# Sensor Choice Reverses Under Operating Condition Shift for Robotic Condition Classification

**Md Omar Al Javed[1], 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 authors. E-mail: asenanayaka@tntech.edu

## Abstract

Vibration and acoustic sensing are established nondestructive evaluation modalities for machine state estimation, and their fusion is now standard practice in condition monitoring. What this practice does not establish is what each modality independently contributes, or whether its in-distribution advantage survives an operating-condition shift, a question with direct cost consequences. This paper treats sensing modality as the experimental variable, evaluating vibration-only, acoustic-only, and fused configurations from a delta robot against one fixed classifier, scored on in-distribution and out-of-distribution accuracy under a held-out payload, combined with a cost model into a Pareto comparison. Vibration-only outperforms acoustic-only in-distribution (98.5% vs. 90.0% balanced accuracy), but under payload shift the two become statistically indistinguishable (54.2% vs. 54.7%), despite acoustic costing half as much, dominating vibration on the frontier. Fusion achieves the highest accuracy in both regimes but the lowest accuracy-per-dollar yield. An eight-classifier comparison shows the same reversal: the strongest in-distribution classifier is not the most shift-robust. We further prove a fused configuration's population risk cannot exceed any single modality it contains, turning an observed single-modality advantage into a predictable finite-sample effect. These results show evaluating modality and classifier choice in-distribution alone yields a different, more expensive recommendation than a shift-aware evaluation supports.

**Keywords:** sensor fusion; nondestructive evaluation; distribution shift; instrumentation cost; robotic condition monitoring; classifier robustness

## 1. Introduction

Vibration and acoustic sensing are the most commonly used nondestructive evaluation (NDE) methods for assessing the operational condition of mechanical and robotic systems without disrupting production [1]. As mechanical and robotic systems are pushed to operate collaboratively and faster, following more varied actions, the NDE sensing systems that monitor them face two significant challenges. First, operating conditions constantly change. The speed, trajectory shape, and payload can vary from one production run to another [2]. This means a model trained under a specific set of conditions often must make predictions in different scenarios it did not encounter during training. This issue, known as out-of-distribution generalization, has historically received less rigorous attention in NDE evaluation practices than in common sensor-fusion NDE, which informs many instrumentation decisions. Second, the instrumentation used to observe the robot carries a real capital cost, and that cost is rarely weighed against the diagnostic value it returns. An industry-grade accelerometer is installed because it is standard practice, not because its incremental accuracy over a cheaper alternative has been measured. These two problems are usually studied separately.

Multi-sensor fusion is now the dominant paradigm for addressing the first problem. Systematic reviews of multi-sensor information fusion for equipment fault diagnosis document a large and rapidly growing body of fusion architectures spanning data, feature, and decision-level integration [3–6], motivated by the observation that a single sensor has limited spatial coverage and cannot by itself capture the full operating signature of complex machinery [7–9]. Concrete instances of this paradigm combine vibration and acoustic (or other heterogeneous) channels via nonlinear dynamical indicators [10], hypergraph fusion [11], multi-attention residual networks [12], and semi-supervised temperature-guided frameworks [13], across application domains including turning [14], additive manufacturing [15], hydraulic systems [16], and complex electromechanical equipment [17]. Researchers have thoroughly investigated condition monitoring for different manufacturing applications [18]. In particular, studies have used wavelet scattering networks with traditional time-domain features for robot fault diagnosis [19]. Additionally, researchers have integrated hybrid pipelines that combine PCA, LDA, and meta-learning techniques into fault-tolerant control systems [20, 21]. Across this literature, the consistent empirical finding is that fusion outperforms any single modality evaluated alone [4, 8–10, 22]. A largely separate literature addresses the fact that operating conditions change. Domain generalization for fault diagnosis has established, most directly in an application-oriented benchmark study [23], that performance under operating conditions unseen during training can diverge sharply from in-distribution performance, and a substantial body of algorithmic work has proposed remedies: domain augmentation networks [24], meta-learning with gradient alignment [25], adversarial domain-augmented generalization [26], feature-decoupling networks [27], relationship-transfer networks [28], and dual-disentanglement cross-domain fusion [29], among others [30, 31]. In addition, another part of the literature treats instrumentation cost directly, such as open-hardware, low-cost vibro-acoustic sensing platforms [32]. Non-intrusive, low-cost data-acquisition systems for machining monitoring [33] demonstrate that consumer-grade sensing can approach research-grade bandwidth at a fraction of the price. In contrast, a distinct engineering literature formalizes sensor selection itself as a cost-constrained multi-objective (Pareto) optimization problem [34–36].

