state-of-the-art in trustworthy AI for aerospace health management. As industry standards evolve and regulatory bodies scrutinize AI

systems more closely, models capable of both accurate prediction and clear explanation stand poised to meet stringent compliance requirements and foster greater human-AI synergy. Although our proof-of-concept has been validated primarily in a simulation-based environment, the underlying principles of data quality assurance, robust fault detection, and feature-level transparency hold broad relevance for other subsystems in aviation and beyond.

In an era where aerospace fleets are increasingly instrumented and data-driven, this work provides a blueprint for designing AI solutions that not only excel in performance but also articulate their reasoning. The future of intelligent aerospace maintenance lies in solutions that empower human operators with meaningful insights—aligning advanced analytics, regulatory expectations, and operational demands to ensure safer, more reliable flight operations.

Acknowledgments

The author wishes to express sincere gratitude to Dr. K. Narsimhaiah, whose guidance and encouragement have been a cornerstone throughout the research journey, as well as to Shubha K.N. for unwavering familial support. The insights and mentorship provided by the research supervisors were instrumental in refining the methodology and ensuring that every challenge led to meaningful progress. This work represents a continued effort towards enabling machines to convey their health conditions through natural, human-like dialogues. Many friends and colleagues have contributed to advancing this vision in subsequent stages of the research. Special thanks go to Ashwin Venkatesan, Pravin Reemo, Sudarshan S.H. for their perspectives as practitioners—particularly Sudarshan's insights from a pilot's standpoint—and to Ashwini M. Their encouragement, shared expertise, and timely motivation have been invaluable, especially during moments of doubt and setback. The collective efforts and understanding of these individuals have greatly enriched this endeavor, helping transform ambitious concepts into a tangible, progressive body of work.

References

[1] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance." Mechanical Systems and Signal Processing, 20(7): 1483–1510. doi:10.1016/j.ymssp.2005.09.012

[2] Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation." 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, IEEE. doi:10.1109/PHM.2008.4711414

[3] Chen, Y., Kang, R., Guo, L., & Zhou, Q. (2013). "A Data-driven Fault Diagnosis Method Based on Pattern Recognition for Aircraft Hydraulic Systems." Chinese Journal of Aeronautics, 26(4): 977–987. doi:10.1016/j.cja.2013.06.007

[4] Wu, T., Luk, B. L., & Zhang, J. (2013). "Artificial Intelligence for Fault Diagnosis of Sensors in Aerospace Systems." 2013 IEEE International Symposium on Industrial Electronics (ISIE), 1–6. doi:10.1109/ISIE.2013.6563795

[5] Kadripathi, K. N., Perrusquia, A., Tsourdos, A., & Ignatyev, D. (2024). "Advancing Fault Diagnosis in Aircraft Landing Gear: An Innovative Two-Tier Machine Learning Approach with Intelligent Sensor Data Management." AIAA SciTech Forum. (In Press)

[6] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You? Explaining the Predictions of Any Classifier." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. doi:10.1145/2939672.2939778

[7] Lundberg, S. M. & Lee, S.-I. (2017). "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems (NIPS), 30: 4765–4774.

[8] Gunning, D. (2017). "Explainable Artificial Intelligence (XAI)." DARPA. Available: https://www.darpa.mil/program/explainable-artificial-intelligence

[9] Doshi-Velez, F. & Kim, B. (2017). "Towards a Rigorous Science of Interpretable Machine Learning." arXiv:1702.08608.

