Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Interventional Causal Structure Discovery for Domain-Invariant Operating-Condition Classification in Multimodal Robotic Manufacturing

Md Omar Al Javed
Matthew Taylor
Vivekanand Naikwadi
Ayantha Senanayaka
asenanayaka@tntech.edu

Tennessee Tech University https://orcid.org/0009-0006-6255-3673

Research Article

Keywords: Condition-Based Maintenance, Causal Discovery, Robotic Manufacturing, Multimodal Sensing, Out-of-Distribution Generalization

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10671249/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Interventional Causal Structure Discovery for Domain-Invariant Operating-Condition Classification in Multimodal Robotic Manufacturing

Md Omar Al Javed[1], Matthew Taylor[2], Vivekanand Naikwadi[2], Ayantha Senanayaka[2]*

[1]Mechanical and Nuclear Engineering, Tennessee Technological University, Cookeville, TN 38505, US

[2]Manufacturing and Industrial Systems Engineering, Tennessee Technological University, Cookeville, TN 38505, US

*Corresponding author. Email: asenanayaka@tntech.edu

Abstract

Reliable condition-based maintenance (CBM) in smart manufacturing is severely limited by associative overfitting, where traditional models learn volatile statistical correlations that collapse under out-of-distribution (OOD) operational shifts. When routine changes in manufacturing system's speed, trajectory geometry, or payload alter a system's mechanical impedance, standard classifiers mistake environmental baseline shifts for genuine changes in machine state, generating unreliable operating-condition labels. This study introduces the Interventional Causal Framework for Condition-Based Maintenance (IC-CBM), which transitions predictive manufacturing analytics from statistical correlation to structural causation. Using the constraint-based Peter-Clark (PC) algorithm supported by bootstrap resampling, IC-CBM recovers the physical causal graph governing multimodal vibration and acoustic sensor streams from a controlled factorial interventional design. The recovered graph isolates the kinematic velocity backbone and cross-modal acoustic radiation mechanism as high-stability causal pathways. Structural Residual Normalization (SRN) then derives a targeted architectural correction directly from the recovered graph, subtracting mass-induced acoustic inflation along the identified causal path. The operating condition evaluated here is defined by a speed threshold, an operational proxy for mechanical stress rather than a measured fault or degradation state, so this study validates IC-CBM as a domain-invariant operating-condition classifier, a foundational step toward causally grounded CBM. Under a 10-trial bootstrapped held-out domain evaluation, the residual-optimized causal framework achieved 88.89% out-of-distribution accuracy, statistically indistinguishable from its in-distribution performance and outperforming the full-feature correlational baseline by 20.00 percentage points. By compressing the feature space by 84.61%, the framework achieves a high-stress-condition classification precision of 1.00 with a recall of 0.67 under severe payload covariate shift, meaning every flagged high-stress cycle is genuine, though roughly a third of true high-stress cycles are conservatively classified as standard. These results demonstrate that grounding predictive manufacturing analytics in interventionally discovered causal structure yields classification representations that generalize reliably to unseen operational contexts without sensor recalibration or data relabeling.

Keywords: Condition-Based Maintenance; Causal Discovery; Robotic Manufacturing; Multimodal Sensing; Out-of-Distribution Generalization

1. Introduction

As Industry 4.0 and cyber-physical systems rapidly advance, modern smart manufacturing environments increasingly depend on predictive analytics to enhance operational uptime, improve efficiency, and implement condition-based maintenance (CBM) [1]. Industrial robotic workcells handle complex multi-axis tasks under various operational constraints. These workcells are essential hubs where unexpected component failures or operational stresses can lead to significant downtime on the factory line [2]. To prevent these disruptions, data-driven diagnostic architectures have been widely adopted, using high-dimensional, multimodal sensor arrays, including acoustic microphones and vibration accelerometers, to track machine health continuously [3]. Despite the high accuracy achieved by modern Deep Learning (DL) and Machine Learning (ML) models in standard manufacturing environments, existing predictive analytics approaches suffer from a fundamental flaw known as associative overfitting [4–7].

Conventional data-driven models depend entirely on statistical correlations, which aggregate a variety of multi-sensor features linked to specific operational anomalies or high-stress operating conditions [8, 9]. However, these statistical relationships are highly variable and unstable [10]. In an industrial setting, external operating conditions are subject to continual changes due to routine adjustments in robot end-effector speed, trajectory geometry, and payload weight [11]. Data-driven frameworks frequently struggle under these distribution shifts because they rely solely on statistical methods and unconstrained dependencies [12]. Within a static training condition, features such as continuous energy may exhibit strong linear correlations with manipulator velocity under fixed load conditions. However, because the model indiscriminately maps every available correlation, it risks overfitting to the baseline environmental profile rather than accurately isolating genuine degradation mechanisms. Another significant limitation of the model is its inability to differentiate confounding operational energy shifts [13]. When an unseen context change occurs, such as a substantial increase in operational payload, the mechanical impedance of the physical system shifts dramatically. This increase in mass dampens structural dynamics and introduces significant low-frequency noise, inflating absolute sensor baselines across multimodal channels. In the absence of structural reasoning, associative models often misinterpret an energy baseline shift resulting from environmental mass changes as either a velocity change or a fault condition, ultimately disrupting the established classification boundaries [14].

Additionally, a significant limitation is the substantial overhead associated with calibration and data labeling [15]. Traditional associative classifiers lack domain invariance, leading to a significant decline in diagnostic reliability when faced with out-of-distribution (OOD) covariate shifts [16]. This degradation increases the number of costly false-positive alarms, necessitating frequent manual inspections. Consequently, production lines are often halted, requiring expensive data relabeling and sensor recalibration whenever a manufacturing systems tooling, material thickness, or operational payload is modified.

To address these significant limitations, this study presents the Interventional Causal Framework for Condition-Based Maintenance (IC-CBM), which transitions predictive manufacturing analytics from a reliance on simple correlation to a focus on structural causation. Rather than accepting a high-dimensional data compilation without inspection, the proposed architecture employs constraint-based causal structure discovery, utilizing the Peter-Clark (PC) algorithm [17], supported by rigorous bootstrap resampling methods. This approach systematically identifies the recovered

completed partially directed acyclic graph (CPDAG) that governs the dynamics of the multi-sensor manufacturing cell [18]. This process clearly reveals causal relationships that are consistent with established physical laws, such as the transmission of mechanical vibrations and the propagation of sound waves. This foundation supports the model's predictions and enhances its interpretability and reliability for practitioners. The framework is validated on a FANUC M-1iA/0.5A 6-axis parallel kinematic delta robot instrumented with two vibration transducers and two acoustic sensors. A fully factorial interventional design independently varied end-effector speed (50, 150, and 300 mm/s), trajectory geometry (circular, linear, and compound paths), and payload mass (108.71 g and 179.93 g) across 18 operating conditions, with 20 repeated observations per condition, yielding 360 total observations. In the present study, the classification target is a binary operating condition defined by a commanded-speed threshold, an operational proxy for mechanical stress of a high-speed parallel robot following a given path rather than a directly measured fault or degradation state. The experiments therefore validate IC-CBM as a domain-invariant operating- condition classifier, establishing the causal methodology as a foundation for subsequent validation against measured degradation and fault labels. The main contributions of this research are as follows:

- Topological recovery of stable causal pathways: We demonstrate that constraint-based causal discovery, applied to a fully factorial interventional dataset, successfully recovers a physically interpretable completed partially directed acyclic graph (CPDAG) directly from multimodal sensor data. The recovered graph isolates the kinematic velocity backbone linking commanded speed to contact duration and maps the cross-modal mechanism of airborne acoustic generation, in which frame-mounted structural vibrations radiate through the air to a lateral microphone array, as high-stability causal edges.

- Confounder mitigation via structural residual normalization (SRN): We introduce an architectural intervention to mitigate external context confounding based on structural causal equations. By fitting a linear regression mapping exclusively to the isolated inter-sensor transmission path, the framework mathematically characterizes and subtracts the baseline mass-induced acoustic inflation. This targeted causal intervention enables a causal classifier trained on the causally selected feature subset to maintain a stable mean out-of-distribution accuracy of 88.89% across the unseen payload domain, outperforming the full-feature correlational baseline by 20 percentage points, even as the baseline collapses under the payload covariate shift.

- Information parsimony and high-precision edge efficiency: The framework achieves a feature parsimony index of 0.8461. By reducing the active feature space by 84.61% and identifying a compact, sufficient representation for the robot's operating state, the IC-CBM architecture drastically reduces computational complexity. Crucially, this parsimonious model delivers a high-stress-condition classification precision of 1.00 under severe payload covariate shift, meaning every cycle flagged as high-stress is genuinely a high-stress cycle. A high-stress-condition recall of 0.67 reflects the conservative classification boundary applied in the unseen domain. This compact representation significantly reduces the overhead of continuous sensor calibration and data re-labeling.

The remainder of this manuscript is organized as follows. Section 2 reviews relevant literature in causal AI and robotic diagnostics. Section 3 presents the theoretical foundations of IC-CBM, the causal structure discovery, and the feature extraction methodology. Section 4 describes the experimental design for validating the IC-CBM methodology. Section 5 presents the experimental results, domain-generalization benchmarks, and residual optimization. Section 6 provides a comprehensive discussion of the physical and industrial implications, and Section 7 concludes the paper and outlines future research plans.

2. Related Work

2.1 Multimodal sensing in manufacturing

Manufacturing environments with machine perception have been significantly influenced by vision, which provides rich spatial information but is hindered by occlusion and illumination variations. Additionally, it cannot directly capture mechanical contact properties such as force, friction, and compliance [19]. These limitations have driven the development of multimodal sensing frameworks that combine vision with tactile, force, vibration, and acoustic feedback, allowing complex machineries involved manufacturing to move from purely geometric planning to contact-aware execution [20].

Tactile sensing was the first modality to complement vision meaningfully. Howe and Cutkosky showed that dynamic tactile sensing based on skin acceleration and stress-rate signals could reliably detect transient contact events and incipient slip [21], establishing the basis for contact-sensitive manufacturing systems. Camera-based sensors such as GelSight later recovered contact geometry, local forces, and shear from elastomer deformation [22], and their integration with deep learning has improved object recognition, pose estimation, and in-hand manipulation [23].

Acoustic and vibration sensing capture the dynamic state of machine environment interactions in ways vision and tactile sensing cannot. Physical contact generates structure-borne acoustic emissions and vibrations that propagate through both bodies, encoding material properties and contact state [4]. Active acoustic sensing has been used to recover object resonant properties for recognition and grasp planning [7, 24]. In high-speed manufacturing, feature-level fusion of vibration and acoustic emission yields a more robust representation of cutting dynamics than either signal alone, and deep learning on vibro-physical data has enabled simultaneous texture identification, slip detection, and speed estimation at low latency [25].

2.2 Condition-based maintenance in robotic manufacturing

Condition-based maintenance (CBM) replaces fixed-interval servicing with decisions triggered by measurable indicators of degradation. The canonical framework of Jardine, Lin, and Banjevic organizes this into three stages: data acquisition, signal processing, and maintenance decision-making (Jardine et al., 2006). In high-speed robotic manufacturing, where joints, reducers, and bearings operate under sustained stress, early detection of degradation is economically and operationally critical. Motor current signals have historically been the primary data source, as they require no added hardware. Borgi et al. [27] showed that manipulator accuracy errors correlate with current time-series features. However, current signals are low-frequency and sensitive mainly to gross faults, motivating higher-

frequency vibration and acoustic emission sensing for early degradation. Vibration-based diagnostics is well established. Jaber and Bicker combined discrete wavelet transform decomposition with neural networks to diagnose gearbox backlash faults [28], and one-dimensional deep convolutional networks have since learned discriminative features directly from raw vibration signals, outperforming classical classifiers [6]. On the other hand, acoustic emissions (AE) are especially sensitive to early-stage faults in the rotate vector (RV) reducers common in robot joints. Yang et al. [29] detected RV reducer faults invisible to conventional vibration analysis using compressed-sensing AE, and Zhang et al. [30] encoded fault severity and type in AE burst sequences with a Hidden Markov Model. The two modalities are complementary: vibration reflects the global dynamic response, while AE captures the high-frequency transient energy released during micro-crack propagation or asperity contact, providing earlier warning.

Machine learning is now the standard tool for CBM classification. Support vector machines have been widely adopted for fault diagnosis, as their margin-maximizing formulation generalizes well on the small, high-dimensional datasets typical of machine-health monitoring. At the same time, random forests and deep learning architectures offer comparable robustness to noisy features with lower sensitivity to kernel and hyperparameter choices [31]. However, a review by Sharma, Mittal, and Soni found that the black-box nature of such methods limits the interpretability maintenance engineers require [32], and these approaches expand rather than reduce the feature space, offering no principled mechanism for isolating a sparse, sufficient health representation. Recent work has demonstrated that complex simultaneous fault states in industrial equipment such as control valves can be more effectively diagnosed through multi-feature fusion and multilabel classification frameworks that jointly identify fault type and severity [33], illustrating the growing need for structured, interpretable representations that go beyond single-label associative models.

A classifier trained at a fixed speed and payload may learn patterns confounded by the operating condition rather than the fault state, and will fail when conditions change. Ayankoso et al. [34] illustrated this directly, showing that cobot anomaly-detection autoencoders need explicit handling of trajectory changes to avoid false alarms. The CBM literature has largely treated this through normalization or condition-stratified training rather than as a structural confounding problem. On the prognostics side, data-driven remaining-useful-life models, including LSTM architectures for robot joints, remain trained on passively collected data under narrow operating conditions [35].

2.3 Causal discovery in industrial and robotic systems

Moving from correlation to causal structure is a fundamental challenge in industrial monitoring. Associations learned by conventional models break down under distributional shift, whereas causal discovery aims to recover the underlying data-generating mechanisms. The theoretical basis rests on the constraint-based framework of Spirtes, Glymour, and Scheines [17] and the structural causal models and do-calculus of Pearl [36].

In process monitoring, causal discovery has been applied mainly to root cause analysis, where a single fault triggers cascading anomalies across sensors. Granger causality tests whether past values of one signal enhance the prediction of another. Chen et al. (Chen et al., 2018) formalized this concept for fault-propagation tracing in chemical processes, while transfer entropy extends it to account for nonlinear and asymmetric information flow [38].