Three limitations identified across this research domain summarized above, and together they define the challenge this paper addresses. First, where single-modality baselines are reported in the fusion literature, they are almost always reported to demonstrate that fusion helps, not to characterize what each channel independently contributes. A result showing fusion beats each modality alone does not establish whether the weaker channel is redundant, marginal, or essential, and does not indicate whether that ranking would survive a change in operating conditions. Second, the domain generalization literature primarily focuses on algorithmic solutions, including new architectures, loss functions, and adaptation procedures, which are applied to a fixed sensing configuration. However, it often overlooks the fundamental empirical question of whether the sensing modality is inherently more or less stable under shifts, regardless of the adaptation algorithm used. Third, the literature on cost-aware instrumentation and accuracy-focused sensor fusion rarely intersects. Low-cost sensing studies typically treat cost as a characteristic of the proposed hardware rather than a variable compared with alternatives. Additionally, to our knowledge, Pareto-based sensor selection methods have not been applied to the decision between using an acoustic sensing chain and a vibration sensing chain to classify robotic operating conditions. Furthermore, these methods do not provide a clear metric for assessing out-of-distribution risk since they mainly concentrate on in-distribution diagnostic information. This paper addresses this gap by evaluating the sensing modality itself while keeping the classifier fixed.

In this study, the sensing configuration is treated as an experimental variable rather than a fixed design choice made once and then reported. A single family of classifiers was evaluated with hyperparameters kept constant across all configurations being compared. The evaluation was conducted using three different feature partitions, including vibration-only, acoustic-only, and a combination of both (fusion). This approach ensures that observed performance differences reflect the sensing channel rather than variations in modeling effort. Each configuration is assessed under two complementary risks: in-distribution accuracy and out-of-distribution accuracy, evaluated under a held-out operating condition. The accuracy is reported

in conjunction with instrumentation costs as part of a Pareto set, rather than as a single accuracy figure justified post hoc. Additionally, in-distribution risk is estimated using repeated stratified cross-validation, while out-of-distribution risk is evaluated through leave-one-covariate-level out transfer. Moreover, the formulation of the problem includes a distribution-free argument explaining why the population risk of a fused configuration cannot exceed that of any single modality it contains, provided the underlying classifier family is closed under coordinate projection. This reframes an observed advantage of a single modality on finite data as an estimation effect, rather than as evidence that the added modality carries no additional information.

The methodology is validated on a fully factorial dataset comprising 360 runs, collected from a delta-configuration parallel robot equipped with a two-channel research-grade vibration sensor and a two-channel consumer-grade condenser microphone pair. This dataset spans 3 end-effector speeds, 3 trajectory profiles, and 2 payload levels, each replicated 20 times. The payload was controlled by intervention rather than merely being observed, allowing it to serve as a genuine held-out condition test, rather than just a resampling split. Three modality configurations are compared. The resulting modality ranking is also cross-checked against five other classifier categories, including boosting ensembles, kernel methods, linear models, instance-based learners, and probabilistic classifiers under an identical evaluation protocol. This ensures the study's central findings do not depend on the specific classifier chosen for the main comparison. Each configuration is evaluated using balanced accuracy, macro-averaged F1, precision, and recall, with 95% confidence intervals obtained through percentile bootstrap over pooled held-out predictions. This allows for direct reading of overlapping or non-overlapping intervals between configurations as evidence for or against statistically significant differences. In addition to accuracy, two derived statistics connect each configuration's risk estimates to deployment-relevant questions: Retained Capacity, the fraction of in-distribution accuracy that persists after the operating condition shift, and Risk Spread, which indicates whether that degradation depends on which condition is withheld. Furthermore, the accuracy is combined with an additive instrumentation-cost model to create a formal cost-accuracy Pareto comparison across the three sensing configurations, with a scalarized yield metric reported as a secondary ranking aid. The comparison study extends this protocol across the multiple classifiers mentioned earlier, isolating whether the modality-level findings are artifacts of the classifiers used to obtain them, a check that the existing fusion and domain-generalization literature rarely considered, to our knowledge, report at this joint modality-and-classifier granularity. The main contributions of this research are as follows:

1. A cost- and shift-aware modality ablation methodology, applicable to any factorial multi-sensor operating-condition dataset, that evaluates sensing configuration as an independent variable against a fixed classifier under both in-distribution and interventional out-of-distribution risk. This is formalized with distribution-free results showing that an observed single-modality advantage on finite data is an estimation effect, not evidence against the value of the added modality.
2. Empirical evidence from a delta-robot vibro-acoustic dataset indicating that the in-distribution ranking of a research-grade vibration sensor over a consumer-grade acoustic pair does not hold after an operating-condition shift. Furthermore, this reversal occurs at the classifier level, such that the strongest in-distribution classifier is not necessarily the most robust to shifts, which changes the instrumentation and modeling recommendations that an in-distribution-only evaluation would support.
3. A reporting scheme that explicitly presents the accuracy-cost trade-off through a formal cost-accuracy Pareto analysis, rather than a single post hoc cost disclosure. This demonstrates that the lower-cost acoustic-only configuration is not as advantageous as initially assumed.

