AAAI Spring Symposium Series (SSS 2026)

# Causal Learning for Fault and Anomaly Detection in Unmanned Aerial Systems

**Atul Rawal**

Research Institute for Tactical Autonomy, Howard University, Washington D.C. USA

## Abstract

Unmanned Aerial Systems (UAS) are vital to tactical autonomy applications, where they operate in contested, communication denied, and high stakes environments with limited human oversight. In such contexts, robust fault detection is not merely a safety concern but a mission critical capability where engine failures or control surface malfunctions can result in asset loss, mission failure, or compromised operational security. Traditional anomaly detection methods rely on correlation-based approaches that identify statistical deviations but fail to reveal the underlying causal mechanisms driving failures. This opacity is particularly problematic for tactical autonomy, where operators and commanders require explainable diagnostics to maintain situational awareness and trust in autonomous systems. We argue that causal learning provides a more principled and interpretable framework for understanding UAS faults. By utilizing causal learning to identify the cause-and-effect relations from telemetry data we can answer not only whether a fault occurred, but what caused it, enabling root cause analysis essential for rapid field assessment and informed decision making. We present preliminary proof-of-concept results using the ALFA (AirLab Failure and Anomaly) dataset to learn causal structures from UAS telemetry data. Our causal graphs reveal directional relationships between sensor readings, control inputs, and failure states, offering actionable insights for predictive maintenance, real time fault diagnosis, and post mission analysis that correlation-based methods cannot provide. We position causal reasoning as a foundational capability for trustworthy tactical autonomy.

## Introduction

Unmanned Aerial Systems (UAS) have become vital assets for tactical autonomy applications, supporting intelligence, surveillance, reconnaissance (ISR), logistics, and contested environment operations(Criollo et al. 2024; Hassan & Rogers 2025) . Unlike commercial or civilian applications where failures result in inconvenience or financial loss, UAS failures in tactical applications carry much severe operational consequences such as compromised missions, lost assets, and potential exposure of sensitive capabilities. The U.S Department of War's (DoW) increasing reliance on autonomous systems demands fault detection mechanisms that are not only accurate but also interpretable and trustworthy.

Current approaches to UAS anomaly detection predominantly employ machine learning methods trained to recognize statistical patterns associated with failures (Wang et al. 2021). While effective at classification, these methods treat the system as a black box, highlighting correlations without revealing why failures occur. This limitation has critical consequences for tactical autonomy. First, these methods offer limited interpretability as human operators cannot understand the causal chain leading to failure, hindering trust and complicating rapid decision making in mission critical applications (Manning et al. 2004). Second, correlation-based models generalize poorly and may fail when encountering novel failure modes not represented in training data, which is a significant risk in adversarial environments where failures may result from enemy action or unforeseen conditions (Rawal et al. 2025). Additionally, without causal knowledge, maintenance remains reactive rather than preventive, unable to target root causes, which reduces fleet readiness and increases logistical burden in forward deployed contexts.

To address these challenges, we propose the utilization of causal discovery methods to identify and highlight the directed causal structure among UAS telemetry variables. Rather than highlighting the correlation based features, causal learning can identify true causal effects of different variables, with profound implications for trustworthy autonomous systems. This shift from correlation to causation aligns with broader DoW initiatives emphasizing explainable AI (XAI) and human machine teaming, where operators must understand and appropriately calibrate trust in autonomous system recommendations. This paper provides an overview of causal learning along with preliminary results for a proof-of-concept example of fault detection using causal discovery.

## Overview of Causal Learning

Causality describes the relationship between an underlying cause and the resulting effect (Rawal et al. 2025) . It forms a foundational concept across scientific inquiry, where causes explain why an outcome occurs and effects describe what is observed. Although causality is frequently mistaken for correlation, the two are not equivalent: correlation captures statistical associations among variables, while causation characterizes a direct cause–effect mechanism. In AI/ML contexts, causal learning focuses on analyzing how changes in one variable influence model predictions when another variable is modified. Here the variable that is actively manipulated is referred to as the treatment, while the variable whose response is measured is known as the outcome. Additional background variables are called covariates, and variables that simultaneously influence both the treatment and the outcome are identified as confounders.