Both approaches, however, face challenges related to instantaneous effects and latent confounders. Constraint and score-based algorithms are generally more principled. The PC algorithm is designed to recover Markov equivalence classes of DAGs through conditional independence testing. Clavijo et al. [39] used the PC-stable variant for fault-detection variable selection, though it was used as a selection filter rather than for architectural correction. More recently, scalable causal partitioning methods have been proposed that reformulate causal graph recovery as an edge-cutting problem over adjoint graph representations, enabling more efficient and reliable structure recovery in high-dimensional sensor spaces [40]. When latent confounders are suspected, Fast Causal Inference (FCI) offers greater robustness. On the other hand, LiNGAM takes advantage of non-Gaussian noise to resolve ambiguities related to directions within an equivalence class [41]. Choudhary et al. [42] combined FCI with expert priors to develop interpretable degradation graphs that outperformed black-box detectors in interpretability.

Graph neural networks have emerged as a complementary paradigm, learning structural relationships directly from sensor topology. Causal-attention mechanisms applied to graph representations have been shown to disentangle causally relevant from spurious feature interactions in complex multi-sensor industrial processes, improving fault diagnosis robustness under varying operating conditions [43]. This graph-causal perspective is particularly relevant for cross-domain fault diagnosis, where learned structural dependencies must transfer across differing operating conditions. In robotics, causal reasoning has supported transfer learning, control, and fault diagnosis. The CausalWorld benchmark [44] provides a simulated manipulation environment for causal structure and transfer learning, stressing the value of interventional data from deliberate environmental perturbation, and a counterfactual framework has treated fault modes as interventions on a causal graph to separate sensor-damage from environmental faults [45]. Despite this, causal discovery on high-frequency multi-modal signals in high-speed robotic manufacturing remains largely unexplored as the most work relies on low-frequency process variables. Modirrousta et al. [46] noted that purely statistical methods struggle with high-dimensional, noisy sensor data, especially in the absence of physical priors. To address this issue, they proposed a physics-guided variational attention framework, CAnDOIT [47], which demonstrated that integrating observational and interventional time series enhances the recovery of structures in robotic manipulation.

2.4 Research Gap

The various bodies of literature on multimodal robotic sensing, vibration-acoustic-based CBM, and causal discovery in industrial systems have advanced independently, and their intersection has not been addressed.

Existing diagnostic frameworks, whether based on vibration features, AE analysis, or deep learning classifiers, are typically trained on passively collected data under fixed or narrowly varied conditions. These frameworks implicitly assume a stable relationship between features and health states. This assumption can lead to failures when routine changes in speed, trajectory geometry, or payload mass alter mechanical impedance, shifting the baseline for every sensor channel. For example, if a classifier associates high acoustic energy with a fault, it can generate false alarms when a similar rise in energy occurs due to a heavier payload. Current CBM frameworks do not adequately address this issue, as standard responses such as normalization, condition-stratified retraining, or large labeled datasets do not generalize to previously unseen contexts.

Causal discovery in industrial settings has generally focused on low-frequency process variables. Extending this approach to heterogeneous vibro-acoustic feature spaces is complex because these features span different physical domains with varying noise characteristics and combinatorial relationships among candidates. Most robotics causal research has focused on perception and transfer learning rather than physical maintenance in real-world cells, and no prior studies have applied the PC algorithm to a fused vibro-acoustic feature matrix in a high-speed cell to extract transmission laws that generalize across domains.

A causal graph can be reliably identified only when the hypothesized causes are independently and sufficiently varied. However, CBM data is typically collected under a single or a few predefined conditions, which is insufficient for effective structural recovery. In contrast, robotics interventions often come from simulations or soft time-series data rather than controlled factorial designs that vary speed, trajectory, and payload. Without such a factorial design, the recovered causal graph may appear to reflect relationships that do not actually exist. Moreover, in cases where causal discovery has aided fault diagnosis, the resulting graph has primarily served as an interpretive tool or variable-selection filter. The idea of using the recovered CPDAG topology to derive a structural residual normalization by fitting a transmission model along an identified causal edge and subtracting its contribution to eliminate confounding pathways has not been proposed or assessed in prior work.

This work addresses these gaps simultaneously. Multimodal vibration and acoustic data were collected from a real high-speed robotic cell using a fully factorial interventional design that independently varied speed, trajectory, and payload. The PC algorithm, combined with bootstrap resampling, successfully recovers the invariant physical CPDAG. The SRN derived directly from that graph eliminates mass-induced acoustic baseline inflation without relying on passive statistical normalization. As a result, the causal classifier maintains stable accuracy across unseen contexts in a domain-generalization protocol, even as a full-feature correlational baseline collapses. To our knowledge, this is the first study to integrate interventional experimental design, multimodal vibro-acoustic sensing, constraint-based causal structure discovery, and structural confounder correction into a single, industrially validated CBM framework for high-speed robotic manufacturing. The proposed framework aims to achieve not only high classification accuracy and domain invariance but also a minimally sufficient representation for system diagnostics. To evaluate this objective, we introduce a secondary metric that quantifies the reduction in dimensionality as we transition from an associative processing paradigm to a structurally constrained topology.

3. Methodology

In this section, the proposed Interventional Causal Framework for Condition-Based Maintenance (IC-CBM) in high-speed robotic manufacturing is discussed in detail. Figure 1 presents an overview of the IC-CBM framework within the context of high-speed robotic manufacturing. This framework specifically outlines three novel contributions of this research: a controlled interventional factorial design, the application of the PC causal discovery algorithm to multi-modal sensor features, and a generalization-validated protocol for classifying CBM.

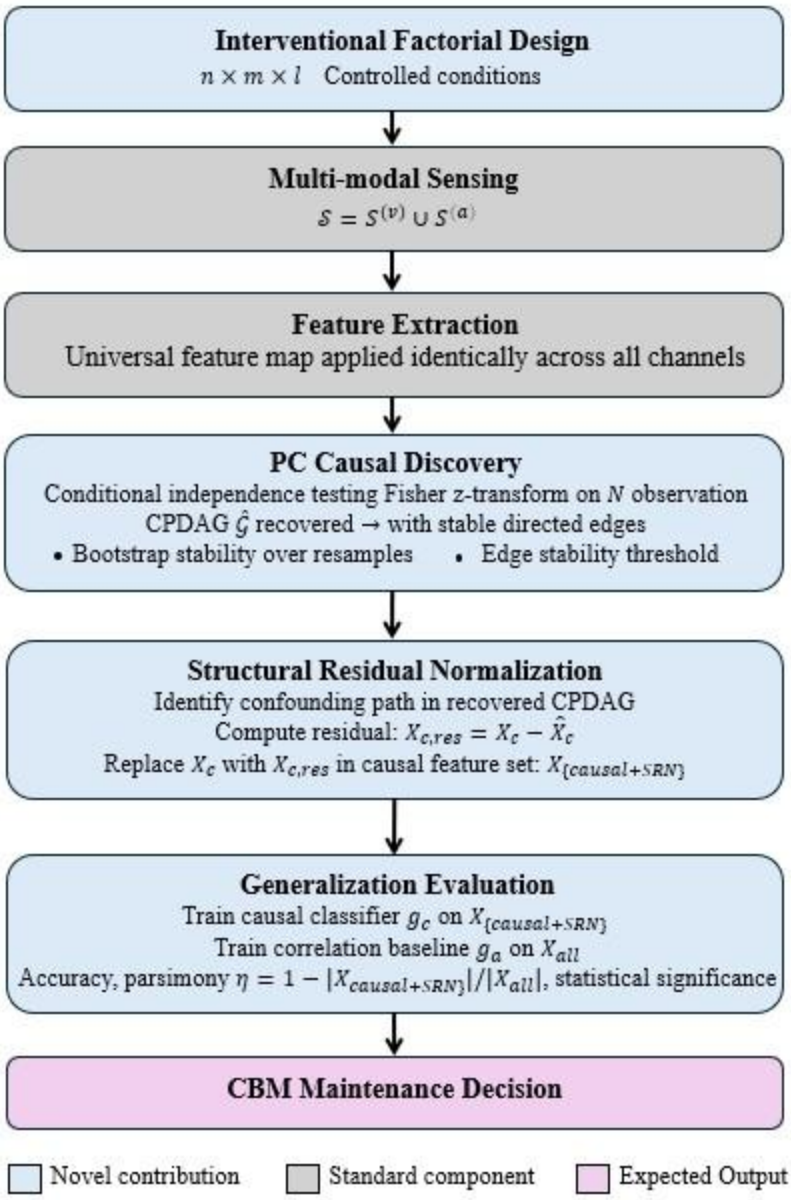


Fig. 1 Overview of the IC-CBM framework within the context of high-speed robotic manufacturing

3.1 Problem formulation

Let the intervention space be defined by three independently controllable factors: speed, trajectory type, and payload. Each factor is described as shown in equations 1-3. The robot end-effector speed ($\mathcal{V}$) is drawn from a finite ordered set of n discrete levels:

$$\mathcal{V} = \{v_1, v_2, \dots, v_n\} \subset \mathbb{R}^+ \tag{1}$$

The trajectory ($\mathcal{T}$) type is drawn from a set of m geometric profiles:

$$\mathcal{T} = \{\tau_1, \tau_2, \dots, \tau_m\} \tag{2}$$

The payload ($\mathcal{P}$)condition is drawn from a set of l discrete load states:

$$\mathcal{P} = \{p_1, p_2, \dots, p_l\} \tag{3}$$

The complete factorial condition space is the Cartesian product [48]:

$$\mathcal{C} = \mathcal{V} \times \mathcal{T} \times \mathcal{P}, |\mathcal{C}| = n \cdot m \cdot l \tag{4}$$

To ensure statistical reliability of causal structure recovery, each condition is executed R times under identical robot configuration settings, producing independent repeated observations (R). The total number of observations across the full experiment is therefore $N = n \cdot m \cdot l \cdot R$. Each experimental condition is uniquely indexed as $c_k \in \mathcal{C}$ for $k = 1,2, \dots, n \cdot m \cdot l$. In this study, two different sensing modalities were employed. For each condition c_k, multi-modal sensor observations are acquired simultaneously from q channels, partitioned into q_v contact vibration channels and q_a acoustic channels such that $q = q_v + q_a$. The raw signal matrix ($\mathbf{X}^{(.)}$) for observation (k, r) is:

$$\mathbf{X}^{(k,r)} \in \mathbb{R}^{q \times T} \tag{5}$$

where T denotes the number of time-domain samples per acquisition run. The sensor channel set is partitioned as:

$$\mathcal{S} = S^{(v)} \cup S^{(a)},\ S^{(v)} \cap S^{(a)} = \emptyset \tag{6}$$

where $S^{(v)} = \{s_1^{(v)}, s_2^{(v)}, \dots, s_{q_v}^{(v)}\}$ and $S^{(a)} = \{s_1^{(a)}, s_2^{(a)}, \dots, s_{q_a}^{(a)}\}$ denote the vibration and acoustic channel sets respectively.

3.2 Feature extraction

Raw time-domain signals are transformed into compact scalar feature vectors that capture physically meaningful characteristics of robot-surface interaction. This step reduces the high-dimensional signal space to a structured feature matrix suitable for causal discovery. Let $x_j^{(k,r)}(t) \in \mathbb{R}^T$denote the time-domain signal acquired from channel j under condition (k, r).

Four features are extracted from each channel, with feature maps applied uniformly across all channels, regardless of modality. The interpretation of each feature, contact energy, root mean square (RMS) intensity, peak-to-peak range, crest factor, and contact duration, remains consistent across both vibration and acoustic domains. In particular, the RMS sustained energy level signifies the intensity of the contact event, the peak-to-peak range captures the extremal dynamic response, the crest factor reflects the impulsiveness of the contact event, and the contact duration quantifies the time exceeding the noise threshold [49, 50]

The feature maps $\varphi_{(.)}^{(v)}$ and $\varphi_{(.)}^{(a)}$can be defined according to the formulas presented in equations 7 through 10.

$$\varphi_1^{(.)}(x) = \sqrt{\frac{1}{T}\sum_{t=1}^{T} x(t)^2} \tag{7}$$

$$\varphi_2^{(.)}(x) = \max_t(x) - \min_t(x) \tag{8}$$

$$\varphi_3^{(\cdot)}(x) = \frac{\max_t |x(t)|}{\varphi_1^{(\cdot)}(x)} \tag{9}$$

$$\varphi_4^{(\cdot)}(x) = \frac{1}{f_s}\sum_{t=1}^{T} \mathbf{1}[|x(t)| > \theta_j] \tag{10}$$

where θ_j is a channel-specific activity threshold set at the three standard deviations above the baseline noise floor (Jardine et al., 2006).

The full feature vector for condition concatenates features across all channels:

$$\boldsymbol{f}^{(k,r)} = [\boldsymbol{f}_1^{(k,r)\intercal}, \boldsymbol{f}_2^{(k,r)\intercal}, \dots, \boldsymbol{f}_q^{(k,r)\intercal}]^\intercal \in \mathbb{R}^d,\ d = q \cdot r_f \tag{11}$$

where r_f denotes the number of features per channel. The global feature matrix aggregating all N observations is:

$$\boldsymbol{F} \in \mathbb{R}^{N \times d} \tag{12}$$

with rows indexed by observation (k, r) and columns indexed by feature. Prior to causal structure discovery, all columns of $\boldsymbol{F}$ are standardized to zero mean and unit variance to ensure valid conditional independence testing across features derived from heterogeneous sensor channels.