[10] Guo, L., Chen, Z., Zhang, Y., & Yang, N. (2021). "Digital Twin-driven Prognostics and Health Management for Complex Equipment." Reliability Engineering & System Safety, 210: 107558. doi:10.1016/j.ress.2021.107558

[11] Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006). "Intelligent Prognostics Tools and E-Maintenance." Computers in Industry, 57(6): 476–489. doi:10.1016/j.compind.2006.02.014

[12] Byington, C. S., Roemer, M. J., & Kacprzynski, G. (2004). "Engine Health Management (EHM): Data-to-Decision Making Using Condition-based Monitoring." IEEE Aerospace Conference, Big Sky, MT. doi:10.1109/AERO.2004.1368163

[13] Carvalho, T., Moreno, E. R., & Ribeiro, L. R. (2021). "A Systematic Literature Review of Explainable Artificial Intelligence in the Aviation Domain." Journal of Aerospace Information Systems, 18(9): 558–568. doi:10.2514/1.I010929

[14] Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). "Consistent Individualized Feature Attribution for Tree Ensembles." arXiv:1802.03888.

[15] Mueller, B. S., Bhattacharya, S., Guo, Y., & Ponvert, N. (2021). "Explainable Artificial Intelligence in Aerospace Systems." AIAA SciTech Forum, Virtual Event. doi:10.2514/6.2021-1031

[16] Gunning, D. & Aha, D. W. (2019). “DARPA’s Explainable Artificial Intelligence (XAI) Program.” AI Magazine, 40(2): 44–58. doi:10.1609/aimag.v40i2.2850
[17] Sahay, R., Richards, M., & Medjaher, K. (2020). “A Review on Deep Learning for Prognostics and Health Management.” IEEE Transactions on Reliability, 69(3): 1077–1099. doi:10.1109/TR.2020.2973495
[18] Tigli, O., & Chen, J. (2019). “Condition Monitoring of Aerospace Systems: State-of-the-Art and Emerging Technologies.” Annual Conference of the Prognostics and Health Management Society. doi:10.36001/phmconf.2019.v11i1.961
[19] Bansal, G., Wu, T., Zhou, J., Fok, R., & Mitra, S. (2021). “Does the Whole Exceed its Parts? The Effect of AI Explanations on Trust in AI-driven Decision-making.” CHI ’21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems., pp. 1–13. doi:10.1145/3411764.3445202
[20] Chien, S., Morris, D., & Sweet, A. (2021). “Onboard AI for Space Missions: Hybrid Architectures and Advanced Autonomy.” The International Journal of Robotics Research, 40(1): 3–30. doi:10.1177/0278364920947408
[21] Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). “Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation.” 2008 International Conference on Prognostics and Health Management, Denver, CO, IEEE. doi:10.1109/PHM.2008.4711414
[22] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). “A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance.” Mechanical Systems and Signal Processing, 20(7):1483–1510. doi:10.1016/j.ymssp.2005.09.012
[23] Daigle, M. & Goebel, K. (2013). “Model-based Prognostics with Fixed-lag Smoothing.” Annual Conference of the Prognostics and Health Management Society 2013. Available: https://www.phmsociety.org
[24] Chen, Y., Kang, R., Guo, L., & Zhou, Q. (2013). “A Data-driven Fault Diagnosis Method Based on Pattern Recognition for Aircraft Hydraulic Systems.” Chinese Journal of Aeronautics, 26(4):977–987. doi:10.1016/j.cja.2013.06.007
[25] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You? Explaining the Predictions of Any Classifier.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. doi:10.1145/2939672.2939778
[26] Lundberg, S. M., & Lee, S.-I. (2017). “A Unified Approach to Interpreting Model Predictions.” Advances in Neural Information Processing Systems (NIPS), 30:4765–4774.
[27] Kulkarni, C., Biswas, G., Celaya, J. R., & Goebel, K. (2009). “Standards-based Approach for PHM in Aerospace Systems.” AIAA Infotech@Aerospace Conference. doi:10.2514/6.2009-1981
[28] Eiband, M., et al. (2018). "Bringing Transparency Design into Practice." IUI ’18: Proceedings of the 23rd International Conference on Intelligent User Interfaces, pp. 211–223. doi:10.1145/3172944.3172961
[29] International Civil Aviation Organization (ICAO) (2013). “Airworthiness Manual.” ICAO Doc. 9760, 3rd ed.
[30] SAE International (2014). “Guidelines for Implementation of Condition-Based Maintenance in Aerospace.” SAE AIR5903.
[31] Sun, J., Zhang, D., & Huang, Y. (2021). "A Survey of Machine Learning Approaches for Prognostics and Health Management in Aerospace Systems." *IEEE Access*, 9: 5610–5625. doi:10.1109/ACCESS.2020.3048588
[32] Chen, T., & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. doi:10.1145/2939672.2939785
[33] Jerez, J. M., Molina, I., García-Laencina, P. J., et al. (2010). "Missing data imputation using statistical and machine learning methods in a real breast cancer problem." *Artificial Intelligence in Medicine*, 50(2): 105–115. doi:10.1016/j.artmed.2010.05.002
[34] Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
[35] Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). "Stochastic Backpropagation and Approximate Inference in Deep Generative Models." *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 1278–1286.