Causal relationships between variables are organized according to Judea Pearl's three level causal hierarchy: association, intervention, and counterfactuals (Pearl 2009, 2010). The first level, association, captures statistical dependencies present in observational data and is the basis of traditional AI/ML models, which rely primarily on correlations without explicit causal structure. The second level, intervention, examines how outcomes change when specific variables are actively manipulated, enabling causal effects to be identified through the underlying causal graph. The highest level, counterfactuals, integrates both association and intervention to reason about hypothetical scenarios by answering questions about what would have happened under alternative conditions. Within AI/ML systems, causal learning can be utilized via two complementary methods: causal inference and causal discovery. Causal inference estimates the magnitude and direction of causal effects, while causal discovery seeks to uncover the causal structure governing relationships among variables in the data.

Two principal formal frameworks are widely used to study causality in both causal inference and causal discovery: structural causal models (SCMs) and the potential outcomes framework. Structural causal models provide a unified and expressive formalism for causal reasoning by combining a causal graph with a set of structural equations (Guo et al. 2020; Pearl 2009, 2010). The causal graph encodes directed relationships among variables including treatments, outcomes, confounders, and covariates where directed edges represent causal influence. For example, an edge $x \longrightarrow y$ denotes that changes in $x$ causally affect $y$ (Guo et al. 2020). A wide variety of methods exist for learning causal structure directly from data through causal discovery. These approaches fall into three categories: constraint-based, score-based, and functional causal model (FCM) based methods (Malinsky & Danks 2018). Constraint-based and score-based approaches infer causal graphs using statistical relationships such as conditional independence tests or optimization of scoring criteria, while FCM-based methods explicitly model causal mechanisms by estimating parameters of structural equations (Yao et al. 2021).

## Proof-of-Concept Example

To highlight the utilization of causal learning for tactical autonomy applications we present a simple proof-of-concept example to highlight the causal variables that impact engine failure for UAS.

### Data Source

The dataset used in this study is derived from the AirLab Failure and Anomaly (ALFA) open-source flight dataset, which provides high fidelity, time synchronized telemetry from real unmanned aerial vehicle (UAV) flight experiments (Keipour et al. 2021). The ALFA dataset contains multi modal sensor and control data recorded during nominal and fault induced flight scenarios, including engine failure events, making it well suited for studying safety critical behavior in autonomous systems. For this work, we focus on flight sequence containing a controlled engine failure, enabling detailed analysis of system dynamics before, during, and after the failure event. The dataset includes telemetry streams such as airspeed, attitude, velocity, altitude, inertial measurements, control inputs, and flight controller state information, all recorded at high temporal resolution. The original ALFA flight data is distributed across multiple CSV files, each corresponding to a distinct telemetry topic (e.g., airspeed, attitude, velocity, inertial measurements, control inputs). To create a clean dataset, all relevant telemetry streams were time aligned and merged using a nearest neighbor join based on a common timestamp. This process ensured consistent synchronization across heterogeneous sensor modalities while preserving the temporal ordering of events. A binary engine failure indicator (fault_flag) was derived from the engine failure status topic, marking the transition point at which the failure occurs. Rows preceding the failure event were labeled as nominal operation, while rows following the event were labeled as post failure. Missing values introduced during synchronization were removed to ensure numerical stability for downstream causal and machine learning analyses.

From the merged dataset, a subset of variables was selected based on their physical relevance to engine performance, aircraft dynamics, and control behavior. These variables include commanded airspeed to capture pilot or autopilot power demand, measured airspeed and forward velocity to represent aerodynamic response, relative altitude to characterize energy state, and attitude measurements to reflect stability and control dynamics. These variables were chosen to enable both causal discovery and causal inference, allowing us to distinguish mission phase correlations from physically meaningful precursors to engine failure. By restricting the analysis to a compact, interpretable set of flight variables, the resulting dataset supports transparent reasoning about cause-and-effect relationships in safety critical autonomous systems.

| Variable | Description |
|---|---|
| Engine Failure | Binary indicator representing an engine failure. |
| Throttle | Throttle command controlling engine power. |
| Thrust/RPM | Estimated propulsion. |
| Vel_meas | Longitudinal body frame velocity measurement. |
| Airspeed | Measured airspeed from onboard sensors. |
| Airspeed_cmd | Target airspeed. |
| Pitch | Measured pitch angle. |
| Lift | Derived lift approximation. |
| Altitude | Aircraft altitude above ground level. |
| Roll_meas | Measured roll angle. |
| Yaw_meas | Measured yaw angle. |
| Roll/Yaw Stability | Lateral and directional stability |
| Gyro x/y/z | Angular velocity along axes. |
| Acc x/y/z | Linear acceleration along axes. |

Table 1: Dataset variables and description.