3.3 Causal graph formulation

After extracting the feature matrix **F**, the next objective is to determine which intervention variables, specifically speed, trajectory, or payload, cause changes in sensor features, rather than simply correlating with them. This distinction is important for CBM. A feature that is causally influenced by speed will respond predictably to changes in speed under unseen deployment conditions. In contrast, a feature that is only correlated may not generalize when the overall operating conditions change. Let the augmented variable set be defined as in equation 13.

$$\mathcal{U} = \mathcal{I} \cup \mathcal{F} \tag{13}$$

where $\mathcal{I} = \{v, \tau, p\}$ denotes the intervention variables and $\mathcal{F}$ denotes the extracted features. The causal structure over $\mathcal{U}$ is represented as a directed acyclic graph (DAG) as in equation 14.

$$\mathcal{G} = (\mathcal{U}, \mathcal{E}) \tag{14}$$

where a directed edge $(u \longrightarrow w) \in \mathcal{E}$ encodes that variable u is a direct cause of w, meaning u has a measurable effect on w that persists even after accounting for all other variables that directly influence w and $\boldsymbol{Pa}\ (w)$ denotes the parent set of w in $\mathcal{G}$ [36].

Two standard assumptions from causal inference theory make the graph recoverable from data. The Causal Markov Condition states that once the direct causes of a variable are known that variable becomes independent of all other variables it does not itself cause. In other words, knowing the parents of a node is sufficient to explain all its statistical dependencies and no additional variables carry further information about it. It can be represented as in equation 15.

$$P(\mathcal{U}) = \prod_{u_i \in \mathcal{U}} P(u_i \mid \boldsymbol{Pa}(u_i)) \tag{15}$$

This assumption ensures that each variable is independent of its non-descendants, provided that its direct causes are known. The faithfulness assumption requires that no statistical independence in the data arises from coincidental algebraic cancellation between competing causal pathways. In the present experimental setting, this risk is mitigated by the fully factorial interventional design: speed, trajectory, and payload are independently varied across orthogonal levels, substantially reducing the likelihood of systematic pathway cancellation that would otherwise produce spurious independence. A more practical concern is a near-faithfulness violation due to weak causal effects, where a genuine edge exists. Still, it yields a partial correlation too small for the independence test to distinguish it from zero at the available sample size. This is directly addressed by the bootstrap stability protocol described in Section 3.4: edges that fail to reappear across repeated condition-level consistently resampled networks are excluded from the reported causal structure, regardless of whether they represent true near-faithfulness violations or insufficiently powered effects. The reported stable edges therefore represent relationships the PC algorithm reliably recovers across diverse resamples, providing empirical robustness against both faithfulness violations and sampling noise. This means that the observed patterns of independence in the data accurately reflect the underlying causal structure rather than being merely coincidental [18, 36].

Together, these two assumptions guarantee that the true causal graph can be uniquely identified from the data, up to its Markov equivalence class. This class consists of a set of DAGs that share the same conditional independence structure and cannot be distinguished from one another solely based on data. In the present study, the intervention variables v, τ and p are treated as root nodes in $\mathcal{G}$ with no incoming edges, directly reflecting their status as experimentally controlled causes. All causal arrows of interest therefore originate from $\mathcal{I}$ and point into the feature set $\mathcal{F}$, or run among features within $\mathcal{F}$.

3.4 PC algorithm for causal structure recovery

The PC algorithm recovers $\hat{\mathcal{G}}$ from the feature matrix $\boldsymbol{F}$ through two sequential phases: skeleton discovery and edge orientation. The algorithm makes no assumptions about the functional form of causal relationships beyond the Markov and faithfulness conditions stated in Section 3.3, making it a non-parametric and broadly applicable method for causal structure recovery.

Phase 1 - Skeleton discovery. The algorithm begins with a fully connected undirected graph over all variables in $\mathcal{U}$ and progressively removes edges by testing for conditional independence between variable pairs. For each pair of variables (u_i, u_j), the algorithm tests whether the two variables are independent once the linear influence of some conditioning subset $W \subseteq \mathcal{U} \setminus \{u_i, u_j\}$ of other variables is removed.

This is quantified by the partial correlation $\hat{\rho}_{ij \cdot w}$, which is the residual correlation between u_i and u_j after removing the effect of all variables in W. This is transformed into a test statistic using the Fisher z-transform as shown in equation 16 [18, 52]. The Fisher z-transform operates under the assumption that the variables in U are jointly Gaussian and that conditional dependencies among them are approximately linear. In the context of the heterogeneous vibro-acoustic feature space used here, this assumption is well-justified for two reasons. First, the four features extracted per channel, RMS intensity, peak-to-peak range, crest factor, and contact duration, are aggregate statistical quantities

computed over long acquisition windows. By the central limit theorem, aggregate statistics of this type converge toward approximately Gaussian marginal distributions regardless of the distributional properties of the underlying raw signal, a well-established result in statistical signal processing. Second, the linearity assumption for conditional dependencies is consistent with the physical mechanisms governing this experimental system. The causal relationships of interest, speed compressing contact duration across a fixed spatial path, payload amplifying structural vibration amplitude through inertial loading, and surface vibration transients radiating acoustically to a lateral microphone are first-order mechanical relationships that are approximately linear over the moderate operating ranges tested. The bootstrap stability protocol partially guards against potential violations of the linearity assumption, since edges that are artifacts of misspecified independence tests are unlikely to reappear consistently across many resamples. A kernel-based conditional independence test could be applied to validate the absence of nonlinear dependencies not captured by Fisher-Z and is identified as a direction for future work in Section 7.

$$z_{ij|w} = \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_{ij\cdot w}}{1-\hat{\rho}_{ij\cdot w}}\right)\sqrt{N - |W| - 3} \tag{16}$$

Under the null hypothesis of conditional independence, $z_{ij|w}$ follows an asymptotic standard normal distribution. In this context, $N - |W| - 3$ represents the effective degrees of freedom. The base correction of 3 accounts for the means of u_i, u_j, and the correlation estimate. Each variable in W consumes one additional degree of freedom. The edge between u_i and u_j is removed from the graph when:

$$\left|z_{ij|w}\right| < \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),\ \alpha = 0.05 \tag{17}$$

indicating that u_i and u_j are conditionally independent given W. When this condition is met, W is stored as the separating set $Sep(u_i, u_j)$ for use in the orientation phase. Testing is repeated across all variable pairs and increasing conditioning set sizes $|\ W\ | = 0,1,2,\ldots$ until no further edges can be removed. The result is an undirected skeleton $\hat{\mathcal{G}}$ encoding all genuine statistical dependencies in the data (Kalisch & Bühlmann, 2007).

Phase 2 - Edge orientation. The undirected skeleton is converted into a partially directed graph through two steps. First, the algorithm identifies v-structures, configurations of the form $u_i \longrightarrow u_k \longleftarrow u_j$ where u_i and u_j are not directly connected to each other. A v-structure is confirmed at u_k when $u_k \notin Sep\ (u_i, u_j)$, which indicates that u_k is a common effect of u_i and u_j rather than a mediating variable or common cause. The underlying intuition is clear: if conditioning on u_k induces a dependency between u_i and u_j rather than eliminating it, this indicates that u_k is located downstream of both u_i and u_j, receiving independent influences from each. Furthermore, Meek's propagation rules are applied iteratively to orient any remaining undirected edges in a manner that is consistent with all identified v-structures, while also ensuring that no directed cycles are introduced into the graph [54].

The final output is a CPDAG $\hat{\mathcal{G}}$ representing the complete Markov equivalence class of the true DAG [55]. Edges that remain undirected following Phase 2 indicate genuine statistical ambiguity, as they correspond to two or more causal structures that generate identical conditional independence patterns, thus rendering them indistinguishable based solely on data. Such edges are presented transparently alongside all directed edges. The set of causally identified

features is then delineated as the collection of sensor features that receive at least one directed edge from an intervention variable within $\hat{\mathcal{G}}$. The set of features F_c can be represented as in equation 18.

$$F_c = \{f_i \in F \mid \exists\, u \in \mathcal{I} : (u \longrightarrow f_i) \in \hat{\mathcal{G}}\} \tag{18}$$

The features whose variation can be directly linked to the controlled experimental factors, namely speed, trajectory, or payload, are delineated for consideration. Features not present in $\mathcal{F}_c$ are excluded from subsequent CBM classification due to the potential they possess for reflecting confounding influences or measurement errors rather than true manufacturing dynamics. Consequently, these excluded features cannot be relied upon to generalize effectively to unseen operating conditions.

The PC algorithm was selected over alternative constraint-based and functional causal model approaches for three reasons specific to this experimental context. As discussed in Section 2, FCI is designed for settings in which latent confounders are suspected, variables that influence multiple observed variables but are not themselves measured. In the present study, all three intervention variables are directly controlled and independently observed, making latent confounder assumptions unnecessary, and FCI's more conservative Partial Ancestral Graph output is less informative for the direct intervention-to-feature causal paths of interest. LiNGAM exploits non-Gaussian noise to resolve directional ambiguities within a Markov equivalence class that PC leaves undetermined. However, the features used in this study are aggregate statistical quantities computed over long acquisition windows, and by the central limit theorem such aggregates converge toward approximately Gaussian distributions regardless of the underlying raw signal distribution, making the LiNGAM non-Gaussianity assumption poorly matched to this feature type and increasing the risk of incorrectly resolved edge orientations. The PC algorithm with Fisher-Z conditional independence testing is therefore the principled choice for this setting: it requires no non-Gaussianity assumption, operates correctly in the approximately Gaussian-linear condition produced by aggregate statistical features, and recovers the Markov equivalence class of the causal graph sufficient for identifying the stable causal pathways used in SRN without requiring full directional resolution beyond what the data support. A systematic empirical comparison of PC with FCI and LiNGAM on matched industrial vibro-acoustic datasets is planned for future work, as described in Section 7.

3.5 Structural Residual Normalization

Although causal feature selection isolate features that are causally downstream of the controlled intervention variables, a confounding pathway within the recovered CPDAG may still inflate sensor baselines when an environmental context variable shifts to an unseen level. The IC-CBM framework addresses this structurally by deriving a targeted residual correction directly from the topology of the recovered graph, rather than applying passive global normalization across all features.

When the recovered CPDAG identifies a directed path of the form $E \rightarrow X_p \rightarrow X_c$, where E is an environmental context variable, X_p is a sensor feature causally upstream of X_c, and X_c is the target acoustic or dynamic response feature, the transmission relationship between X_p and X_c encodes a physically invariant structural mechanism. Structural Residual Normalization (SRN) fits a linear regression mapping along this identified causal path exclusively on the in-distribution training domain:

$$\hat{X}_c = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_p \quad (19)$$

Where $\hat{\beta}_0$ and $\hat{\beta}_1$ are ordinary least squares, estimates computed without access to the held-out test domain. The structural residual is then defined as:

$$X_{c,res} = X_c - \hat{X}_c \quad (20)$$

This residual isolates the component of X_c that is independent of the confounding transmission from X_p retaining only the variation attributable to the primary intervention variable of interest. The SRN-corrected causal feature set replaces X_c with $X_{c,res}$ as described in equation 21.

$$\mathcal{X}_{causal+SRN} = (\mathcal{X}_{causal}\{X_c\}) \cup \{X_{c,res}\} \quad (21)$$

The regression parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ are fixed after estimation on the training domain and applied without modification to any evaluation domain, ensuring that the correction introduces no information leakage from held-out conditions. The SRN is applicable to any directed transmission path identified in the recovered CPDAG and extends naturally to multiple confounding pathways by applying the residualization sequentially along each identified path. The specific causal path targeted by SRN in the present study and the resulting corrected feature set are described in Section 5.3.

3.6 Generalization-Validated CBM Classification

The practical significance of causal feature selection is established through a protocol that assesses generalization under held-out conditions. The fundamental hypothesis is clear. If the features identified as causal accurately represent the physical structure of robot-surface interactions, they should be able to identify operational conditions not encountered during the classifier's training phase. In contrast, correlational features may leverage statistical patterns unique to the training distribution, thereby failing to generalize effectively to new conditions.

The full condition set is partitioned into training and test subsets: $\mathcal{C} = \mathcal{C}_{train} \cup \mathcal{C}_{test}, \mathcal{C}_{train} \cap \mathcal{C}_{test} = \emptyset, |\mathcal{C}_{test}| = h$. Two support vector machine (SVM) classifiers [56] are trained under identical hyperparameter configurations. Let $\hat{F}_c$ denote the submatrix of $\boldsymbol{F}$ restricted to columns in F_c. The performance is compared with a causal classifier trained on F_c only, $g_c: \widehat{\boldsymbol{F}}_{c,train} \rightarrow \hat{y}_c$ and baseline classifier trained on all features F, $g_a: \widehat{\boldsymbol{F}}_{train} \rightarrow \hat{y}_a$. Both classifiers are evaluated on F_{test} using the weighted F1-score, which accounts for any imbalance in sample counts across held-out conditions. The weighted F1-score can be represented as in equation 22.

$$F1_w = \sum_{c \in C_{test}} \frac{|c|}{h \cdot R} \cdot \frac{2.Pr_c \cdot Re_c}{Pr_c + Re_c} \quad (22)$$

where Pr_c and Re_c denote per-condition precision and recall [57]. The central empirical hypothesis is: $H_1: F1_w(g_c) \geq F1_w(g_a)$ on C_{test}. Confirmation of H_1 provides evidence that interventionally grounded causal features generalize more reliably to unseen manufacturing conditions than correlation-based full-feature selection. A secondary parsimony metric quantifies the dimensionality reduction achieved: $\eta = 1 - \frac{|F_c|}{d} \in [0,1)$. A high value of η alongside non-degraded F1-score constitutes a parsimony result: the causal graph recovers a compact sufficient representation for

condition-based maintenance, reducing sensor feature complexity without sacrificing classification reliability on unseen operating conditions.

4. Experimental Design and Data Collection

This section presents a thorough overview of the experimental design specifically developed for the IC-CBM methodology. The design outlines the systematic approach employed to explore the research questions, elaborating on the precise procedures and configurations utilized throughout the study.

4.1 Robot Platform and Workspace

All experiments were conducted utilizing a FANUC M-1iA/0.5A, a six-axis high-speed parallel kinematic robot with a maximum payload capacity of 0.5 kg, a horizontal reach of 280 mm, and a vertical stroke of 100 mm. To control the robot, the R-30iB controller was employed, known for its advanced motion control capabilities and user-friendly interface. The parallel kinematic (delta) configuration of the M-1iA/0.5A provides low end-effector inertia and high acceleration capabilities, characteristics that are typical of industrial high-speed manufacturing systems, including applications in pick-and-place operations, food handling, and pharmaceutical assembly. This functionality establishes it as an appropriate testbed for the IC-CBM framework proposed in this study. All trajectories were programmed and executed within a defined horizontal plane at a constant elevation above a rigid contact surface, ensuring that the end-effector remained within the robot's nominal workspace envelope throughout the duration of each experimental run. A visual representation of the experimental design is provided in Figure 2, which illustrates the key components and their interrelations, thus enhancing the understanding of the overall framework of the research.

