A Cyber-Physical Testbed for Hierarchical Fault Isolation in Bilateral Systems: Multi-Domain Fusion with Attention-Derived Laterality Attribution

Jagadeesh Harinarayanan

[1]*Independent Researcher, Chicago, IL, USA*

jagadeeshharinarayanan@gmail.com

Abstract

Prognostics and Health Management (PHM) for complex cyber-physical systems is constrained by a fundamental scarcity of labeled fault data, as catastrophic failures are operationally irreproducible and ethically impossible to induce at scale. To bridge this gap, this study utilizes a dual-engine aerospace platform as a high-fidelity case study, detailing a reproducible cyber-physical testbed leveraging the DCS World (Digital Combat Simulator) Su-25T simulation environment for hierarchical fault isolation. A robust real-time telemetry architecture extracts 25 sensor channels at a 30 Hz scheduler frequency via Apache Kafka and PostgreSQL, facilitating programmatic fault injection across a structured 3×3 altitude-speed grid (1,000 ft, 8,000 ft, and 20,000 ft altitude bands; 350, 400, and 450 kts true airspeed (TAS)). The resulting dataset comprises 369 labeled missions spanning four distinct thermodynamic states: left-engine fire, right-engine fire, dual-engine fire, and a nominal baseline. Under a validated injection protocol with randomized onset (T + 60s to T + 80s), two research questions are addressed.

RQ1 evaluates multi-domain sensor fusion through an XGBoost framework trained on a 789-feature vector encompassing time-domain statistics and cross-channel asymmetry indicators.

RQ2 explores temporal fault laterality via a PyTorch Transformer encoder processing 450-timestep post-fault sequences. Model interpretability is achieved through SHapley Additive exPlanations (SHAP) domain contribution analysis and Attention-derived Domain Activation Profiles (DAPs). The XGBoost pipeline achieves high diagnostic performance (Macro F1 = 0.9914) , while the Transformer achieves Macro F1 = 0.9784 for laterality identification, maintaining robustness via leave-2-out cross-validation (mean F1 = 0.975).

A novel methodological finding reveals that the Transformer identifies failure lateralization by attending to stable healthy-side channels, suggesting that stable healthy-side signals provide a lower-variability contrastive reference for laterality discrimination in bilateral systems.

1. Introduction

Prognostics and Health Management (PHM) for complex, bilateral cyber-physical systems faces a fundamental data acquisition challenge: catastrophic failure events are rare by design, operationally irreproducible, and ethically impossible to induce on occupied platforms at the scale required for deep supervised learning. The consequence is a persistent industry-wide reliance on degradation models derived from bench tests, reduced-order analytical models, or cross-platform transfer assumptions—none of which adequately reproduce the integrated multi-domain sensor response of an in-flight engine fire across varied aerodynamic conditions.

Simulation-based testbeds offer a structured alternative. The NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset established that simulation-generated labeled degradation trajectories are a legitimate foundation for PHM model development (Saxena et al., 2008). More broadly, the systematic use of modern, physics-based flight simulators for multi-domain fault injection across a true flight-envelope grid, coupled with real-time streaming data pipelines, has not been widely reported in contemporary PHM literature.

This work exploits the DCS World Su-25T simulation environment as a controlled cyber-physical testbed. The aircraft's integrated Lua scripting interface enables automated programmatic fault injection, while its avionics model produces correlated multi-domain sensor responses—spanning thermal, mechanical, aerodynamic, and motion parameters—that are physically consistent with real-world dual-engine fire phenomenology. A key scope condition is acknowledged upfront: the thermodynamic representation of engine fire propagation in DCS World is simplified relative to certified engine test data. The primary contribution is methodological, not parameter-calibrated.

Transformer architectures have demonstrated exceptional performance on multivariate time series classification in PHM contexts (Wu et al., 2020; Zerveas et al., 2021), driven by their ability to model long-range temporal dependencies via multi-head self-attention. However, the interpretability of attention-based models in fault diagnosis remains an open question. Raw attention weights alone are not sufficient as explanations; this work addresses that gap by combining averaged attention weights with input magnitude to construct Attention-derived Domain Activation Profiles (DAPs)—an interpretable approximation of channel-level importance, which we validate through consistency across leave-2-out folds (Section 5.4) rather than a formal faithfulness guarantee.