The remainder of this paper is organized as follows. Section 2 reviews the multimodal sensor fusion, domain generalization, and cost-aware sensing literature summarized above in more depth. Section 3 gives the general, dataset-independent mathematical formulation of the proposed modality-ablation methodology. Section 4 describes the experimental platform and dataset and reports the in-distribution, out-of-

distribution, feature-attribution, and classifier-comparison results. Section 5 discusses the implications of these results for instrumentation decisions and positions them against the limitations identified. Section 6 concludes and outlines directions for future work.

## 2. Related Work

### 2.1 Multimodal sensing for machine and robot condition monitoring

Multimodal sensing, such as combining vibration and acoustic channels is now standard practice in condition monitoring. A recent systematic review of multi-sensor information fusion for equipment fault diagnosis [37] surveys fusion architectures across data, feature, and decision-level integration, and a broader synthesis focused on multimodal sensor fusion for fault diagnosis and prognostics catalogues reporting-standard pitfalls across nearly 1,200 studies published since 2015 [3]. Comprehensive reviews of rotating machinery indicate that multi-sensor fusion has become the dominant approach. This shift derives from the understanding that a single sensor offers limited coverage and mounting-location constraints that a complementary second sensor can alleviate [8]. Concrete fusion architectures demonstrate the breadth of the field. They include feature-level fusion that employs a novel nonlinear dynamical indicator across multi-sensor signals from rotating machinery [10], hypergraph-based sensor fusion for fault diagnosis [38], and adaptive multivariate feature-mode decomposition utilizing multi-attention residual fusion networks [39]. Additionally, semi-supervised, temperature-guided fusion frameworks are designed for machinery with limited labeled data [13]. These techniques span various applications, including turning processes [14], additive manufacturing [15], health assessment of hydraulic components [16], and handling complex electromechanical equipment involving replacement-part effects [17]. Fusion is generally implemented through early concatenation, intermediate parallel-subnetwork merging, or late decision combination, and reported results consistently favor fusion over single-modality baselines across this body of work [4, 7, 8, 22, 37, 40].

In the context of robotic manipulators, a review of condition monitoring and fault diagnosis approaches for industrial robots [18] surveys various methods, including vibration analysis, current signature analysis, and acoustic emission techniques. These methods are applied to joints, gearboxes, and reducers. Additionally, more focused studies examine vibration-based diagnostics to assess the integrity of base mounting in industrial robots, as well as the use of autoencoders for anomaly detection in a three-degree-of-freedom (3-DOF) delta robot [41]. In the context of parallel robots, fault diagnosis has been explored using wavelet scattering networks and compared with time-domain features across various methods such as wide neural networks, linear SVM, logistic regression, and kernel naive Bayes classifiers [19]. Additionally, a hybrid PCA/LDA/meta-learning feature-extraction pipeline has been implemented within an active fault-tolerant control loop [20]. However, the existing literature rarely investigates the sensing modality as an independent variable. Furthermore, it infrequently reports on operating-condition classification tasks, as opposed to only addressing fault/no-fault or fault-type classification tasks [21].

### 2.2 Ablation Studies and Evaluation Under Operating Condition Shifts

In the multimodal fusion literature, when researchers report single-modality baselines, they usually aim to show that combining modalities is more effective than using each one alone. However, they do not reveal each channel's independent contribution. This distinction is important for making informed decisions about instrumentation. Simply proving that fusion outperforms individual modalities does not clarify whether the weaker channel is redundant, marginal, or essential, nor does it indicate if that ranking would hold under different operational conditions. This issue is not unique to condition monitoring. In general machine-learning evaluation, it is well established that comparing classifiers requires controlled, repeated evaluations with documented statistical tests [42]. Relying solely on a single point estimate is not sufficient for reliable comparisons [43].