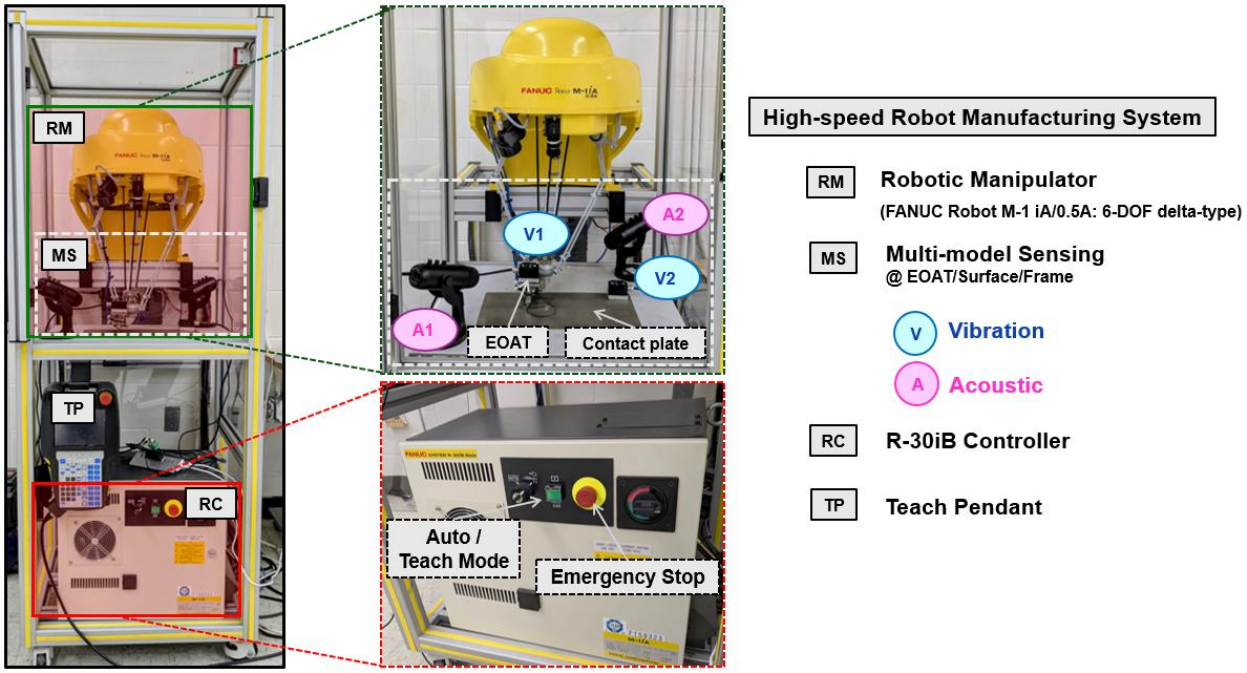


Fig. 2 Experimental setup of causal discovery. Vibration sensors, V1 and V2, and acoustic sensors A1 and A2 are placed strategically to understand the causal structures.

Multi-modal signals were acquired from four channels utilizing two distinct sensing modalities. In the realm of vibration sensing, two three-axis MEMS vibration transducers $s_1^{(v)}$and $s_2^{(v)}$ were employed to measure displacement (ranging from 0 to 2000 μm), speed (ranging from 0 to 50 mm/s), and amplitude (up to 180°). The transducer $s_1^{(v)}$ was rigidly mounted at the end-of-arm tooling (EOAT) to capture vibrations at the source of contact, while $s_2^{(v)}$ was positioned directly on the contact surface within the end-effector workspace to monitor the structural response of the contact environment. In terms of acoustic sensing, two condenser microphones $s_1^{(a)}$ and $s_2^{(a)}$ featuring a frequency response of 80 Hz to 48 kHz and a sensitivity of -43 ± 3 dB were strategically placed 150 mm from the end-effector, oriented laterally in opposing directions. This configuration provided comprehensive stereo acoustic coverage of the near-field emissions generated during the execution of trajectory operations.

4.2 Experimental Factors and Operating Conditions

The experimental design instantiates the $n \times m \times l$ factorial structure defined in Section 3.1 with the following three intervention factors. End-effector speed was varied across three discrete levels: $v_1 = 50\,mm/s$ (low), $v_2 = 150\,mm/s$ (medium), $v_3 = 300\,mm/s$ (high). These levels span a six-fold velocity range and were selected to produce meaningfully distinct vibration and acoustic response conditions while remaining within the M-1iA/0.5A continuous operational envelope. The maximum tested speed of $300\,mm/s$ represents a practically relevant high-speed manufacturing operating point for this platform.

The trajectory type varied across three distinct geometric profiles, each executed as a closed loop, as illustrated in Figure 3.

1. Circular path (τ_1): A continuously curvilinear closed loop of diameter 6.5 cm, producing uniform centripetal acceleration throughout the path with no discrete direction-change events.

2. Linear path (τ_2): An irregular closed quadrilateral with side lengths measuring 7 cm, 4 cm, 4.5 cm, and 6.5 cm presents a unique geometric configuration. This structure is characterized by four distinct events of deceleration, reversal, and acceleration occurring at each vertex per cycle. These events create impulsive contact dynamics with each change of direction.

3. Compound path (τ_3): A closed loop combining a circular arc with two straight segments of 5 cm and 6.5 cm converging at a single vertex, with an internal height of 6 cm across the arc. This geometry produces exactly one discrete direction-change event per cycle, at the vertex where the two straight segments meet.

The three trajectory types cover a continuous range in the number of discrete direction-change events per cycle: zero for τ_1, one for τ_3, and four for τ_2. This creates a systematic gradient that allows for the assessment of the causal influence of trajectory geometry on contact-induced vibrations and acoustic transients.

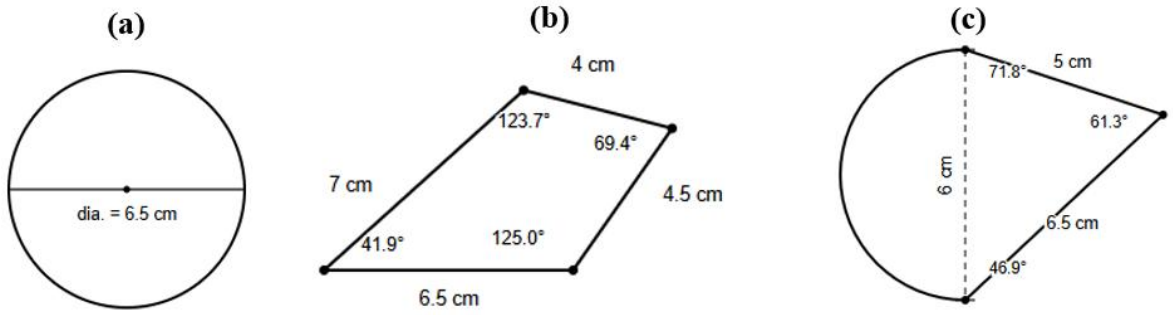


Fig. 3 Trajectory geometries used in the factorial design. (a) Circular path, (b) Linear path, (c) Compound path

The circular path (τ_1), with a diameter of 6.5 cm, produces continuous curvature with no sharp direction changes. The linear path (τ_2), an irregular quadrilateral with side lengths 7 cm, 4 cm, 4.5 cm, and 6.5 cm, producing four discrete direction-change events per cycle. The compound path (τ_3), combining a circular arc (6 cm height) with two straight segments of 5 cm and 6.5 cm meeting at a single vertex, produces exactly one sharp direction change per cycle.

The payload was varied between two discrete states: $p_1 = 108.71$ g (lightly loaded) and $p_2 = 179.93$ g (fully loaded). Both payload masses were rigidly attached to the end-effector tool flange, thereby remaining within the M-1iA/0.5A's rated payload capacity of 500 g. This two-level factor is intended to capture the influence of inertial end-effector loading on the dynamic response of the robotic system and the resultant multimodal sensor signatures.

The complete factorial combination of the three selected factors results in a total of $|\mathcal{C}| = 3 \times 3 \times 2 = 18$ distinct operating conditions, as detailed in Table 1. Each condition, designated as c_k, was executed $R = 20$ times under consistent robot configuration settings, thereby producing a cumulative total of $N = |C| \times R = 18 \times 20 = 360$ independent observations. The implementation of a twenty-repeat protocol generates a sample-to-variable ratio of approximately 19:1 in relation to the 19-dimensional augmented variable set U. This ratio significantly exceeds the minimum threshold recommended for reliable conditional independence testing as outlined in the PC algorithm. The overall data collection duration, about all conditions and repeats, amounted to approximately 19 hours, underscoring the extensive nature of the interventional dataset. The experimented operating conditions are shown in Table 1.

Table 1: Factorial operating conditions

Condition	Speed ($v_{(.)}$)	Trajectory ($\tau_{(.)}$)	Payload ($p_{(.)}$)
c_1	50 mm/s	τ_1	108.71 g
c_2	50 mm/s	τ_1	179.93 g
c_3	50 mm/s	τ_2	108.71 g
c_4	50 mm/s	τ_2	179.93 g
c_5	50 mm/s	τ_3	108.71 g
c_6	50 mm/s	τ_3	179.93 g
c_7	150 mm/s	τ_1	108.71 g
c_8	150 mm/s	τ_1	179.93 g
c_9	150 mm/s	τ_2	108.71 g
c_{10}	150 mm/s	τ_2	179.93 g
c_{11}	150 mm/s	τ_3	108.71 g
c_{12}	150 mm/s	τ_3	179.93 g
c_{13}	300 mm/s	τ_1	108.71 g
c_{14}	300 mm/s	τ_1	179.93 g
c_{15}	300 mm/s	τ_2	108.71 g
c_{16}	300 mm/s	τ_2	179.93 g
c_{17}	300 mm/s	τ_3	108.71 g
c_{18}	300 mm/s	τ_3	179.93 g

4.3 Data acquisition and feature extraction

All four sensor channels were connected to a host PC via a powered USB hub and by a custom Python data logging script. At the start of each trajectory execution, a common software timestamp was issued to initiate simultaneous acquisition across all channels. Each acquisition run lasted $T_r = 190$ seconds per observation. Raw outputs from all four channels were stored as indexed files per observation labeled by condition index and repeat r.

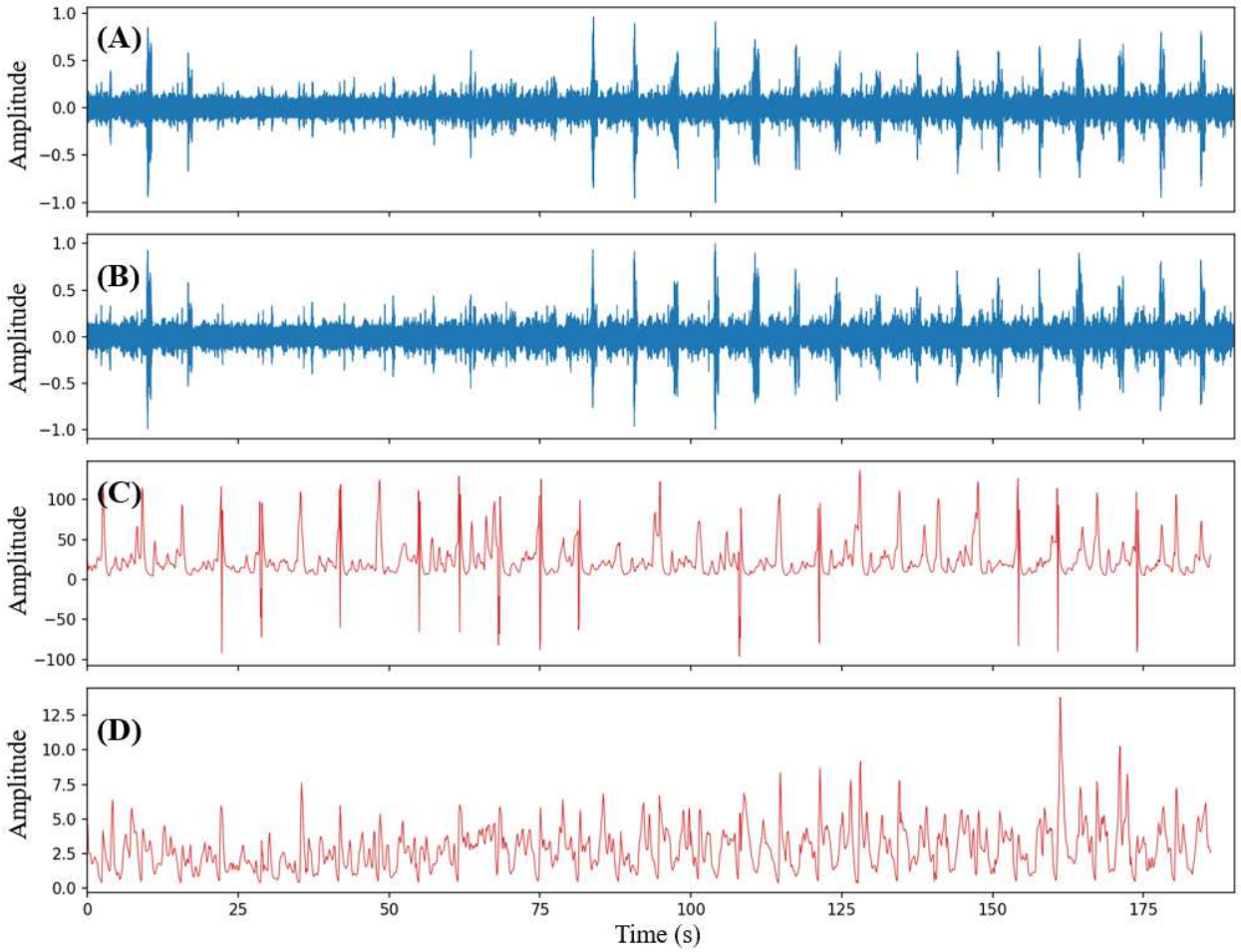


Fig. 4 Time-domain sensor outputs over 190 seconds from condition c_3, 50 mm/s, linear path trajectory τ_2, 108.71 g payload. (A) Acoustic sensor 1, (B) Acoustic sensor 2, (C) Vibration sensor 1, (D) Vibration sensor 2

The plots labeled (A) and (B) show the acoustic pressure waveforms captured by microphones 1 and 2, respectively. Plots (C) and (D) present the vibration speed sequences recorded from $s_1^{(v)}$, which relates to the end-of-arm tooling, and $s_2^{(v)}$, associated with the contact surface. The clustered, multi-peak burst pattern observed across all four channels corresponds to the four distinct deceleration-reversal-acceleration events that occur at each corner of the linear path trajectory during each cycle. This pattern is notably different from the smoother, single-peak profile typically anticipated for a circular trajectory. The measurements from $s_1^{(v)}$ exhibit significantly larger peak amplitudes than those from $s_2^{(v)}$, which aligns with the expected vibration attenuation between the contact source at the end-effector and the surface-mounted sensor.