Several intermediate variables used in the causal analysis are not directly measured in the ALFA dataset and are therefore approximated using physically motivated proxies derived from available telemetry. Roll/Yaw Stability is computed as a scalar instability index obtained from measured roll and yaw attitude angles together with inertial measurements, aggregating angular rate magnitudes from gyroscope signals and lateral accelerations to capture lateral–directional instability. Lift is approximated using a scaled aerodynamic proxy based on measured airspeed and pitch angle, assuming pitch is reported in radians, and defined as $\hat{L}(t) = (\text{airspeed\_meas}(t))^2 \sin(\text{pitch\_meas}(t))$, which preserves the dominant dependence of lift on airspeed squared and effective angle of attack without requiring aircraft specific parameters. Thrust and propeller RPM are approximated using a control demand proxy derived from measured throttle command and commanded airspeed, defined as $\hat{T}(t) = 0.5\,\text{rc\_ch3}(t) + 0.5\,\text{airspeed\_cmd}(t)$, capturing relative propulsion loading imposed by both low level actuation and high level speed demand. All derived variables are dimensionless and used solely as interpretable intermediate constructs for causal modeling rather than direct physical measurement

## Methods

To infer the causal structure underlying engine failure events in autonomous flight operations, we employed a constraint-based causal discovery approach using the Peter Clark (PC) algorithm. The PC algorithm is well suited for safety critical autonomy settings as it constructs causal graphs directly from conditional independence relations observed in the data while making explicit assumptions about causal sufficiency and faithfulness. The goal of this analysis was to recover the causal graph that captures plausible cause effect relationships between propulsion, aerodynamic, kinematic, and stability variables and the binary outcome Engine Failure.

The PC algorithm identifies the causal skeleton by iteratively testing conditional independence between pairs of variables given subsets of remaining variables. We employed Fisher's Z test for conditional independence, which is appropriate for continuous variables under the assumption of approximate linear Gaussian relationships. Given sample size constraints and to reduce false positives, the maximum conditioning set size was bounded. Independence tests were performed at a significance level of $\alpha = 0.05$, balancing sensitivity and robustness. To ensure physical plausibility and operational realism, domain knowledge from flight dynamics and autonomy engineering was used to guide interpretation of the resulting graph. Edges violating known physical laws were excluded from final visualization, while directionality consistent with both data and domain knowledge was retained.

## Results & Discussion

Figure 1 presents the resulting causal graph for the dataset highlighting the features impacting engine failure for UAS in tactical autonomy applications. It highlights a structured decomposition of engine failure risk into propulsion, aerodynamics, and stability driven pathways, reflecting the underlying physics and control logic of autonomous flight. Rather than exhibiting a single dominant causal chain, Engine Failure emerges as a downstream outcome influenced by multiple interacting subsystems. This multi path structure is consistent with real world autonomy failures, where degradation in propulsion, loss of lift, or instability can independently or jointly precipitate failure events. Importantly, the graph highlights Engine Failure as a sink node receiving directed influence from several intermediate state variables, reinforcing that failure is often not an instantaneous event but the culmination of upstream state degradation. This structure supports post hoc analysis, counterfactual reasoning, and early warning detection by identifying which subsystems exhibit causal influence prior to failure onset.