The contributions of this work are fourfold:

(1) Engineering of a reproducible cyber-physical fault injection framework producing 369 labeled flights across a 3×3 altitude-speed grid via a real-time 30 Hz Kafka–PostgreSQL telemetry pipeline;

(2) An XGBoost pipeline achieving Macro F1 = 0.9914 with SHAP domain contribution analysis confirming physics-consistent feature importance hierarchy;

(3) A Transformer-based fault laterality classifier achieving Macro F1 = 0.9784, with Attention-derived Domain Activation Profiles revealing that fault identification is driven by healthy-side channel contrast—a novel and previously unreported laterality mechanism; and

(4) Generalization validation via leave-2-out cross-validation across all nine flight-envelope conditions, yielding a mean generalization Macro F1 of 0.975.

2. CYBER-PHYSICAL SIMULATION TESTBED

2.1 Platform Description and Aerodynamic Fidelity

The Su-25T serves as a twin-turbofan platform within the DCS World simulation environment, characterized by high-fidelity thermodynamic modeling of independent engine states and bifurcated hydraulic architectures. Its integrated avionics suite, which includes master warning (MW) logic, operates at a 30 Hz scheduler frequency to generate multi-domain sensor outputs that exhibit significant physical correlation. Notably, the cyber-physical model captures the asymmetric aerodynamic phenomenology resultant from unilateral thrust loss, including uncommanded lateral roll moments, yaw deviations, and structured airspeed decay. Such cross-domain coupling is paramount for the development of robust diagnostic frameworks capable of hierarchical fault laterality isolation.

A definitive scope condition regarding simulation fidelity is acknowledged. While the DCS World thermodynamic representation of fire propagation—encompassing thermal gradients and RPM attenuation—remains physically plausible, it is simplified compared to certified production turbofan data. Consequently, the primary contribution of this research is methodological rather than parameter-calibrated, focusing on the architectural integrity of the diagnostic pipeline.

2.2 Real-Time Telemetry Pipeline Architecture

Sensor data is extracted from the DCS World simulation engine via a custom Lua export script hooking into the aircraft avionics interface at the scheduler frequency. Each tick produces a nested JSON payload containing all available sensor channels, which is published to an Apache Kafka topic. A Python Kafka consumer deserializes each payload via a recursive flattening algorithm, collapsing nested sensor JSON objects into a flat dictionary, which is written to a PostgreSQL table (raw_telemetry within su25t_telemetry_v1) using JSONB storage.

Post-mission, records are extracted from JSONB to tabular format and stored as Parquet files in a Medallion data lake (bronze → silver → gold layers). End-to-end pipeline latency is maintained under 100 ms per tick. This architecture mirrors operational aircraft health monitoring patterns and supports straightforward adaptation to live edge-inference sensor feeds.

3. DATASET GENERATION AND FEATURE ENGINEERING

3.1 Flight Envelope Grid and Dataset Stratification

The cyber-physical testbed generated a dataset of 369 unique labeled flights. Fault flights were executed across a 3×3 altitude-speed grid: altitude bands at 1,000 ft (low-altitude maneuvering), 8,000 ft (medium-altitude cruise), and 20,000 ft (high-altitude operations); speed tiers at 350, 400, and 450 kts TAS. For each of the nine grid cells, all fault types were executed. This grid forces classifiers to generalize across drastically different air densities and control-surface dynamic pressures rather than memorizing altitude- or speed-specific fault signatures.

The 369 flights were partitioned into training and test sets, stratified across the grid. Table 1 presents the balanced experimental design.

Table 1. Dataset stratification by mission type.

Mission Type	Train Flt.	Test Flt.	Total Flt.
Fault-Free Baseline	126	27	153
Left Engine Fire	56	16	72
Right Engine Fire	58	14	72
Dual Engine Fire	55	17	72
Total	**295**	**74**	**369**

3.2 Fault Injection Protocol and Sensor Mapping

The experimental design utilizes four distinct classification categories: left-engine fire (L), right-engine fire (R), dual-engine fire (D), and a fault-free baseline (B). These engine fire conditions are programmatically initiated through the DCS World scripting framework, which triggers the corresponding thermodynamic states within the avionics model. This results in physically correlated sensor fluctuations, including thermal spikes, RPM degradation, hydraulic system pressure drops, and integrated master warning (MW) system activation. A rigorous fault injection protocol was applied across the 369-flight dataset to ensure repeatability: (i) autopilot remains active from takeoff through T+57 s to stabilize the flight trajectory; (ii) manual control is assumed at T+57 s followed by a 3-second stabilization period; (iii) the fault event is randomly injected between T+60 s and T+80 s; and (iv) the MW system flag, synchronized with the onset of the fire, establishes the canonical temporal reference point (T = 0) for subsequent data analysis.