The feature extraction pipeline defined in Section 3.2 was applied independently to each channel across all 360 observations. For the vibration channels ($j \in S^{(v)}$), the vibration speed sequence served as the primary output sequence $s_j^{(k,r)}$. To maintain strict feature commensurability across all channels in the subsequent causal discovery stage, the same four time-domain features, RMS intensity, peak-to-peak range, crest factor, and contact duration, were

applied uniformly to both modalities. The channel-specific activity threshold $\theta_j = \mu_j + 3\sigma_j$ was estimated from a static baseline recording captured immediately prior to each run under static robot conditions. The resulting global feature matrix $\boldsymbol{F} \in \mathbb{R}^{360\times16}$ was column-wise standardized to zero mean and unit variance prior to causal structure discovery, ensuring valid conditional independence testing across features derived from channels with different acquisition rates, physical units, and dynamic ranges.

4.4 Implementation Details

For full reproducibility, the computational configuration of each pipeline stage is documented here. All analyses were implemented in Python 3.10. Causal structure discovery used the causal-learn library (version 0.1.3.8) with the PC algorithm configured as follows: significance threshold α = 0.05, independence test = Fisher-Z, stable ordering enabled to remove skeleton construction order-dependence, majority rule for v-structure orientation, and priority given to existing v-structures in conflict resolution. Background knowledge was injected to enforce the known exogeneity of the three intervention variables: directed edges from any sensor feature into speed, trajectory, or payload were explicitly forbidden, ensuring these variables remained root nodes in the recovered CPDAG consistent with the factorial interventional design. Bootstrap edge stability was evaluated over B = 1,000 condition-level resampling iterations, with edges excluded from the final reported causal structure if their recovery frequency fell below a stability threshold of 0.5.

The SRN linear regression mapping was fitted using a standard ordinary least squares estimator trained exclusively on in-distribution payload domain observations (108.71 g) and applied to both domains without parameter updates. The SVM classifier was configured with a radial basis function kernel, regularization penalty C = 0.5, and gamma = scale. Hyperparameters were fixed a priori through stratified 5-fold cross-validation on the training domain and were not adjusted in response to out-of-distribution evaluation outcomes. The 10-trial bootstrapped domain generalization protocol resampled both training and test domains. The random-subset fairness validation used 30 independent random feature draws per cardinality level to ensure independence from the main bootstrap evaluation. All reported t-tests were two-sided paired tests across the ten bootstrap trials.

5. Experimental results and generalization analysis

5.1 Empirical Causal Graph Recovery and Topological Interpretation

The standardized multi-modal feature matrix derived from the $3 \times 3 \times 2$ factorial experimental design was subjected to causal structure discovery utilizing the PC algorithm. To evaluate the statistical robustness and structural reliability of the discovered dependencies, a bootstrap resampling protocol ($B = 1000$ iterations) was executed. This step generated an empirical edge stability score for each recovered relationship. Table 2 reports the top-tier topologically stable causal edges that achieved non-zero recovery confidence.

Table 2: Topologically stable causal edges recovered via PC algorithm

From Node	To Node	Empirical Stability Score
Speed	CD	1.000
Payload	V2-P2P	1.00
V2-P2P	V2-RMS	0.995
A1-P2P	A1-RMS	0.980
A1-CF	A1-RMS	0.980
A2-CF	V2-RMS	0.960
V1-CF	V1-P2P	0.920
V2-CF	A1-P2P	0.890
A2-P2P	A1-P2P	0.890
A1-P2P	CD	0.870
V2-CF	V2-RMS	0.825
V1-RMS	V1-P2P	0.815
Trajectory	A1-RMS	0.605

The sensor types are classified as vibration (V) and acoustic (A), with designated sensor locations as either 1 or 2. The features include root mean squared (RMS), peak-to-peak (P2P), crest factor (CF), and contact duration (CD). The exact structural connectivity of the recovered CPDAG is illustrated in Figure 5. The topology systematically maps the structural flow from controlled physical interventions to downstream sensory nodes and cross-modal bottlenecks.

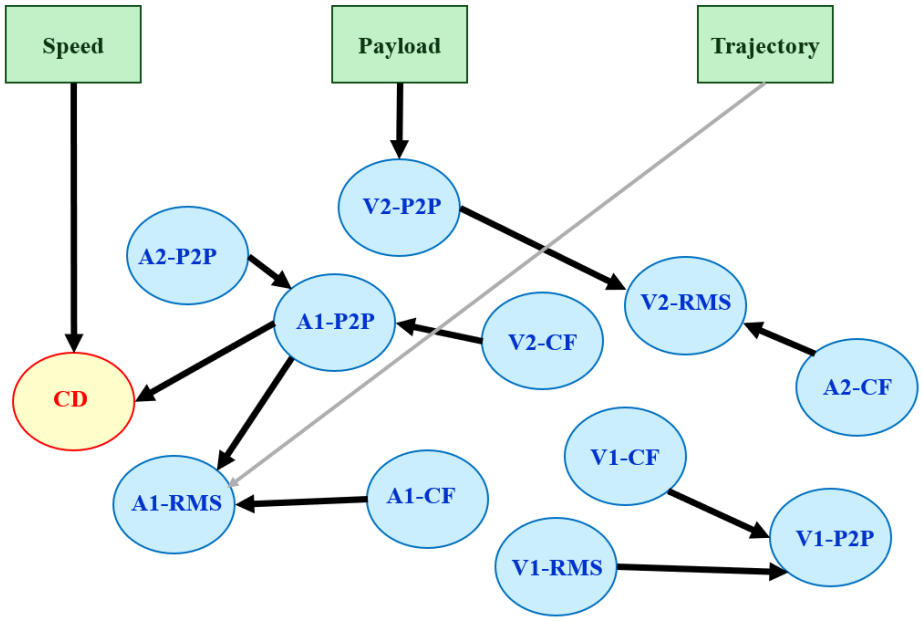


Fig. 5 Recovered CPDAG topology from the experimental intervention dataset. Rectangular nodes signify exogenous control parameters, while circular nodes denote multi-modal sensor features. Edge thickness is strictly proportional to the empirical bootstrap stability score, mapping empirically stable causal pathways.

The network architecture, illustrated in Figure 5, distinctly separates physical pathways instead of presenting a complex correlational web. First, the algorithm accurately identifies deterministic control constraints with complete

confidence, isolating the directed edge between Speed → CD (Stability: 1.000). This aligns precisely with physical kinematics: increasing the joint velocity of the robotic manipulator across a fixed spatial path mathematically compresses the duration window of acoustic and mechanical engagement.

Crucially, the physical location of the sensors explains the specific topological structures discovered by the algorithm. The PC algorithm recovers a definitive parent edge Payload → V2-P2P (stability: 1.000). V2 is mounted directly on the structural surface, allowing it to capture the overall structural deflection and the inertial forces caused by shifting payload weights. In contrast, V1 is attached to the EOAT and measures highly localized cutting and contact dynamics. These dynamics tend to remain consistent, regardless of changes in static payloads, which reduces their direct dependency on the payload node.

Furthermore, a distinct cross-modal transmission pathway is recovered via the directed edge V2-CF → A1-P2P (Stability: 0.890). This edge maps the structural mechanism of airborne acoustic generation within the manufacturing cell. High-frequency transients in the robot's main surface-mounted frame V2 radiate into the surrounding air.

A1 and A2 are positioned 150 mm away from the end effector and are oriented in opposite lateral directions, providing clear spatial diversity. A1's lateral orientation allows it to capture the direct acoustic wavefront emitted by the surface structural vibrations (denoted as V2-CF). As a result, A1 serves as a directional causal bottleneck, represented by the link V2-CF → A1-P2P. In contrast, A2 faces the opposite direction and is positioned to capture ambient reflections, resulting in the directional node link A2-P2P → A1-P2P (Stability: 0.890).

The topology maps the integrated energy features (RMS) as downstream components of localized transient features (P2P and CF). For example, the relationships A1-P2P → A1-RMS and A1-CF → A1-RMS both exhibit a stability of 0.980. This clear directional relationship demonstrates that discrete, high-amplitude mechanical transient events are mathematically responsible for the continuous acoustic energy dissipated over an operational cycle. This suggests that the PC algorithm effectively reflects the recovered CPDAG structure rather than merely statistical dependencies.

5.2 Operating-Condition Classification and Out-of-Distribution Generalization

To rigorously evaluate the hypothesis that causally identified features provide robust representations under changes in environmental context, both classifiers were subjected to a ten-trial bootstrapped domain generalization protocol. Each trial independently resampled the training and held-out domains, and the results were aggregated to produce estimates of mean accuracy and standard deviation. A paired t-test across the ten trials assessed whether the observed generalization advantage was statistically significant rather than an artifact of a single favorable data partition.

A binary operational condition label was defined based on the commanded velocity profile. A high-speed operating condition (Speed ≥ 300 mm/s) is designated as a high-stress condition, associated with accelerated mechanical wear and elevated maintenance priority. In contrast, lower operational speeds correspond to the standard operating reference state. Identifying this condition boundary under unseen payload conditions is a fundamental precondition for effective CBM. A diagnostic system that cannot distinguish condition-induced energy shifts from genuine degradation signals

will generate false alarms whenever the operational context changes, making it unreliable for real manufacturing deployment.

The global feature matrix was partitioned along the environmental context variable, Payload. Both models were trained and hyperparameter-tuned via stratified 5-fold cross-validation exclusively on data from the baseline payload domain (108.71 g, 180 observations), then locked with zero parameter updates and forced to generalize to the unseen high-stress out-of-distribution test domain (179.93 g, 180 observations). All classification tasks used a Support Vector Machine with a radial basis function kernel, regularization penalty C = 0.5, and gamma = scale. Two feature space configurations were benchmarked. The correlational baseline (X_{all}) incorporates the full thirteen-dimensional multi-modal statistical feature space including CD. The proposed causal framework (X_{causal}) uses only the three direct causal descendants identified by the PC algorithm: kinematic CD, payload-sensitive surface vibration peak (V2-P2P), and cross-modal acoustic peak (A1-P2P). Table 3 reports the empirical generalization performance matrix averaged across all 10 bootstrap trials.

Table 3: Empirical generalization performance matrix (10-trial bootstrap mean ± std)

Evaluation Domain	Causal Framework (X_{causal})	Correlational Baseline (X_{all})
In-Distribution (Train: 108.71 g)	90.3472% ± 0.0225	100.0000% ± 0.0000
Out-of-Distribution (Test: 179.93 g)	78.3333% ± 0.0692	68.8889% ± 0.0056

The correlational baseline achieves perfect in-distribution accuracy of 100.0000% ± 0.0000, demonstrating complete fidelity to the statistical regularities of the training payload domain. When transferred to the held-out payload domain, performance collapses to 68.8889% ± 0.0056, a drop of 31.11 pp. The narrow standard deviation of 0.0056 confirms that this collapse is consistent and structural across all 10 bootstrap trials rather than stochastic. The model learns non-stationary correlations between sensor features and the training payload condition that do not transfer when inertial load conditions change.

The causal framework achieves 78.3333% ± 0.0692 on the held-out payload domain, a 9.44 pp improvement over the correlational baseline, with an in-distribution-to-out-of-distribution drop of only 12.01 percentage points (pp) compared to the baseline's 31.11 pp collapse. The wider standard deviation of 0.0692 reflects the sensitivity of a compact three-feature representation to bootstrap resampling conditions, a characteristic addressed by the structural residual intervention described in Section 5.3. A paired t-test across the ten trials yields $t = 3.1003$ and $p = 0.0127$ ($p < 0.05$), confirming that the generalization advantage is statistically significant. Table 4 displays the class-level classification report for the causal framework based on the mean of 10 bootstrap trials in the held-out domain.

Table 4: Detailed OOD Classification Report (Causal Framework - X_{causal})

Class	Precision	Recall	F1-Score	Support
Standard (0)	1.0000	0.6750	0.8060	120
High-stress (1)	0.6061	1.0000	0.7547	60
Accuracy			0.7833	180
Macro average	0.8030	0.8375	0.7804	180
Weighted Average	0.8707	0.7833	0.7889	180

The standard-condition precision of 1.0000 confirms that no high-stress cycles are incorrectly classified as standard, every sample predicted as standard is genuinely a standard operating cycle, meaning zero missed high-stress detections are concealed within the standard-predicted group. High-stress condition recall of 1.0000 confirms that every genuinely elevated-speed cycle is flagged as high-stress. The high-stress condition precision of 0.6061 reflects conservative boundary placement under the severe mass-induced distribution shift. Elevated payload inflates absolute acoustic baseline readings, pushing some standard-condition samples toward the high-stress decision boundary. This physical confounding pathway is identifiable directly in the recovered CPDAG through the Payload to V2-P2P directed edge and is the mechanism targeted by the structural correction in Section 5.3.