One dominant causal pathway proceeds from Throttle ⟶ Thrust/RPM ⟶ Airspeed ⟶ Lift ⟶ Altitude ⟶ Engine Failure. This chain reflects the physical dependency between propulsion commands, generated thrust, forward velocity, aerodynamic lift, and the ability to maintain altitude(Ghimire et al. 2023; Wang et al. 2023). Disruptions along this pathway such as thrust loss or insufficient airspeed propagate forward, reducing lift and altitude margins and increasing the likelihood of engine failure or mission termination (Çuhadar & Dursun 2015). The presence of both airspeed measurement and velocity measurement feeding into the Airspeed node indicates that airspeed is causally informed by multiple sensor modalities, highlighting redundancy in state estimation. From a trustworthiness perspective, this suggests that degraded or biased measurements in

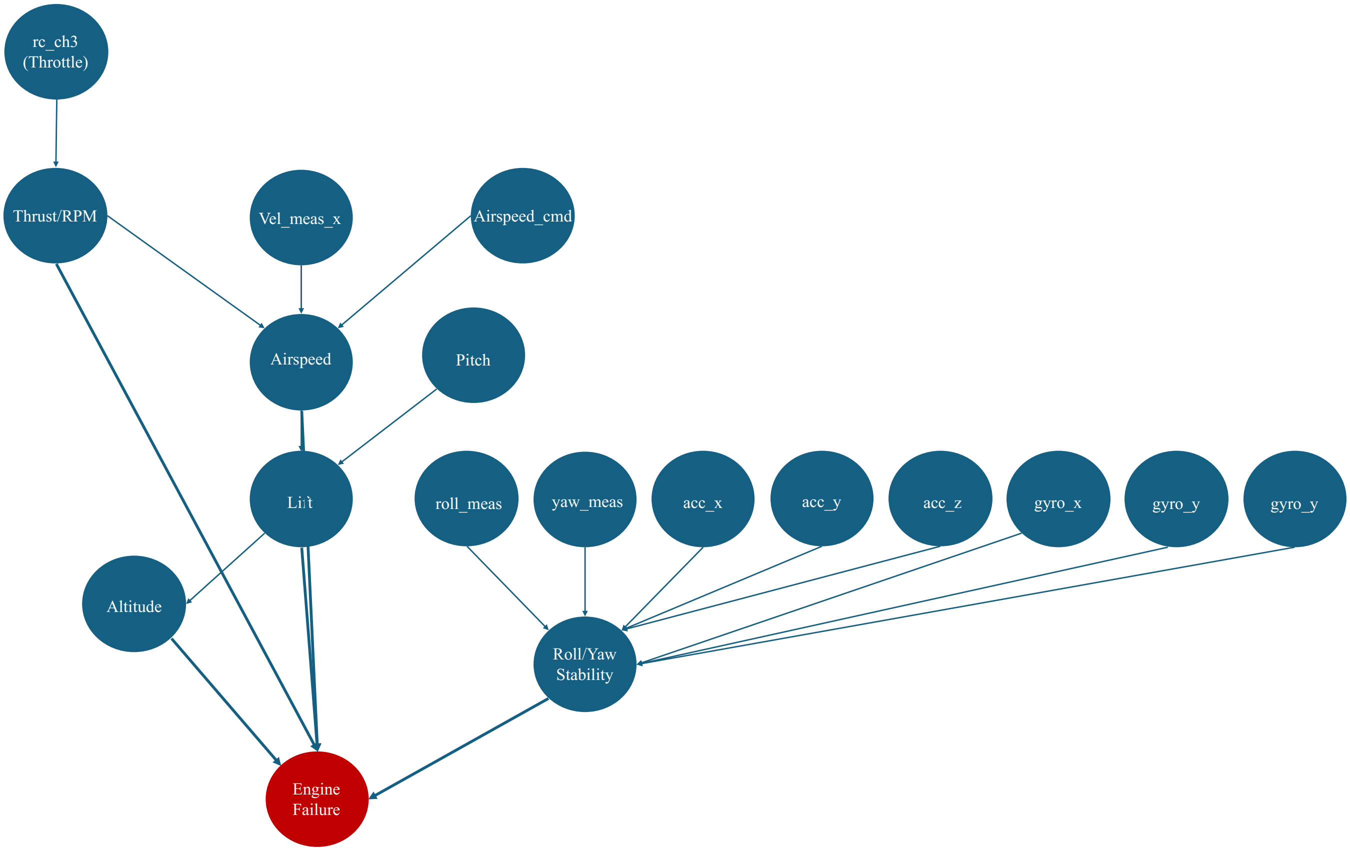

Figure 1: Causal graph highlighting the variables affecting engine failure for UAS.

one modality may still propagate risk through shared downstream variables, emphasizing the importance of causal monitoring of propulsion effectiveness.