In addition, these 0studies are primarily evaluated based on data from in-distribution conditions. However, a significant and rapidly growing body of research on domain generalization for fault diagnosis shows that performance under unseen operating conditions can differ markedly from in-distribution performance [23]. Several algorithmic approaches address this issue. These techniques encompass domain augmentation networks for real-time diagnosis under unfamiliar conditions [24], meta-learning-based domain generalization methods featuring gradient alignment and semantic matching, adaptation to varying situations with limited data [25], and adversarial domain-augmented generalization [26]. Other methods involve feature-decoupling domain generalization networks [27], relationship-transfer domain generalization [28], and approaches like multimodal cross-domain mixed fusion with dual disentanglement [29]. Additionally, strategies for test-time adaptation in evolving operating conditions [31] and open-set domain generalization address both distribution and label shift [30].

### 2.3 Instrumentation cost as an evaluation axis

Low-cost and non-intrusive instrumentation has been pursued as an explicit design goal, particularly for small and medium-sized manufacturers and for retrofit onto active equipment. An open-hardware low-cost condition-monitoring platform combining a MEMS ultrasonic microphone with a MEMS accelerometer demonstrates that a consumer-grade acoustic channel can approach research-grade accelerometer bandwidth at a fraction of the cost [32]. Non-intrusive, low-cost data-acquisition hardware has similarly been developed for machining-process health monitoring using acoustic emission and accelerometers mounted in the work holding rather than on the tool itself [44]. In both cases, cost is reported as a property of the proposed system rather than as an axis of comparison.

Sensor selection and placement have been formally addressed as a multi-objective optimization problem in structural health monitoring and systems engineering. Research has explored various approaches, including Pareto-optimal sensor placements that balance cost and diagnostic information for offshore structures [35], non-dominated sorting genetic algorithms for multi-objective sensor placement in grid structures [36], and greedy algorithms that mix inexpensive and expensive sensors within cost constraints [45]. Additionally, nondominated-solution greedy sensor selection has been utilized for optimal experimental design [46], and a general review of sensor optimization objectives, including diagnostic performance, robustness, and cost-effectiveness, has been conducted [47]. Prior work has also investigated visualization techniques for Pareto-frontier analysis in laboratory instrumentation network design and upgrade decisions [48].

### 2.4 Classifier families and explainability

There are several well-established categories of classifiers that have been employed for the classification of machine and robotic faults, specifically using vibration and acoustic data. Among these categories, bagged tree ensembles, commonly implemented through random forests, serve as robust models that enhance predictive accuracy by aggregating multiple decision trees [49]. Another prevalent approach is the use of boosted tree ensembles, which adhere to the gradient-boosting framework, allowing for refined models that correct errors of previous trees iteratively [50]. Additionally, the support vector machine (SVM) framework leverages kernel classifiers, which can efficiently handle both linear and non-linear classification problems by transforming input features into higher-dimensional spaces [51]. Furthermore, the nearest-neighbor method, a widely recognized instance-based classification technique, compares new data points to those in the training set and classifies them based on proximity [52].

Moreover, recent advances in model-agnostic explainability have significantly improved the interpretability of machine learning models in condition monitoring, notably through Shapley Additive Explanations (SHAP) [53]. This methodology has been increasingly implemented in various studies focused on feature attribution in condition monitoring. For instance, a study on bearing fault diagnosis utilizing SHAP revealed

that specific vibration statistics, such as skewness and shape factor, play a crucial role in influencing model predictions [54]. Additionally, researchers have effectively combined LIME (Local Interpretable Model-agnostic Explanations) and SHAP for fault diagnosis in audio sensor data, providing insights across multiple classifiers, including gradient boosting, SVM, naive Bayes, and random forests [55]. SHAP has also been instrumental in severity classification for wind turbine gearboxes, explicitly linking feature attribution results to actionable recommendations on sensor placement [56]. This integrative approach underscores the importance of model explainability in enhancing both diagnostic accuracy and practical implementation in fault monitoring systems.

### 2.5 Research gap summary

The distinct literature categories reviewed is advance directly on this problem. Multimodal sensor fusion for condition monitoring, domain generalization under operating-condition shift, and cost-aware or Pareto-based sensor selection, and each is well developed on its own terms (Sections 2.1 - 2.4). Their intersection is rarely discussed. In this domain of literature, no work identified varies sensing modality as the independent variable against a fixed classifier and jointly evaluates in-distribution accuracy, interventional out-of-distribution risk, and instrumentation cost on a robotic manipulator platform. The cost-aware sensor-selection literature illustrates this separation. Pareto-based formulations of sensor placement and selection are well established. Still, each instance evaluates cost against in-distribution diagnostic performance alone, never against performance under the covariate shift the sensor will encounter post-deployment. A configuration judged cost-effective by that literature's own criteria could therefore be cost-ineffective under the exact operating-condition variation that domain-generalization literature shows is common in practice.