Empirical validation determined that the T+57 s autopilot disengagement is critical; earlier transitions resulted in a 140 m pitch-up deviation prior to fault onset, introducing significant aerodynamic artifacts. Six sensor channels were omitted from the modeling phase due to hardware-specific limits or simulation constants (e.g., the radar altimeter reaches saturation at 1,502 m AGL, providing no informative variance for the medium- and high-altitude bands). The final feature space comprises 25 usable channels, categorized into four physical domains as detailed in Table 2. For the Transformer-based sequence classification task (RQ2), the left and right fuel flow sensors were further excluded; exploratory data analysis revealed that their variance was dominated by digital throttle response noise rather than thermodynamic engine fire phenomenology, resulting in 23 channels for temporal sequence modeling.

Table 2. Physical domain mapping of sensor channels.

Physical Domain	Sensor Channels (RQ1: 25 channels; RQ2: 23 channels)
Thermal	sensors_temp_l, sensors_temp_r, sensors_fire_l, sensors_fire_r
Mechanical	sensors_rpm_l, sensors_rpm_r, sensors_hyd_fail
Aerodynamic	sensors_aoa, sensors_roll, sensors_pitch, sensors_tas, sensors_alt_baro, sensors_vsi, sensors_mach, sensors_g_load
Motion	sensors_roll_rate, sensors_pitch_rate, sensors_yaw_rate, sensors_yaw, sensors_vel_x, sensors_vel_y, sensors_vel_z, sensors_master_warn

3.3 Feature Engineering

For RQ1, a 5-second window was selected to span multiple oscillation cycles of the fastest-varying channels (e.g., yaw rate), while a 2.5-second stride balances temporal resolution against the computational and multiple-testing cost of finer strides is applied to each flight at 30 Hz (150 discrete timesteps per window). For each window, 30 per-channel time-domain statistics are computed across all 25 channels (750 features), producing descriptors including mean, standard deviation, RMS, skewness, kurtosis, slope, peak-to-mean ratio, zero-crossing rate, and inter-quartile range. Additionally, 39 physically motivated cross-channel asymmetry features are computed per window. For any bilateral sensor pair the normalized asymmetry index is defined as:

$$\Delta s = (sL - sR) / (sL + sR + \varepsilon) \tag{1}$$

The term ε represents a minimal constant designed to mitigate division-by-zero errors occurring during thermodynamic engine shutdown phases. Significant asymmetry indicators encompass RPM differentials, exhaust gas temperature variances, fire system activation disparities, and cross-channel interaction products, such as the product of RPM decay and thermal elevation. This feature engineering process yields a high-dimensional vector of 789 features for each sliding window, resulting in a comprehensive tabular dataset comprising 20,131 observations. As detailed in Table 3, the dataset exhibits a pronounced class imbalance, with the fault-free baseline accounting for more than 71% of the total extracted windows.

Table 3. RQ1 sliding-window dataset class distribution.

Class Label	Windows	% Dataset	Test Support
normal (Class 0)	14,319	71.1%	2,747
l_engine_fault (Class 1)	1,530	7.6%	345
r_engine_fault (Class 2)	1,597	7.9%	342
dual_engine_fault (Class 3)	2,685	13.3%	709
Total	**20,131**	**100.0%**	**4,143**

For RQ2, the T = 0 to T+15 s post-fault window is extracted from each of the 216 fault flights (72 per class), producing a fixed-length PyTorch tensor of shape (216, 450, 23).

4. Algorithmic Methodology

4.1 RQ1: Multi-Domain Fault Diagnosis via Gradient Boosting

The XGBoost multi-class framework (Chen & Guestrin, 2016) undergoes optimization using a 30-iteration RandomizedSearchCV protocol targeting the hyperparameter manifold {n_estimators, max_depth, learning_rate, subsample, colsample_bytree, min_child_weight, gamma}, validated through 5-fold stratified cross-validation. The optimal configuration encompasses: n_estimators = 300, max_depth = 7, learning_rate = 0.01, subsample = 0.9, colsample_bytree = 0.9, min_child_weight = 1, and gamma = 0.3. To mitigate the pronounced class imbalance documented in Table 3, inverse-frequency sample weights are integrated during the iterative tree-fitting process. The learning objective is defined by the minimization of the softmax multi-class logarithmic loss function:

$$L = -1/N \times \Sigma i\ \Sigma k\ yik \cdot \log(\hat{p}ik) \tag{2}$$

In this formulation, y_{ik} serves as the binary indicator for membership of sample i in class k, whereas $\hat{p}_{ik}$ represents the softmax probability predicted by the model ensemble. Post-training, the SHAP TreeExplainer (Lundberg & Lee, 2017) is utilized on the independent test partition to derive feature-level attribution scores, facilitating an evaluation of the physics-consistent importance hierarchy within the learned model.