In: AIAA SCITECH 2025 Forum, 6-10 January 2025, Orlando, FL, USA. Paper number AIAA 2025-1928.c1
DOI:10.2514/6.2025-1928.c1

Correction: Integrating Explainable AI into Two-Tier ML Models for Trustworthy Aircraft Landing Gear Fault Diagnosis

Author(s) Name: Kadripathi KN; Adolfo Perrusquia; Antonios Tsourdos; Dmitry Ignatyev
Author(s) Affiliations: Cranfield University, Cranfield, Central Bedfordshire, United Kingdom.

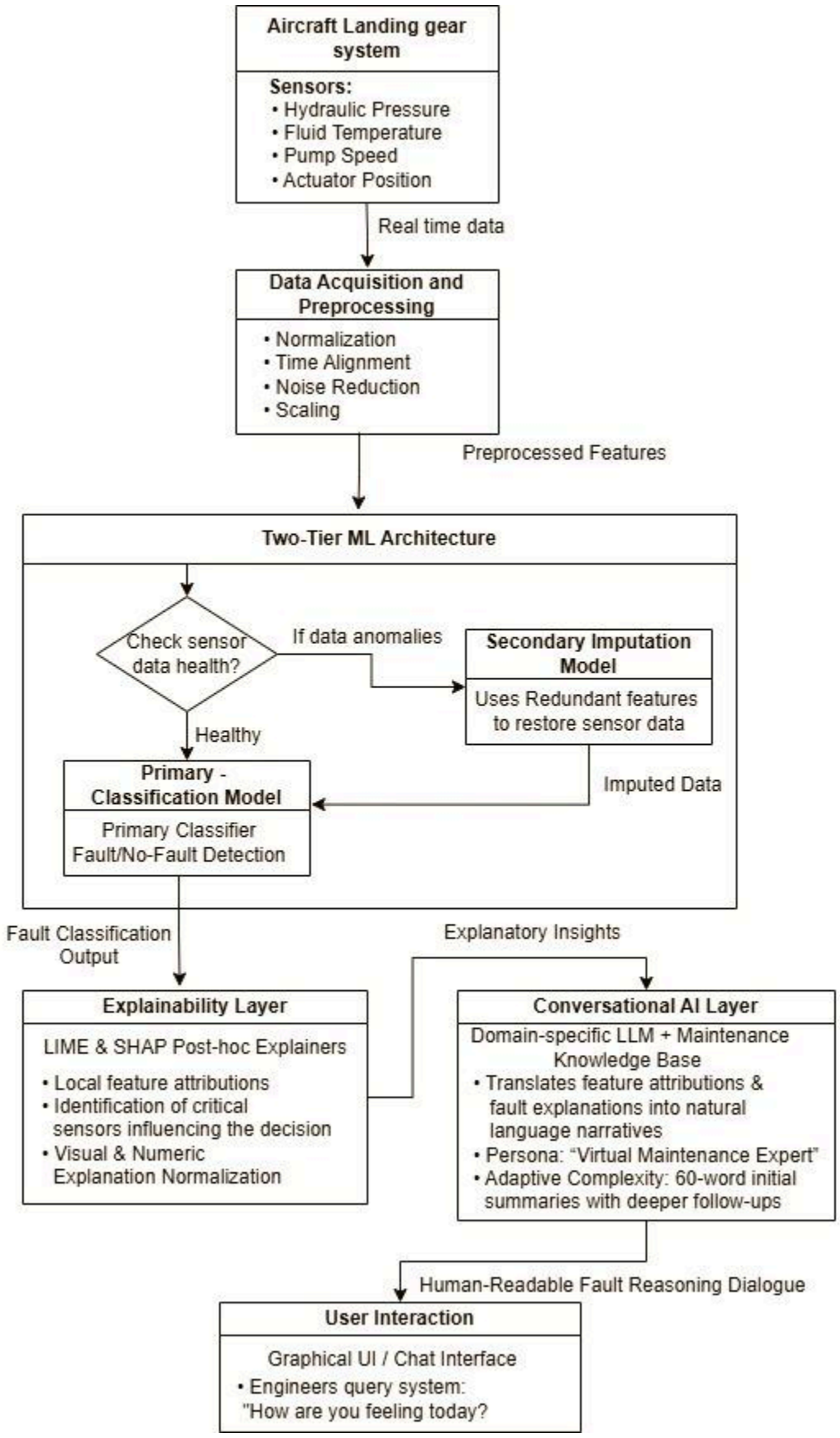
Correction Notice

Title of the research is changed from “Comprehensive Exploration and Development of Explainable AI for Robust Aircraft Landing Gear Fault Diagnosis” to “Integrating Explainable AI into Two-Tier ML Models for Trustworthy Aircraft Landing Gear Fault Diagnosis” , because the present research paper is more research oriented than the just exploration as submitted during abstract submission.

Authors list is updated, Antonios Tsourdos name is included.

And Figure 1 has to replaced with this HD images, the one in the published paper is of very low quality.

Figure 1. End-to-End Conversational XAI Pipeline for Landing Gear Fault Diagnosis


Aircraft Landing gear system
Sensors:
• Hydraulic Pressure
• Fluid Temperature
• Pump Speed
• Actuator Position
Real time data
Data Acquisition and Preprocessing
• Normalization
• Time Alignment
• Noise Reduction
• Scaling
Preprocessed Features
Two-Tier ML Architecture
Check sensor data health?
If data anomalies