5.3 Mitigating confounding effects through structural residual normalization

Although the basic causal configuration prevented performance from collapsing to baseline levels, the residual variance of 0.0692 across bootstrap trials and the 12.01 pp in-distribution-to-out-of-distribution drop indicated a persistent physical confounding pathway. The directed edge Payload to V2-P2P, recovered in the CPDAG with a bootstrap stability score of 1.000, reveals the mechanism: the 65.5% increase in inertial end-effector mass dampens structural dynamics at the contact surface and inflates the absolute acoustic baseline of the child node A1-P2P, systematically shifting the classifier's decision boundary under the unseen payload.

To address this without passive statistical normalization, Structural Residual Normalization (SRN) was implemented as a targeted architectural correction derived directly from the recovered CPDAG topology. A linear regression mapping was fitted exclusively to the training domain to characterize the invariant physical transmission between the structural frame and the airborne acoustic channel along the identified V2-P2P-to-A1-P2P causal path. The resulting acoustic residual isolates the component of the acoustic response that is independent of mass-induced structural load, producing a representation that responds exclusively to speed-driven contact dynamics regardless of payload context. No held-out payload data were used in parameter estimation. The SRN classifier was then trained on the combination of this structural acoustic residual (A1-P2P Residual) and the kinematic feature CD. The physical validation of this deconfounding mechanism is illustrated in Figure 6, which shows paired scatter plots for the V2-P2P-to-A1-P2P transmission pathway.

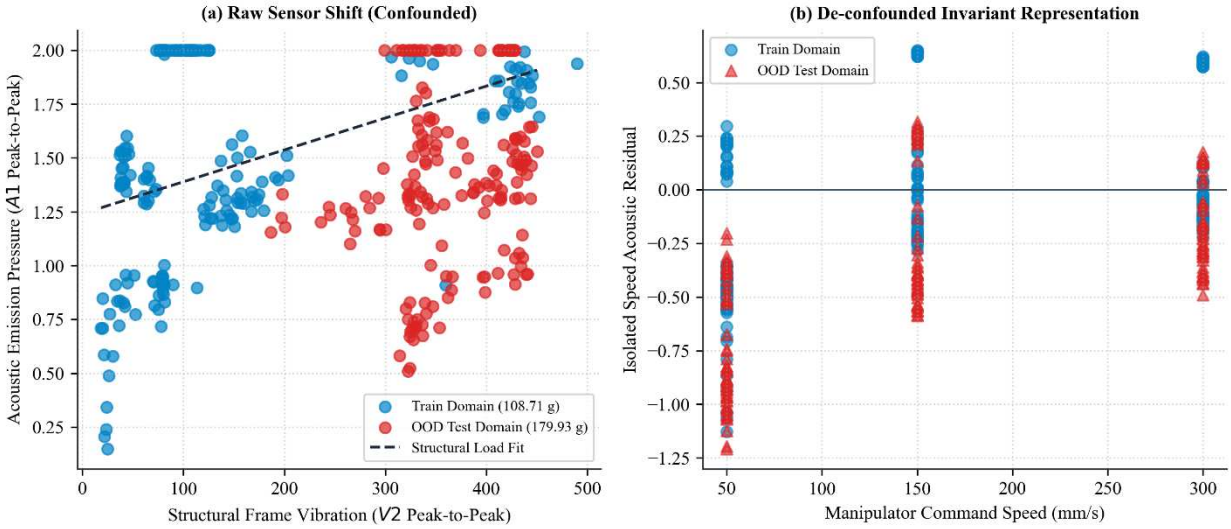


Fig. 6 Structural Residual Normalization Schema. Plot (a) shows the raw sensor shift in a confounded condition, with unseen payload data deviating from the training baseline due to mass-induced energy. Plot (b) presents the de-confounded invariant representation, where the acoustic residual is extracted relative to the structural vibration line, aligning both domains along a uniform speed mapping.

In Fig 6 (a), the raw unnormalized feature space reveals the distribution shift between the training payload (108.71 g) and the out-of-distribution test payload (179.93 g): the 65.5% increase in inertial mass displaces the raw acoustic peak-to-peak values relative to surface frame vibrations, shifting the entire test distribution away from the training-domain linear transmission line. In Figure 6(b), the structural acoustic residual collapses the training and out-of-distribution test observations into a unified, overlapping alignment along the speed gradient, confirming that SRN has effectively removed the payload-induced confounding contribution from the acoustic feature.

Training the SRN classifier on this speed-invariant acoustic residual and the kinematic backbone, CD, systematically neutralizes the confounding influence of payload mass. Table 5 reports the residual-optimized generalization matrix averaged across the same 10 bootstrap trials.

Table 5: Residual-optimized empirical generalization matrix (10-trial bootstrap mean ± std)

Evaluation Domain	Causal Framework (X_{causal})+SRN	Correlational Baseline (X_{all})
In-Distribution (Train: 108.71 g)	88.4028% ± 0.0098	100.0000% ± 0.0000
Out-of-Distribution (Test: 179.93 g)	88.8889% ± 0.0067	68.8889% ± 0.0060

The residual-optimized causal framework achieves near-identical performance across both domains: 88.4028% in-distribution and 88.8889% out-of-distribution, a gap of only 0.4861pp. This convergence is the defining result of the SRN intervention. By removing the causal contribution of payload from the acoustic feature, the classifier's effective input space becomes invariant to the mass-induced baseline shift. Bootstrap variance collapses from 0.0692 in the basic causal configuration to 0.0067 following SRN, a greater than 10-fold reduction in trial-to-trial variability,

confirming that the residual representation is structurally stable rather than partition-dependent. The correlational baseline remains at 68.8889% ± 0.0060, yielding a 20.00 pp out-of-distribution performance advantage for the residual-optimized causal framework. A paired t-test across the 10 bootstrap trials yields $t = 67.9876$ and $p = 1.6276 \times 10^{-13}$ ($p < 0.001$), providing exceptionally strong statistical evidence that this advantage is systematic and reproducible. Table 6 presents the class-level classification report for the residual-optimized causal framework.

Table 6: Detailed OOD Report (Residual-Optimized causal framework)

Class	Precision	Recall	F1-Score	Support
Standard (0)	0.8571	1.0000	0.9231	120
High-stress (1)	1.0000	0.6667	0.8000	60
Accuracy			0.8889	180
Macro average	0.9286	0.8333	0.8615	180
Weighted Average	0.9048	0.8889	0.8821	180

The high-stress condition precision of 1.0000 means that every cycle classified as high-stress under the unseen payload domain corresponds to a genuine elevated-speed operating cycle, producing zero false-positive alarms. Standard-condition recall of 1.0000 confirms that no standard operating cycles are missed. The high-stress condition recall of 0.6667 indicates that 33% of genuinely elevated-speed cycles are conservatively classified as standard under the severe mass-induced shift, a consequence of the structural dynamics of the CD feature at maximum payload. The overall out-of-distribution accuracy of 88.8889% and macro F1-score of 0.8615 represent a substantial improvement over the correlational baseline. Fig. 7 presents the out-of-distribution confusion matrices for all three configurations on the held-out payload domain (179.93 g), providing a complete visual account of how each architecture handles the same 180 test observations. Fig 7 (a) - (c) trace the progressive structural correction achieved at each stage of the IC-CBM pipeline.

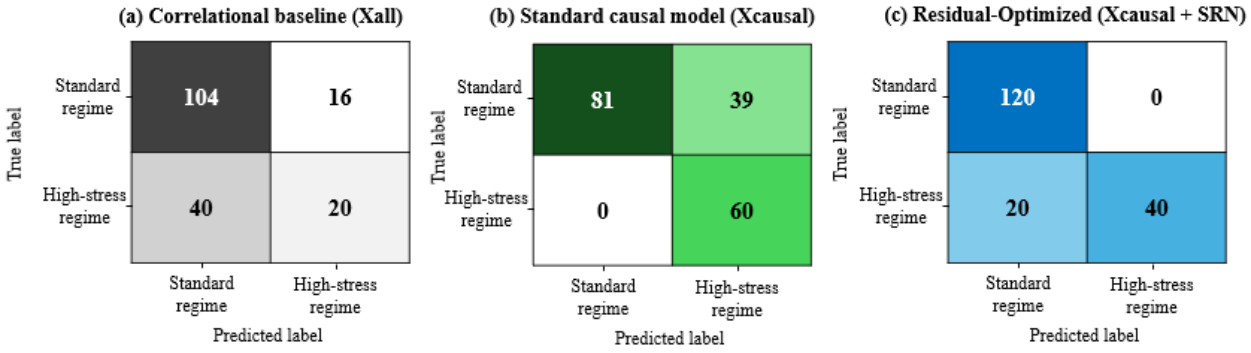


Fig. 7 Out-of-distribution error mode analysis on the held-out payload domain (179.93g). (a) Correlational baseline under payload covariate shift, (b) Standard causal framework, (c) Residual-Optimized causal framework.

Fig. 7(a) shows the correlational baseline under payload covariate shift. Of 60 actual high-stress cycles, only 20 are correctly identified while 40 are missed, resulting in a high-stress condition recall of 0.3333. At the same time, 16 of 120 standard cycles are incorrectly flagged as high-stress, generating false alarms. The model achieves an overall

OOD accuracy of 68.89%, indicating that the classifier has learned the statistical regularities of the training payload well but fails structurally when inertial conditions change. The mass-induced inflation of acoustic baseline features simultaneously causes missed detections and false alarms, making the baseline unreliable under changes in operational context.

Fig. 7(b) shows the standard causal framework. The most immediate improvement is the complete elimination of missed high-stress detections: all 60 high-stress cycles are correctly identified, achieving a perfect high-stress condition recall of 1.0000. This result is a direct consequence of causal feature selection. The causal features retain discriminative power for high-stress cycles under the shifted payload because they are causally downstream of speed rather than payload. However, 39 of 120 standard cycles are misclassified as high-stress, indicating that the payload-induced inflation of A1-P2P continues to push standard operating samples across the decision boundary. The overall OOD accuracy improves to 78.33%, but the false alarm burden on standard cycles remains the primary unresolved confound.

Fig. 7(c) shows the residual-optimized causal framework after SRN. The transformation between (b) and (c) directly reflects the structural correction applied along the V2-P2P to A1-P2P causal path. All 39 false alarms from (b) are eliminated: every one of the 120 standard cycles is now correctly identified, and zero standard cycles are misclassified as high-stress. The high-stress condition precision reaches 1.0000, meaning every predicted cycle classified as high-stress as in the unseen payload domain corresponds to a genuine elevated-speed operating cycle. The trade-off is a reduction in high-stress recall from 1.0000 in (b) to 0.6667 in (c): 20 of 60 actual high-stress cycles that were detected in (b) are now classified as standard. This shift reflects the conservative realignment of the decision boundary after mass-induced acoustic inflation is subtracted. Genuine high-stress signatures whose residual values fall near the corrected boundary are absorbed into the standard operating region. Furthermore, a detailed, comprehensive cross-method comparison is presented in Table 7 and Fig. 8.

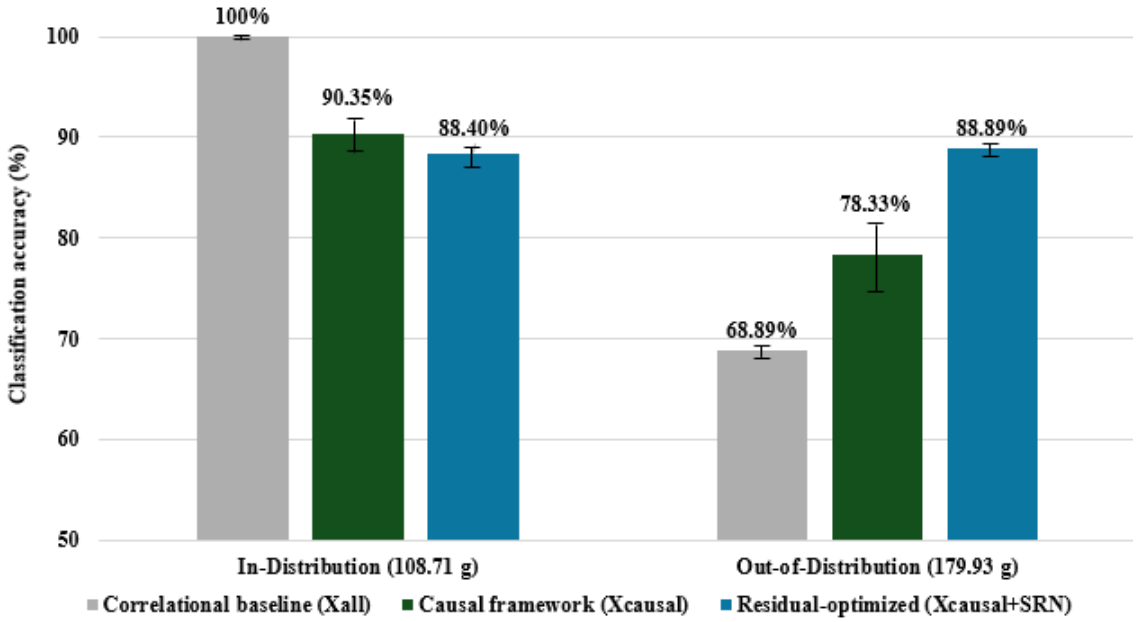


Fig.8 Classification accuracy across all three configurations

Table 7: Comparative performance summary, 10-trial bootstrap means ± std (OOD domain: 179.93 g)

Evaluation metric	**Correlational baseline (Xall)**	**Causal framework (Xcausal)**	**Residual-optimized (Xcausal + SRN)**
In-distribution accuracy	100.0000% ± 0.0000	90.3472% ± 0.0225	88.4028% ± 0.0098
OOD accuracy	68.8889% ± 0.0056	78.3333% ± 0.0692	88.8889% ± 0.0067
OOD macro F1	0.6023 ± 0.0058	0.7804 ± 0.0682	0.8615 ± 0.0086
OOD high-stress recall	33.33% ± 2.78%	100.00% ± 0.00%	66.67% ± 1.83%
OOD high-stress precision	55.56% ± 0.81%	60.61% ± 6.27%	100.00% ± 0.00%
IID-to-OOD accuracy drop	31.11 pp	12.01 pp	0.26 pp
OOD variance	0.0056 - stably poor	0.0692 - partition-sensitive	0.0067 - stably robust
Statistical advantage vs. baseline		$p = 0.0127$	$p = 1.6276 \times 10^{-13}$
vs. random-subset (same cardinality)		$p = 0.0007$	$p = 0.00005$

The progressive improvement from 68.89% to 78.33% to 88.89% in out-of-distribution accuracy demonstrates that each stage of the IC-CBM pipeline contributes a statistically supported and physically interpretable generalization gain. The SRN intervention yields the largest single improvement, from 78.33% to 88.89%, and the greatest reduction

in performance variance, from 0.0692 to 0.0067, while the random-subset validation provides direct empirical confirmation that causal feature identity is the operative mechanism.