4.2 RQ2: Fault Laterality via Transformer Attention

The PyTorch Transformer architecture utilizes a linear input projection (23 → 64 dimensions), sinusoidal positional encoding, and two stacked Transformer encoder blocks. Each block features 4-head multi-head self-attention (d_model = 64, dim_ff = 128, dropout = 0.1). The self-attention operation is:

$$Attention(Q, K, V) = softmax(QK^T / \sqrt{dk})\ V \tag{3}$$

Following global average pooling, a linear classifier maps the 64-dimensional latent representation to 3 output classes. Training used a cross-entropy loss with Adam optimizer (lr = 1e-3), batch size 32, early stopping with patience 15 on validation loss (best model at epoch 10).

For interpretability, feature importance was derived directly from the model's native self-attention mechanism rather than relying on backpropagated gradients. Specifically, the multi-head self-attention weights were extracted during the forward pass and averaged across all attention heads and encoder layers. This process collapsed the attention matrices to isolate the specific temporal windows most heavily weighted by the network. To determine physical channel importance, the raw input sequence was multiplied by these aggregated attention scores. The resulting attention-weighted inputs were then grouped into six lateralized domains (Thermal_L, Thermal_R, Mechanical_L, Mechanical_R, Aerodynamic, Motion) to produce the Attention-Derived Domain Activation Profile (DAP).

5. Results

5.1 RQ1: Fault Detection and Classification Performance

The XGBoost classifier achieves Macro F1 = 0.9914 on the held-out test set of 4,143 windows. Per-class performance is highly consistent: normal (F1 = 1.00), l_engine_fault (0.99), r_engine_fault (0.99), dual_engine_fault (0.99). The confusion matrix (Figure 1) demonstrates near-perfect separation; the isolated misclassifications occur exclusively in the ambiguous pre-fault transition windows within the 5-second horizon immediately preceding MW activation.

Two ablation experiments confirm result integrity. First, the non-overlapping window evaluation (stride = window size = 5 s) yields F1 = 0.9924 on 2,094 windows, demonstrating that window overlap does not inflate the reported score. Second, retraining on baseline-only normal windows (excluding pre-fault windows from fault flights) yields F1 = 0.9925, confirming no class boundary contamination from the two-population normal class.

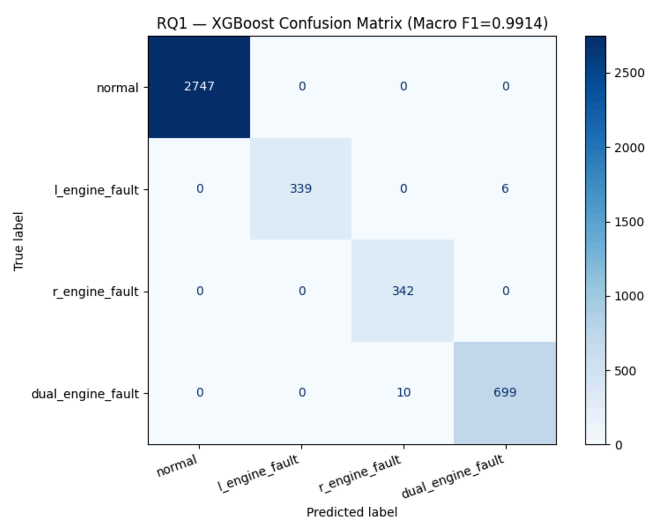

Figure 1. XGBoost confusion matrix on held-out test set.