The second major causal mechanism highlighted through the Roll/Yaw Stability node, which aggregates inertial and attitude signals such as gyroscopes, accelerometers, roll and yaw measurements. The direct edge from Roll/Yaw Stability to Engine Failure indicates that loss of lateral or directional stability constitutes an independent failure mode, even in the absence of immediate propulsion loss. This is particularly relevant for autonomous systems in contested environments, where control authority loss or oscillatory behavior can rapidly escalate into unrecoverable states. Notably, the graph separates pitch and lift from roll/yaw stability, suggesting that longitudinal and lateral dynamics contribute to failure through distinct causal mechanisms. This separation enables targeted intervention strategies. For example, stabilizing yaw dynamics without modifying propulsion while preserving interpretability of the autonomy stack.

Overall, the causal graph provides a mechanistic explanation of how engine failure can arise from different subsystems, rather than treating failure as a black box classification outcome. By explicitly modeling causal dependencies, the graph supports explainable failure attribution, counterfactual analysis and proactive risk mitigation. For tactical autonomy, this structure enables early warning and decision support, allowing autonomy managers to identify which causal pathway is deteriorating and intervene before failure manifests. More broadly, the graph demonstrates how causal discovery can bridge sensor level telemetry and mission level safety outcomes, advancing trustworthy, interpretable, and operationally relevant AI for autonomous systems.

While the proposed causal analysis provides interpretable insights into the mechanisms contributing to engine failure, several limitations must be acknowledged. First, the causal graph is learned from observational data and therefore depends on the standard assumptions of the PC algorithm, including causal sufficiency, faithfulness, and the absence of hidden confounding. In operational autonomous systems, unobserved factors such as environmental disturbances, actuator degradation, software scheduling effects, or sensor calibration drift may influence system behavior but are not explicitly captured in the available telemetry (Hedayati et al. 2024). Violations of these assumptions can lead to missing, spurious, or misoriented edges in the inferred causal structure. Additionally, the PC algorithm identifies a Markov equivalence class of causal graphs rather than a single uniquely identifiable DAG. Although domain knowledge from flight dynamics and autonomy engineering was used to interpret and constrain edge directions, certain causal relationships remain statistically indistinguishable given the data. As a result, the inferred graph should be interpreted as a plausible causal explanation consistent with both observed dependencies and physical constraints, rather than a definitive ground truth representation of the system.

An additional limitation arises from the use of derived and latent variables that were approximated from raw sensor measurements. Higher level constructs such as lift, thrust/RPM, and roll/yaw stability were computed using simplified, physically motivated approximations rather than direct measurements. These latent variables aggregate multiple signals and implicitly assume stable aerodynamic coefficients, linearized dynamics, and noise free sensor fusion. In reality, nonlinear effects, cross axis coupling, and sensor biases may violate these assumptions, introducing uncertainty into both the causal discovery process and subsequent inference (Guo et al. 2025). Errors in these approximations may propagate through the causal graph and influence the inferred strength or directionality of downstream relationships.

Finally, the analysis is limited to a single engine failure sequence and platform configuration, which constrains the generalizability of the results. The estimated causal relationships may differ across aircraft types, autonomy stacks, control architectures, or environmental conditions. Moreover, temporal aggregation and synchronization of high frequency telemetry may obscure fast transient effects or feedback loops that are critical during failure onset. Future work should incorporate multi mission analyses, interventional or simulation based validation, and uncertainty aware causal discovery methods to improve robustness and applicability in operational settings.