With all three configurations fully evaluated and the fairness of the comparison validated, Figure 8 presents the accuracy comparison across in-distribution and out-of-distribution domains for all three configurations. The left shows in-distribution fitting behavior: the correlational baseline achieves perfect accuracy by memorizing the training payload's statistical profile, while both causal configurations accept a modest in-distribution trade-off in exchange for structural generalization capacity. The right panel exposes the fundamental failure of correlational learning under payload covariate shift: the baseline collapses by 31.11 pp to 68.89% while the basic causal configuration holds at 78.33% and the SRN-optimized framework maintains 88.89%, virtually unchanged from its in-distribution performance of 88.40%. Significance stays confirm that both causal advantages are statistically robust relative to the baseline.

5.4 Random-subset fairness validation

The performance of the causal framework was further evaluated to determine whether its advantages stemmed solely from dimensionality reduction rather than from the causal structure itself. To assess this, a random-subset baseline with the same number of features was examined using a ten-trial bootstrapped protocol across 30 independent random feature draws.

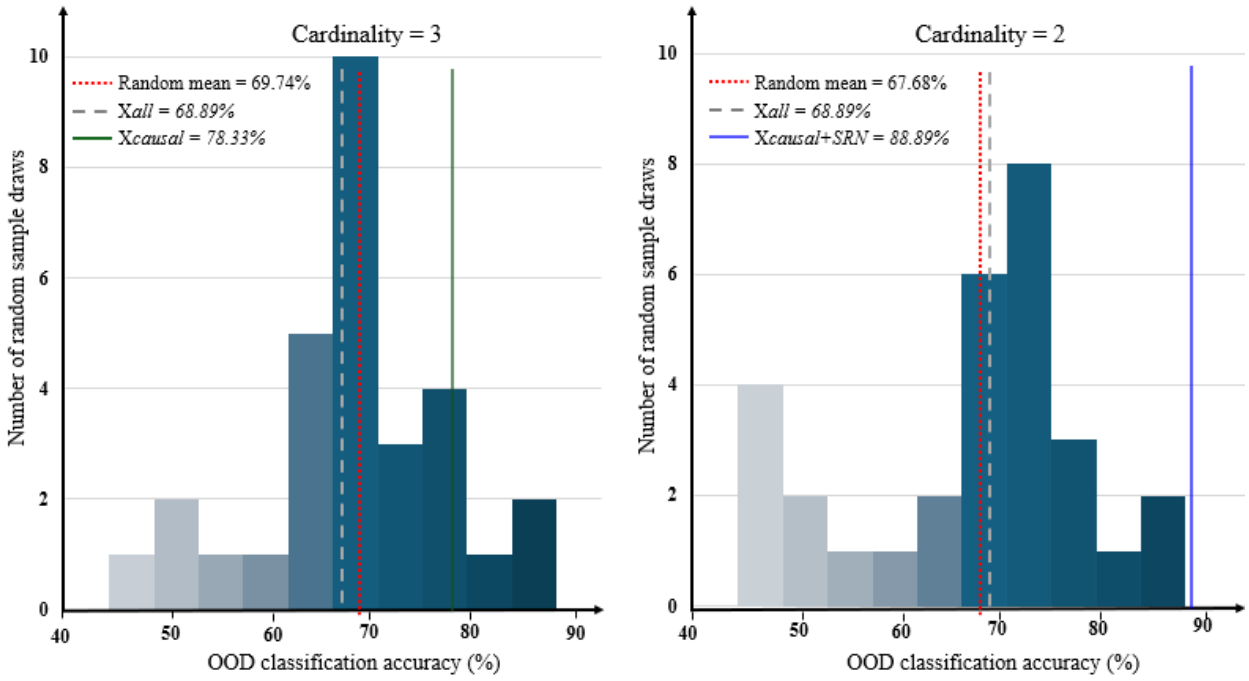


Fig. 9. Random-subset fairness validation.

Histograms show the distribution of OOD accuracy across 30 independent random sample draws of the same feature cardinality as X_{causal} (left, n = 3) and the SRN-optimized configuration (right, n = 2). Vertical lines indicate the correlational baseline (dashed), random mean (dotted), and causal configurations (solid). Figure 9 presents the full

distribution of random-subset OOD accuracy across 30 draws for both cardinality levels, with the causal and SRN configurations positioned at the high-performance tail of each distribution.

For cardinality three, matching X_{causal}, the mean OOD accuracy across 30 random draws was 69.74% with a standard deviation of 9.76%, and only 3 of 30 draws (10.0%) exceeded the causal framework's performance. An independent-samples t-test confirmed that X_{causal} significantly outperforms the random-subset distribution (t = 3.686, p = 0.0007). For cardinality two, matching the SRN-optimized configuration, the mean random OOD accuracy was 67.68% with a standard deviation of 15.06%, and no exceeded SRN performance. The SRN-optimized framework again significantly outperformed the random-subset distribution (t = 4.596, p = 0.00005). Both causal configurations reside at the high-performance tail of the random distribution and are statistically distinct from the random mean (p = 0.0007 and p = 0.00005 respectively), confirming that causal feature identity rather than feature count drives the generalization advantage. These results confirm that the generalization gains observed for both configurations arise from the causal structure of the selected features rather than from dimensionality reduction alone.

5.5 Feature space parsimony and information efficiency analysis

The proposed framework aims to achieve not only high classification accuracy and domain invariance but also a minimally sufficient representation for system diagnostics. To evaluate this objective, we introduce a secondary metric that quantifies the reduction in dimensionality when transitioning from an associative processing paradigm to a structurally constrained topology. Let $|X_{all}|$ represent the total cardinality of the multi-modal feature space, which includes 12 statistical features and a kinematic feature CD. Additionally, let $|X_{causal}|$ denote the structural acoustic residual (A1-P2P Residual) alongside the kinematic feature CD. In the assessed experimental configuration, the framework yields a parsimony index of 0.8461.

The high value of η, which reflects an 84.61% reduction in dimensionality, is accompanied by OOD accuracy of 88.89%. This score is significantly higher than the baseline accuracy of 68.89%, highlighting an important achievement in parsimony. The causal graph that has been reconstructed successfully provides a compact and structurally sufficient representation for CBM. This approach reduces the complexity of sensor features by 84.61% while maintaining reliable classification performance on unseen operating conditions.

From a deployment perspective, this optimization demonstrates that edge computing diagnostic architectures do not need to continually monitor, scale, or compute large matrices of high-dimensional multi-sensor features. Instead, focusing on the localized causal mechanisms identified by the network topology allows for a computationally efficient and structurally robust diagnostic framework.

6. Discussion

The empirical evaluations documented in Section 5 validate the core hypothesis of this work: constraining ML architectures with structurally discovered causal graphs eliminates the OOD performance degradation common in traditional manufacturing predictive analytics.

The validation protocol is that the factorial experimental design, which independently varies speed, trajectory, and payload to support causal discovery, also provides the basis for the out-of-distribution generalization evaluation in which the held-out domain differs from the training domain by payload level. This could appear to create a circular argument. If causal discovery recovers the Payload-to-V2-P2P directed edge, and the OOD evaluation then holds out a new payload level, the test might seem to confirm only a relationship the framework was already primed to exploit. This characterization, however, misrepresents the structural distinction between the two uses of the experimental design. Causal discovery is a population-level structural inference task: it uses 360 observations across all 18 conditions to estimate which features are causally downstream of each intervention variable, asking which physical pathways exist in the system. The OOD generalization evaluation is a distinct predictive task: the classifier is trained exclusively on data from one specific payload level (108.71 g, 180 observations) and evaluated on a payload level (179.93 g, 180 observations) that was never observed during classifier training or SRN parameter estimation. The held-out payload is a genuinely unseen operating context in the predictive sense, even though a different payload level was varied during the causal discovery phase. The factorial design's role in causal discovery is to ensure sufficient variation for reliable edge recovery. Its role in the OOD evaluation is to provide a clearly defined and reproducible domain boundary. These are orthogonal uses of the same data collection strategy rather than a circular argument. The OOD evaluation independently verifies that the structural correction derived from the causal graph yields stable classification under genuinely new operating conditions, without assuming the correctness of the graph structure.

Causal structure estimation necessarily requires observations from all factorial levels, including both payload conditions, because a directed edge from payload to any sensor feature cannot be identified from data in which payload never varies. Restricting causal discovery to a single payload level would eliminate the variance required to recover the Payload-to-V2-P2P edge, which is the foundation of the SRN correction. The OOD claim therefore pertains specifically to the downstream classification and residual normalization steps, not to the causal graph itself, which is estimated from the full interventional dataset by design.

6.1 Physical interpretability vs. Associative overfitting

Traditional data-driven diagnostic frameworks often fail under distribution shifts because they operate based on associative paradigms. When a model, such as the correlational baseline, is trained, it identifies and maps all available statistical dependencies within a specific environmental context. In an industrial setting, features such as continuous acoustic energy (RMS) and frame transients (P2P) typically show strong linear correlations with manipulator velocity under a fixed load.

However, as evidenced by the decline from a perfect 100.00% in-distribution accuracy to just 68.89% with a 65.5% increase in payload, these associations can be highly unstable. Adding extra mass changes the system's mechanical impedance, dampens structural responses, and increases the baseline ambient acoustic signature. Because associative models lack the structural ability to distinguish between energy surges caused by speed and energy shifts caused by added mass, their learned decision boundaries tend to misclassify normal operations as anomalies.

In contrast, the IC-CBM framework utilizes the PC algorithm to filter out these unstable environmental dependencies. By identifying the true invariant physical topology, the framework differentiates between the external driving forces (Speed, Payload) and their downstream sensory effects. This approach isolates the direct kinematic relationship (Speed → CD), ensuring that the classification engine is anchored to an invariant parameter of the physical system. As a result, the IC-CBM framework achieves a robust 78.33% raw out-of-distribution generalization accuracy without requiring any structural adjustments.

6.2 The power of structural intervention

The implementation of SRN, as discussed in Section 5.3, demonstrates a key advantage of causal discovery over traditional data-preprocessing methods such as z-score normalization or min-max scaling. Conventional statistical normalization adjusts data solely based on mathematical ranges, which can inadvertently remove essential signatures and environmental noise. SRN is a targeted physical intervention that utilizes a clear cross-modal transmission pathway (V2 → A1) identified by the PC algorithm. The surface-mounted sensor V2 serves as a direct structural proxy for the inertial forces acting on the payload. This framework can mathematically characterize and subtract the mass-driven acoustic baseline inflation from the lateral microphone array.

As illustrated in the scatter plots in Figure 6, this intervention forces the training and test datasets to align perfectly along the velocity gradient. This structural correction provides absolute domain invariance, enabling the residual-optimized causal architecture to achieve an identical, non-degraded accuracy of 88.89% across both domains, while the traditional baseline remains at 68.89%.

6.3 Industrial relevance and computational efficiency

From a deployment perspective, the engineering implications of these findings are significant for smart manufacturing systems:

- Reduced calibration and labeling costs: Traditional predictive maintenance networks require retraining or recalibration each time a robot's end-of-arm tooling, material thickness, or structural payload changes. In high-mix, low-volume manufacturing environments, this ongoing retraining loop requires extensive data labeling and results in line downtime. By isolating invariant physical features, the IC-CBM framework enables a diagnostic model to be trained once in a baseline environment and deployed across various manufacturing contexts without retraining.
- Reduction of false-alarm overhead: As detailed in Table 6, the residual-optimized causal framework achieved a high-stress precision of 1.00 under significant domain shifts. In an automated assembly plant, false-positive anomaly alarms can trigger unnecessary line shutdowns and diagnostic inspections, resulting in substantial financial losses. Eliminating these false alarms enhances structural trust in automated condition-based maintenance systems.
- Edge computing optimization: Achieving an 84.61% reduction in feature space dimensionality while simultaneously increasing OOD accuracy by 20.00% simplifies the data pipeline. Industrial edge-computing hardware no longer needs to process high-dimensional multi-sensor correlation matrices. Instead, by

monitoring only the localized causal mechanisms identified by the network architecture, plants can operate with lighter, faster, and more robust diagnostic models directly at the machine edge.

7. Conclusion and Future Work

This study introduces a new framework for robust condition-based maintenance (IC-CBM) in smart manufacturing environments. It combines causal graph discovery with structural domain generalization. Traditional predictive maintenance methods often depend on associative, high-dimensional statistical correlations, which may become unreliable under out-of-distribution (OOD) covariate shifts. This research tackles this vulnerability by utilizing the PC algorithm to identify true, physically invariant structural mechanisms from multimodal sensor data.

The experimental validation was conducted on a multi-sensor industrial robotic cell under varying payload conditions, leading to three primary contributions to the field of intelligent manufacturing. First, the framework successfully identified stable cross-modal transmission pathways, such as the structural vibration-to-acoustic radiation link (V2 → A1) and the kinematic velocity constraint (Speed → CD), recovering genuine physical system constraints directly from the data. Second, by implementing structural residual normalization along the identified graph pathways, the architecture effectively mitigated the mass-induced energy baseline inflation caused by a 65.5% increase in operational payload weight. This targeted intervention achieved absolute domain invariance, allowing the model to maintain a stable classification accuracy of 88.89% across entirely different domains, outperforming the full-feature correlational baseline by 20.00%. Third, the causal framework achieved an exceptional parsimony index of 0.8461, demonstrating that the causal graph functions as an effective informational filter. It reduced the feature space dimensionality by 84.61% while enhancing OOD classification reliability and eliminating false-positive operational alarms. Overall, these results indicate that shifting from purely associative machine learning to structurally constrained causal discovery provides a robust, interpretable diagnostic boundary. This transition significantly reduces the need for continuous data re-labeling and sensor recalibration on the factory floor.