SHAP domain contribution analysis (Figure 2) reveals the physical causality hierarchy learned by the model. For the normal class, mean |SHAP| values across all four fault-discriminative domains are effectively zero, reflecting that nominal flight identification is driven by the collective absence of fault-specific signals rather than positive activation of any domain; this produces the structureless attribution profile visible in Figure 2. For fault classes, the

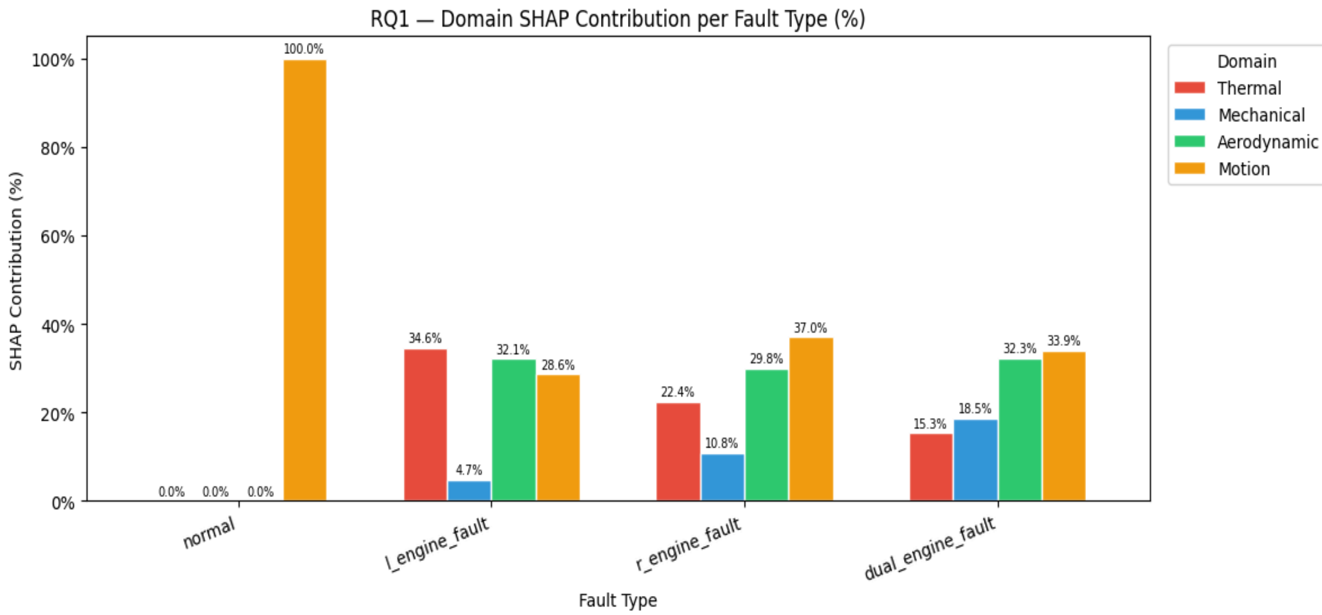

Figure 2. SHAP domain contributions by fault class (% of total mean |SHAP|)

Thermal domain dominates for l_engine_fault (34.6%) and r_engine_fault (22.4%) predictions, consistent with the direct thermal signature of engine fire. The top individual feature globally is fire_asymmetry (mean |SHAP| = 0.635), followed by sensors_master_warn__mean (0.454) and fire_both_active (0.346). SHAP shift analysis at fault onset shows Thermal domain contribution increases by +314.5% from pre-fault to post-fault windows, with Mechanical following at +54.4%—confirming the fire-to-RPM-decay propagation sequence.

No direct benchmark comparison is reported because, to our knowledge, no existing published baseline addresses bilateral fault-laterality detection on a comparably structured simulated flight-envelope testbed; the closest related settings (e.g., C-MAPSS-style degradation benchmarks) do not include a laterality dimension. The internal ablations in Section 5.1 (non-overlapping window evaluation, F1 = 0.9924; baseline-only retraining, F1 = 0.9925) serve as consistency checks in the absence of an external benchmark.

5.2 RQ2: Identification of Fault Laterality via Transformer Architectures

The PyTorch Transformer architecture achieves a Macro F1 of 0.9784 across the hierarchical laterality isolation task. This performance is validated against a held-out test partition comprising 47 missions, stratified into 16 left-side, 14 right-side, and 17 dual-engine fire events. Categorical analysis reveals optimal precision for unilateral left-side failure modes (F1 = 1.00), while maintaining high-fidelity identification for right-side and bilateral thermodynamic states (F1 = 0.97 each). Examination of the resultant confusion matrix (Figure 3) indicates that diagnostic discrepancies are localized to the decision boundary separating dual-engine signatures from unilateral events. These outliers are physically consistent with observed phenomenology; the simultaneous triggering of bilateral thermal telemetry during dual-engine onset produces a high-entropy state that precedes the development of structured aerodynamic asymmetry indicators.