## Conclusion

This work demonstrates how causal learning can be applied to autonomous flight telemetry to move beyond black box failure detection toward mechanistic, interpretable understanding of engine failure in tactical autonomy systems. By learning a physically grounded causal graph from observational data, the proposed approach exposes multiple interacting pathways that contribute to failure outcomes. Rather than treating engine failure as an isolated classification event, the resulting causal structure reveals how upstream control commands and state degradations propagate through the system and culminate in mission level safety risks. The causal graph provides a transparent and actionable representation of system behavior that supports explainable diagnostics, counterfactual reasoning, and principled intervention analysis. For tactical autonomy applications, this enables early identification of which subsystem is causally responsible for rising failure risk, allowing autonomy managers or supervisory agents to select targeted mitigation strategies such as thrust reallocation, speed envelope adjustment, or stability recovery. Importantly, causal learning separates intent from capability, a distinction that is critical in contested or degraded operational environments where sensors, actuators, or control authority may be compromised.

Beyond failure analysis, this work highlights the broader role of causal learning as an enabling technology for trustworthy and resilient tactical autonomy. By providing interpretable causal models aligned with physical dynamics, causal learning offers a foundation for robust decision support, safety assurance, and adaptive autonomy under uncertainty. While future work is needed to extend these results across platforms, missions, and failure modes, this study demonstrates that causal learning can serve as a unifying framework for understanding, diagnosing, and mitigating risk in complex autonomous systems operating in high consequence environments.

## References

Criollo, L., Mena-Arciniega, C., & Xing, S. 2024. Classification, military applications, and opportunities of unmanned aerial vehicles. *Aviation*, *28*(2), 115-127.

Çuhadar, İ., & Dursun, M. 2015. The Aspects, Reasons and Outcomes of an Unmanned Air Vehicle Crash Caused By Engine Failure. *International Journal of Aerospace System Engineering*, *2*(1), 1-5.

Ghimire, R., Corbetta, M., & Palanisamy, R. P. 2023. Fault Detection and Diagnosis Techniques for Electric UAV Powertrain System. 2023 IEEE Aerospace Conference,

Guo, A., Mu, S., Zhou, Z., & Tang, J. 2025. Propagation Characteristics of Measurement Error of Low-Cost Sensors in Solar-Powered Unmanned Aerial Vehicle (UAV) Flight Control. *International Journal of Aerospace Engineering*, *2025*(1), 8247533.

Guo, R., Cheng, L., Li, J., Hahn, P. R., & Liu, H. 2020. A survey of learning causality with data: Problems and methods. *ACM Computing Surveys (CSUR)*, *53*(4), 1--37.

Hassan, W. A. Z. W., & Rogers, R. A. R. P. 2025. The Evolution of the Use of Unmanned Aerial Vehicle (UAV) in Modern Warfare and Its Multi-Faceted Impact. *Jurnal Wacana Sarjana*, *9*(3), 15-34.

Hedayati, M., Barzegar, A., & Rahimi, A. 2024. Fault diagnosis and prognosis of satellites and unmanned aerial vehicles: A review. *Applied Sciences*, *14*(20), 9487.

Keipour, A., Mousaei, M., & Scherer, S. 2021. Alfa: A dataset for uav fault and anomaly detection. *The International Journal of Robotics Research*, *40*(2-3), 515-520.

Malinsky, D., & Danks, D. 2018. Causal discovery algorithms: A practical guide. *Philosophy Compass*, *13*(1), e12470.

Manning, S. D., Rash, C. E., Leduc, P. A., Noback, R., & McKeon, J. 2004. The role of human causal factors in US

Army unmanned aerial vehicle accidents. *US Army*, *20040409*, 013.

Pearl, J. 2009. Causal inference in statistics: An overview. *Statistics surveys*, *3*, 96--146.

Pearl, J. 2010. The foundations of causal inference. *Sociological Methodology*, *40*(1), 75--149.

Rawal, A., Raglin, A., Rawat, D. B., Sadler, B. M., & McCoy, J. 2025. Causality for trustworthy artificial intelligence: status, challenges and perspectives. *ACM Computing Surveys*, *57*(6), 1-30.

Wang, J., Liu, Y., & Song, H. 2021. Counter-unmanned aircraft system (s)(C-UAS): State of the art, challenges, and future trends. *IEEE Aerospace and Electronic Systems Magazine*, *36*(3), 4-29.

Wang, Y., Liu, H., & Liu, K. 2023. Improved Thrust Performance Optimization Method for UAVs Based on the Adaptive Margin Control Approach. *Mathematics*, *11*(5), 1176.

Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., & Zhang, A. 2021. A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *15*(5), 1--46.