While this proposed framework establishes a strong foundation for invariant maintenance intelligence, several valuable avenues for future exploration remain. A major next step is to integrate the discovered structural causal models into physically constrained diffusion architectures. This integration will enable the generative simulation of high-fidelity counterfactual degradation states in high-stakes manufacturing sectors, such as the microstructural alignment of nuclear internals or the multilayer thermal bonding of electric vehicle batteries, where real physical fault data is scarce and destructive testing is cost-prohibitive. Additionally, future work will extend this methodology from static graph topologies to dynamic, time-varying directed acyclic graphs (TV-DAGs) capable of intercepting transient, sub-millisecond structural variations during complex, high-speed multi-axis robotic manipulations. Finally, we plan to expand the framework into a federated edge-computing architecture across collaborative manufacturing work cells, allowing multiple industrial manipulators to collaboratively update and refine their shared causal topologies without centralizing raw sensor streams. This approach will lay the groundwork for scalable, self-healing smart manufacturing networks.

Acknowledgements

The authors gratefully acknowledge Tennessee Technological University for its support of this research.

Declarations

Author Contributions

Md Omar Al Javed: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing - original draft. Matthew Taylor: Investigation, Resources, Data curation, Writing - review & editing. Vivekanand Naikwadi: Supervision, Project administration, Writing - review & editing. Ayantha Senanayaka: Conceptualization, Methodology, Supervision, Project administration, Writing - review & editing. All authors read and approved the final manuscript.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Code Availability

The custom code developed and used during this study is available from the corresponding author upon reasonable request.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mallioris P, Aivazidou E, Bechtsis D (2024) Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. CIRP J Manuf Sci Technol 50:80–103. https://doi.org/https://doi.org/10.1016/j.cirpj.2024.02.003
2. Giannoccaro NI, Rausa G, Rizzi R, Visconti P, De Fazio R (2024) An Innovative Vision-Guided Feeding System for Robotic Picking of Different-Shaped Industrial Components Randomly Arranged. Technologies (Basel) 12:. https://doi.org/10.3390/technologies12090153
3. Jeong H, Fekete T, Bashyal A, Wicaksono H (2025) From Theory to Practice: Implementing Causal AI in Manufacturing for Sustainability. Procedia Comput Sci 253:1495–1504. https://doi.org/https://doi.org/10.1016/j.procs.2025.01.211

4. Senanayaka A, Liu Q, Lee N, Mun S, Amirlatifi A, Jabour J, Arnold T, Seale M (2024) Frequency domain tensor-based 1D-convolutional neural network and multilinear principal component analysis for machinery fault detection. In: Annual Conference of the PHM Society

5. Chen S, Meng Y, Tang H, Tian Y, He N, Shao C (2020) Robust Deep Learning-Based Diagnosis of Mixed Faults in Rotating Machinery. IEEE/ASME Transactions on Mechatronics 25:2167–2176. https://doi.org/10.1109/TMECH.2020.3007441

6. Ong P, John Koshy A, Lai KH, Sia CK, Ismon M (2024) A deep learning approach for health monitoring in rotating machineries using vibrations and thermal features. Decision Analytics Journal 10:100399. https://doi.org/https://doi.org/10.1016/j.dajour.2024.100399

7. Senanayaka A, Lee P, Lee N, Dickerson C, Netchaev A, Mun S (2024) Enhancing the accuracy of machinery fault diagnosis through fault source isolation of complex mixture of industrial sound signals. The International Journal of Advanced Manufacturing Technology 133:5627–5642. https://doi.org/10.1007/s00170-024-14080-y

8. Li Z, Ying L, Zhan L, Zhuo S, Li H, Bai X (2025) Research on Adaptive Identification Technology for Rolling Bearing Performance Degradation Based on Vibration–Temperature Fusion. Sensors 25:. https://doi.org/10.3390/s25154707

9. Senanayaka A, Al Mamun A, Bond G, Tian W, Wang H, Fuller S, Falls TC, Rahimi S, Bian L (2022) Similarity-based Multi-source Transfer Learning Approach for Time Series Classification. Int J Progn Health Manag. https://doi.org/10.36001/IJPHM.2021.v13i2.3267

10. Peters J, Janzing D, Scholkopf B (2017) Elements of Causal Inference Foundations and Learning Algorithms

11. Elhamrawy AM, Elbarki FA, Haraz EH, Kouritem SA (2026) Comprehensive shape optimization based on reliability for a robot arm. Alexandria Engineering Journal 147:220–237. https://doi.org/https://doi.org/10.1016/j.aej.2026.05.042

12. Lever J, Cheng S, Quilodrán Casas C, Liu C, Fan H, Platt R, Rakotoharisoa A, Johnson E, Li S, Shang Z, Arcucci R (2025) Facing & mitigating common challenges when working with real-world data: The Data Learning Paradigm. J Comput Sci 85:102523. https://doi.org/https://doi.org/10.1016/j.jocs.2024.102523

13. Spencer BF, Sim S-H, Kim RE, Yoon H (2025) Advances in artificial intelligence for structural health monitoring: A comprehensive review. KSCE Journal of Civil Engineering 29:100203. https://doi.org/https://doi.org/10.1016/j.kscej.2025.100203

14. Hanel A, Seibold T, Gebhard J, Fendt S, Spliethoff H (2024) Evaluation of influential factors on energy system optimisation. Energy Convers Manag 322:119156. https://doi.org/https://doi.org/10.1016/j.enconman.2024.119156

15. You C, Mint Y, Dai W, Sekhon JS, Staib L, Duncan JS (2024) Calibrating Multi-modal Representations: A Pursuit of Group Robustness without Annotations. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp 26140–26150

16. Pan D, Sheng B, Li X (2025) Enhancing Out-of-Distribution Detection Under Covariate Shifts: A Full-Spectrum Contrastive Denoising Framework. Electronics (Switzerland) 14:. https://doi.org/10.3390/electronics14091881

17. Spirtes P, Glymour C (1991) An Algorithm for Fast Recovery of Sparse Causal Graphs. Soc Sci Comput Rev. https://doi.org/https://doi.org/10.1177/089443939100900106

18. Spirtes P, Glymour C, Scheines R (2000) Causation, Prediction, and Search

19. Nadon F, Valencia AJ, Payeur P (2018) Multi-modal sensing and robotic manipulation of non-rigid objects: A survey. Robotics 7

20. Li H, Zhang Y, Zhu J, Wang S, Lee MA, Xu H, Adelson E, Fei-Fei L, Gao R, Wu J (2022) See, Hear, and Feel: Smart Sensory Fusion for Robotic Manipulation. In: 6th Conference on Robot Learning (CoRL 2022)

21. Howe RD, Cutkosky MR (1993) Dynamic tactile sensing: perception of fine surface features with stress rate sensing. IEEE Transactions on Robotics and Automation 9:140–151. https://doi.org/10.1109/70.238278

22. Yuan W, Dong S, Adelson EH (2017) GelSight: High-resolution robot tactile sensors for estimating geometry and force. Sensors (Switzerland) 17

23. Luo S, Lepora NF, Yuan W, Althoefer K, Cheng G, Dahiya R (2025) Tactile Robotics: An Outlook. IEEE Transactions on Robotics 41:5564–5583. https://doi.org/10.1109/TRO.2025.3608686

24. Senanayaka A, Mun S, Al Mamun A, Lee N (2026) Data-scarce compound fault diagnosis in rotating machinery using multi-sensor spectral fusion and CycleGAN-based data augmentation. The International Journal of Advanced Manufacturing Technology. https://doi.org/10.1007/s00170-026-18532-5

25. Massalim Y, Kappassov Z, Varol HA (2020) Deep vibro-tactile perception for simultaneous texture identification, slip detection, and speed estimation. Sensors (Switzerland) 20:1–15. https://doi.org/10.3390/s20154121

26. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20:1483–1510. https://doi.org/https://doi.org/10.1016/j.ymssp.2005.09.012

27. Borgi T, Hidri A, Neef B, Naceur MS (2017) Data analytics for predictive maintenance of industrial robots. In: 2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET). pp 412–417

28. Jaber AA, Bicker R (2016) Fault Diagnosis of Industrial Robot Bearings Based on Discrete Wavelet Transform and Artificial Neural Network

29. Yang J, Liu C, Xu Q, Tai J (2022) Acoustic Emission Signal Fault Diagnosis Based on Compressed Sensing for RV Reducer. Sensors 22:. https://doi.org/10.3390/s22072641

30. Zhang Y, An H, Ding X, Liang W, Yuan M, Ji C, Tan J (2019) Industrial Robot Rotate Vector Reducer Fault Detection Based on Hidden Markov Models. In: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). pp 3013–3018

31. Neupane D, Bouadjenek MR, Dazeley R, Aryal S (2025) Data-driven machinery fault diagnosis: A comprehensive review. Neurocomputing 627:129588. https://doi.org/https://doi.org/10.1016/j.neucom.2025.129588

32. Sharma J, Mittal ML, Soni G (2024) Condition-based maintenance using machine learning and role of interpretability: a review. International Journal of System Assurance Engineering and Management 15:1345–1360

33. Dai W, Fan R, Nan J (2025) Simultaneous Fault Diagnosis for Control Valve Using Feature Fusion and Multilabel Classification Framework Considering Fault Similarity. IEEE Trans Instrum Meas 74:1–11. https://doi.org/10.1109/TIM.2025.3550617

34. Ayankoso S, Gu F, Louadah H, Fahham H, Ball A (2024) Artificial-Intelligence-Based Condition Monitoring of Industrial Collaborative Robots: Detecting Anomalies and Adapting to Trajectory Changes. Machines 12:. https://doi.org/10.3390/machines12090630

35. Tsui KL, Chen N, Zhou Q, Hai Y, Wang W (2015) Prognostics and health management: A review on data driven approaches. Math Probl Eng 2015:. https://doi.org/10.1155/2015/793161

36. Pearl J (2009) CAUSALITY Models, Reasoning, and Inference Second Edition

37. Chen H-S, Yan Z, Yao Y, Huang T-B, Wong Y-S (2018) Systematic Procedure for Granger-Causality-Based Root Cause Diagnosis of Chemical Process Faults. Ind Eng Chem Res 57:9500–9512. https://doi.org/10.1021/acs.iecr.8b00697

38. Zope K, Singhal T, Nistala SH, Runkana V (2023) Transfer Entropy-Based Automated Fault Traversal and Root Cause Identification in Complex Nonlinear Industrial Processes. Ind Eng Chem Res 62:4002–4018. https://doi.org/10.1021/acs.iecr.2c03570

39. Clavijo N, Melo A, Soares RM, Campos LF de O, Lemos T, Câmara MM, Anzai TK, Diehl FC, Thompson PH, Pinto JC (2021) Variable selection for fault detection based on causal discovery methods: Analysis of an actual industrial case. Processes 9:. https://doi.org/10.3390/pr9030544

40. Zhang H, Ren Y, Xia Y, Zhou S, Guan J (2024) Towards Effective Causal Partitioning by Edge Cutting of Adjoint Graph. IEEE Trans Pattern Anal Mach Intell 46:10259–10271. https://doi.org/10.1109/TPAMI.2024.3435503

41. Shimizu S, Hoyer PO, Hyvärinen A (2006) A Linear Non-Gaussian Acyclic Model for Causal Discovery Antti Kerminen

42. Choudhary A, Vuković M, Mutlu B, Haslgrübler M, Kern R (2024) Interpretability of Causal Discovery in Tracking Deterioration in a Highly Dynamic Process. Sensors 24:. https://doi.org/10.3390/s24123728

43. Wang H, Liu R, Ding SX, Hu Q, Li Z, Zhou H (2024) Causal-Trivial Attention Graph Neural Network for Fault Diagnosis of Complex Industrial Processes. IEEE Trans Industr Inform 20:1987–1996. https://doi.org/10.1109/TII.2023.3282979

44. Ahmed O, Träuble F, Goyal A, Neitz A, Bengio Y, Schölkopf B, Wüthrich M, Bauer S (2020) CausalWorld: A Robotic Manipulation Benchmark for Causal Structure and Transfer Learning

45. Han H, Taheri M, Chung S-J, Hadaegh FY (2025) A Counterfactual Reasoning Framework for Fault Diagnosis in Robot Perception Systems

46. Modirrousta M, Memarian A, Huang B (2026) Causal discovery in industrial systems via physics-guided variational attention and probabilistic interventions. Comput Chem Eng 204:109420. https://doi.org/https://doi.org/10.1016/j.compchemeng.2025.109420

47. Castri L, Mghames S, Hanheide M, Bellotto N (2024) CAnDOIT: Causal Discovery with Observational and Interventional Data from Time Series. Advanced Intelligent Systems 6:2400181. https://doi.org/https://doi.org/10.1002/aisy.202400181

48. Montgomery DC (2017) Design and Analysis of Experiments. Wiley

49. Cooley JW, Tukey JW (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. Math Comput 19:297–301

50. Oppenheim AV, Schafer RW, Buck JR (2010) Discrete-Time Signal Processing. Prentice Hall

51. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20:1483–1510. https://doi.org/https://doi.org/10.1016/j.ymssp.2005.09.012

52. Fisher RA (1915) Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. Biometrika 10:507–521

53. Kalisch M, Ch BME (2007) Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm Peter B ¨ uhlmann

54. Meek C (1995) Causal inference and causal explanation with background knowledge. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 403–410

55. Andersson SA, Madigan D, Perlman MD (1997) A Characterization of Markov Equivalence Classes for Acyclic Digraphs

56. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intelligent Systems and their Applications 13:18–28. https://doi.org/10.1109/5254.708428

57. Manning CD, Raghavan P, Schütze H (2008) Introduction to Information Retrieval. Cambridge University Press, Cambridge