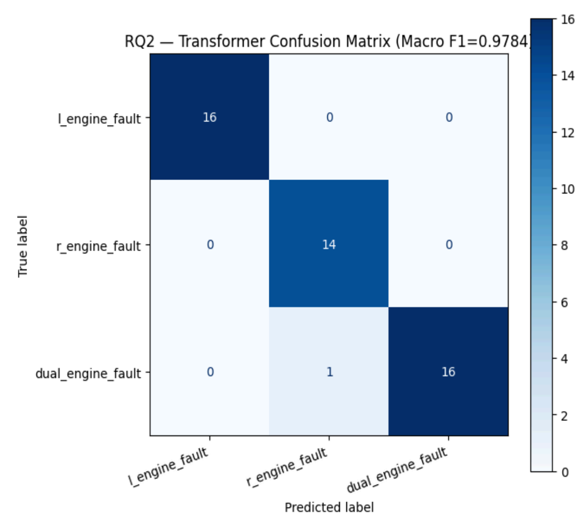

Figure 3. Transformer confusion matrix on fault laterality test set.

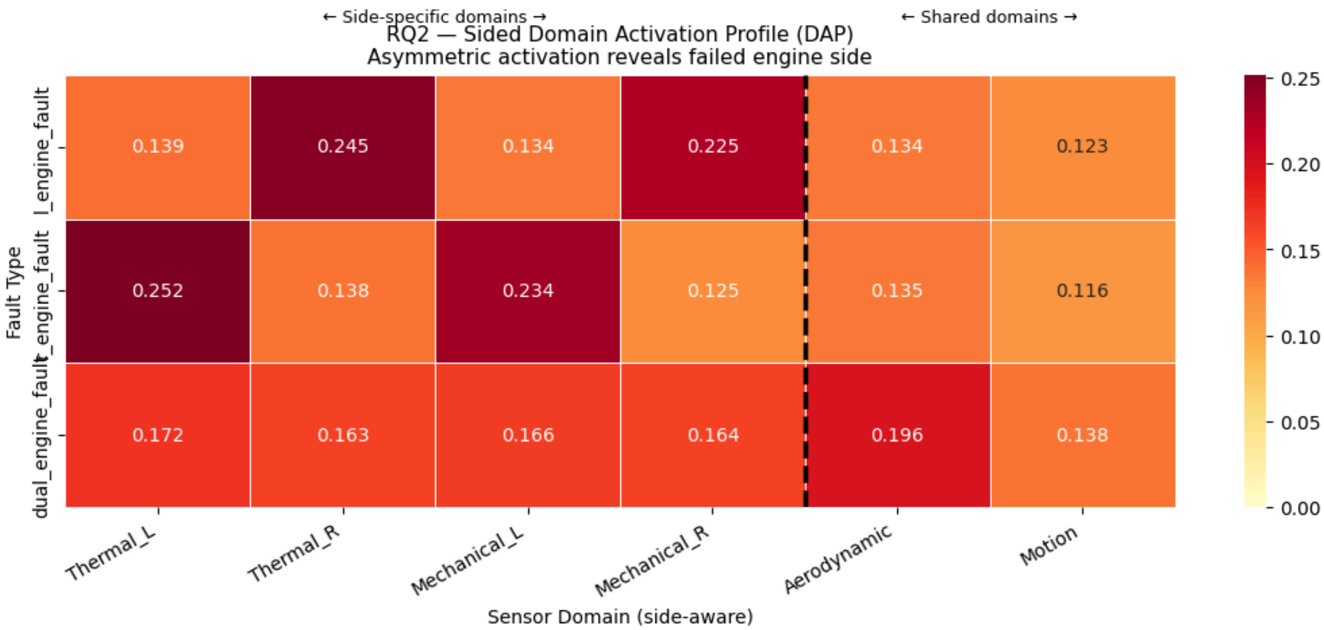

Figure 4.Attention derived Sided DAP. The heatmap shows that steady, healthy-side channels receive peak attention during unilateral failures, driving fault laterality decisions.

5.3 Attention-Derived Domain Activation Profiles

The Attention-derived DAP (Figure 4) constitutes the central empirical finding of this research. Table 4 presents the normalized attribution scores for the thermal and mechanical lateralized domains.