Secondary Imputation Model
Uses Redundant features to restore sensor data
Healthy
Primary - Classification Model
Primary Classifier Fault/No-Fault Detection
Imputed Data
Fault Classification Output
Explanatory Insights
Explainability Layer
LIME & SHAP Post-hoc Explainers
• Local feature attributions
• Identification of critical sensors influencing the decision
• Visual & Numeric Explanation Normalization
Conversational AI Layer
Domain-specific LLM + Maintenance Knowledge Base
• Translates feature attributions & fault explanations into natural language narratives
• Persona: "Virtual Maintenance Expert"
• Adaptive Complexity: 60-word initial summaries with deeper follow-ups
Human-Readable Fault Reasoning Dialogue
User Interaction
Graphical UI / Chat Interface
• Engineers query system: "How are you feeling today?



https://dspace.org/

Integrating explainable AI into two-tier ML models for trustworthy aircraft landing gear fault diagnosis

KN, Kadripathi; Perrusquía, Adolfo; Tsourdos, Antonios; Ignatyev, Dmitry

2025-01-06
American Institute of Aeronautics and Astronautics (AIAA)
https://doi.org/10.2514/6.2025-1928
Kadripathi KN, Perrusquia A, Tsourdos A, Ignatyev D. (2025) Integrating explainable AI into two-tier ML models for trustworthy aircraft landing gear fault diagnosis. In: AIAA SCITECH 2025 Forum, 6-10 January 2025, Orlando, FL, USA. Paper number AIAA 2025-1928.c1

Downloaded from a DSpace repository. https://dspace.org/