For l_engine_fault, the largest attention attributions are concentrated in Thermal_R (0.245) and Mechanical_R (0.225)—the healthy right-side channels—rather than in Thermal_L (0.139) and Mechanical_L (0.134), corresponding to the actively degrading engine. This pattern is precisely mirrored for r_engine_fault: Thermal_L (0.252) and Mechanical_L (0.234) dominate over their right-side counterparts. For dual_engine_fault, attribution is distributed near-symmetrically across Thermal_L (0.172) and Thermal_R (0.163), consistent with bilateral degradation. The thermal laterality asymmetry index for l_engine_fault is −0.277 (healthy side dominance) and for r_engine_fault is +0.293 (healthy side dominance).

Table 4. Attention-derived DAP normalized thermal and mechanical attributions per fault class.

Fault Type	Thermal _L	Thermal _R	Mech _L	Mech _R
l_engine_fault	0.139	0.245	0.134	0.225
r_engine_fault	0.252	0.138	0.234	0.125
dual_engine_fa ult	0.172	0.163	0.166	0.164

The DAP profiles also reveal substantial attribution to the Motion (0.123–0.138) and Aerodynamic (0.134–0.135) domains for unilateral faults. However, SHAP onset analysis (Section 6.2) shows that these domains exhibit near-zero change at fault onset: Motion contribution decreases by 27.1% post-fault, and Aerodynamic increases by only 1.5%. This contrasts sharply with the Thermal domain surge of +314.5%. The high static attribution of Motion and Aerodynamic reflects their role as stable flight-state encoding channels — they carry high overall information content but do not discriminate fault presence or laterality. The Transformer correctly weights them as contextual background. The fault-discriminative signal is concentrated in the Thermal and Mechanical asymmetry, confirming that laterality identification is driven by the differential between left- and right-side channels rather than by aggregate motion state.

5.4 Cross-Envelope Generalization via Leave-2-Out Validation

To evaluate architectural robustness across varied aerodynamic regimes, a leave-2-out (L2O) cross-validation protocol was executed, encompassing all 36 permutations of held-out operating conditions within the 3x3 grid. The XGBoost framework maintains a mean Macro F1 of 0.961 (ranging from 0.936 to 0.979), empirically confirming that the diagnostic pipeline prioritizes fundamental fault phenomenology over the memorization of altitude- or speed-specific sensor artifacts across the 1,000–20,000 ft and 350–450 kts envelope.

Crucially, the Transformer architecture exhibited parallel robustness during L2O evaluation, achieving a mean Macro F1 of 0.975 (ranging from 0.936 to 1.000) on the hierarchical laterality task. This confirms that the underlying Attention-Derived DAP mechanism—specifically the model's reliance on healthy-side sensor stability—is an invariant thermodynamic property of the bilateral system rather than an overfit artifact of specific altitude or airspeed configurations.

6. DISCUSSION

6.1 Contrastive Basis for Healthy-Side Laterality

The central finding—that the Transformer identifies the failed engine by attending to the stable healthy-side channels—can be explained via information theory. When the left engine fires, sensors_temp_l experiences a violent non-linear spike and sensors_rpm_l undergoes rapid, highly variable decay. From an information-theoretic perspective, this high-entropy signal contains substantial diagnostic value for determining that a fault has occurred (explaining Thermal domain dominance in RQ1). However, this high-entropy, chaotic signature has low discriminative value for laterality in a symmetric system: catastrophic fire in the left engine produces a signature that, in isolation, is difficult to distinguish from fire in the right engine.

Conversely, the healthy engine telemetry remains in a low-entropy, highly deterministic state. The Transformer's multi-head self-attention inherently seeks correlated, deterministic patterns to minimize classification loss over the 450-timestep sequence. The model appears to exploit this asymmetry: when global aerodynamic consequences (uncommanded roll, yaw rate anomalies in sensors_roll_rate and sensors_yaw_rate) are present while the right-side thermal and mechanical channels remain aligned with stable flight, the learned representation associates this pattern with a left-side fault. The healthy side provides the contrastive reference frame the classifier learns to rely on for laterality discrimination—mirroring the comparative exclusion logic used by human operators: the engine that does not show the anomaly is healthy.

This mechanism is structurally analogous to differential diagnosis in medical diagnostics and the use of control channels in fault detection and isolation (FDI) theory (Isermann, 2006) . It implies an important design principle for attention-based PHM in symmetric systems: healthy-side sensors are not redundant—they function as a stable contrastive reference baseline that the model exploits to resolve laterality, complementing rather than replacing the diagnostic value of the fault-side signal itself.

Analytical redundancy and parity-relation methods in classical FDI theory similarly exploit a known-good reference—typically a model-predicted value—against which a faulty channel is compared (Isermann, 2006). The present finding extends this principle in two ways: the reference is a physical, sensed healthy-side channel rather than a model prediction, and the comparison is learned implicitly through attention rather than specified analytically. This positions the DAP mechanism as a data-driven analogue to redundancy-based FDI rather than an unrelated construct.

6.2 Physics-Consistent Attribution and Interpretability

The SHAP analysis for RQ1 validates that the learned feature importance hierarchy is physically consistent with engine fire causality. Primary fault indicators, including fire_asymmetry and sensors_master_warn__mean, exert dominant global influence. The Thermal domain's contribution (34.6% for left-side failures) represents the initial ignition event, followed by mechanical thrust decay and subsequent airframe aerodynamic response. The observed +314.5% surge in Thermal domain SHAP values at fault onset confirms that the classifier accurately tracks the fire-to-airframe propagation sequence rather than spurious correlations.

6.3 Simulation Constraints and Scalability

Performance metrics (F1 > 0.97) are necessarily bounded by the structured nature of the DCS World simulation relative to real-world environments containing electromagnetic interference and sensor drift. However, the underlying healthy-side laterality mechanism is a structural property of symmetric physical systems driven by entropy differentials. Consequently, this mechanism is expected to remain robust under noisy conditions. Future research should prioritize hardware-in-the-loop (HITL) validation and the integration of real-time edge inference on live telemetry streams.

7. Conclusion

This study establishes a reproducible cyber-physical testbed for dual-engine fault isolation utilizing the DCS World Su-25T platform. Leveraging a 369-flight dataset and a real-time Kafka telemetry architecture, the XGBoost framework achieved a Macro F1 of 0.9914, while a Transformer encoder demonstrated exceptional laterality identification (Macro F1 = 0.9784). Generalization was confirmed via L2O cross-validation across the flight envelope grid.

Crucially, the Attention-derived Domain Activation Profiles indicate that fault localization is driven by attention to stable, healthy-side channels. This counter-intuitive mechanism implies that healthy-side sensor fidelity is a first-order requirement for deep laterality discriminators in symmetric cyber-physical systems. The work demonstrates that high-fidelity PHM model development and interpretability research can be effectively supported by simulation-derived labeled data in domains where operational fault data is physically impossible to collect.

Nomenclature

Δs	normalized bilateral asymmetry index
sL	left-side sensor channel value
sR	right-side sensor channel value
ε	regularization constant for asymmetry stability
L	softmax cross-entropy loss
N	number of training samples
yik	binary class membership indicator, sample i class k
$\hat{p}$ik	softmax predicted probability, sample i class k
Q, K, V	query, key, value projection matrices
dk	key projection dimension
T	number of timesteps per sequence window
yc	pre-softmax logit for class c
AGL	above ground level
AoA	angle of attack
DAP	domain activation profile
DCS	digital combat simulator
FDI	fault detection and isolation
HITL	hardware-in-the-loop
Hz	hertz
kts	knots (nautical miles per hour)
L2O	leave-2-out cross-validation
MW	master warning
PHM	prognostics and health management
ReLU	rectified linear unit
RPM	revolutions per minute
RQ	research question
SHAP	SHapley additive exPlanations
TAS	true airspeed

References

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* (pp. 1–9). IEEE. doi:10.1109/PHM.2008.4711414

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*.

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021). A transformer-based framework for multivariate time series representation learning. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 2114–2124). doi:10.1145/3447548.3467401

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

Isermann, R. (2006). *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer.

Biography

Jagadeesh Harinarayanan was born in Tamil Nadu, India. He received the B.E. degree in Mechanical Engineering from Anna University, Chennai, India, and the M.S. degree in Business Analytics from Robert H. Smith School of Business, University of Maryland, College Park, MD, USA, in 2024, with a GPA of 3.76.

He has approximately four years of professional experience across the United States and India as a Data Engineer, with expertise spanning Azure, Snowflake, Databricks, dbt, Apache Kafka, and PostgreSQL. He currently works as a contract Data Engineer based in Chicago, IL, USA. He holds the Microsoft Certified: Azure Data Engineer Associate (DP-203) credential.

His research interests include prognostics and health management for industrial and systems engineering, simulation-based fault data generation, multivariate time series classification, transformer architectures for PHM, and the application of stochastic degradation modeling to maintenance scheduling optimization. He is pursuing independent research toward a doctoral program, targeting systems engineering and industrial engineering programs with a focus on PHM and decision-making under uncertainty.