This gap has both a methodological and topical dimension. Studies on condition monitoring and fault diagnosis often do not utilize a factorial or quasi-factorial design for operating conditions. Such designs are essential for supporting both in-distribution evaluations and interventional out-of-distribution assessments within a single study. Additionally, these studies rarely report general-purpose statistical safeguards that are well established in applied statistics and machine learning. These safeguards include percentile bootstrap confidence intervals [57], balanced accuracy that accounts for class or condition heterogeneity, and leakage-safe hyperparameter tuning, which prevents selecting a classifier based on the same data used for its final performance evaluation [58]. These are not new methods. They are standard practice that is inconsistently carried into this domain, and their absence makes it difficult to evaluate claims that one sensing configuration or classifier "outperforming" another in the existing literature. A further consequence is theoretical rather than statistical. Because single-modality baselines in the fusion literature are typically reported only to demonstrate that fusion helps, no prior work distinguishes an observed single-modality shortfall that reflects a genuine information gap from one that is a finite-sample estimation artifact of fitting a higher-dimensional fused hypothesis class from limited data.

## 3. Methodology

We treat sensing modality as an experimental variable: the same fixed classifier is evaluated on each feature subset, so that any accuracy difference is attributable to the sensing channel rather than to the model. We define modality configurations and prove a distribution-free nesting result showing that fusion cannot increase the population risk attainable by any single-modality configuration it contains (Section 3.2). In-distribution and out-of-distribution risk under covariate shift are formalized as two distinct, complementary quantities estimable from a single factorial dataset (Sections 3.4 - 3.5), and combined with shift-robustness and cost-efficiency metrics, retained capacity, risk spread, and a formal Pareto-dominance criterion that connect risk to instrumentation decisions (Section 3.6). Every reported statistic carries a bootstrap confidence interval (Section 3.7). Section 3.8 combines the full procedure into a general workflow (Figure 1), applicable to any multi-sensor operating-condition dataset with a factorial or quasi-factorial covariate design.

### 3.1 Problem setting and notation

Let a physical process generate, for each trial $i = 1,\ldots,N$, a set of raw sensor signals $s_i$ drawn from $N_S$ synchronized channels, an operating-regime label $y_i$ belonging to a finite label set $Y$, and a vector of operating covariates $c_i$ that describe the conditions under which trial $i$ was executed (e.g. applied load, feed rate, trajectory type). The dataset is

$$D = \{ (s_i, y_i, c_i) \}, i = 1,\ldots,N \quad (1)$$

and is assumed to follow a factorial or quasi-factorial design over the covariates: the set of realized covariate levels $C = \{ c(1),\ldots,c(L) \}$ is discrete, and each level is populated by multiple trials, so that both within-level resampling (Section 3.4) and across-level transfer (Section 3.5) are estimable from the same dataset without additional data collection.

A feature-extraction map $\varphi$ reduces each raw channel to a fixed-length descriptor, and concatenating across channels gives the full feature vector for trial $i$:

$$x_i = \varphi(s_i) \in \mathbb{R}^d, \quad x_i = [\, x_i^{(1)}; x_i^{(2)}; \ldots; x_i^{(M)} \,] \quad (2)$$

where the $d$ features are partitioned, by sensor origin, into $M$ non-overlapping modality blocks $x_i^{(1)}, \ldots, x_i^{(M)}$ (e.g. an acoustic block, vibration block, etc.). The $\varphi$ itself is left general, any per-channel statistic that is well-defined on a raw signal segment of finite length may be used, including but not limited to the amplitude-domain statistics in Table 3.1.

### 3.2 Modality configurations and the fused hypothesis class

A modality configuration $m$ selects a subset of the $M$ blocks. Suppose that $M$ for the set of configurations under study, typically each block individually combined their concatenation, e.g. $M = \{ 1, 2, \ldots, M, fused \}$. Let $\sigma m(x)$ denote the restriction of a feature vector $x$ to the blocks used by configuration $m$, so that $\sigma m(x) \in \mathbb{R}^{d_m}$ with $d_m \leq d$ and, for the fused configuration, $d\_fused = d\,(the\ concatenation\ of\ every\ block, x_i^{(1)} \oplus x_i^{(2)} \oplus \ldots \oplus x_i^{(M)})$.

For each configuration $m$, a classifier $h_\theta{}^{(m)}$: $X^m \rightarrow Y$ is drawn from a single fixed hypothesis-class family $H_\theta$, parameterized by hyperparameters $\theta$ that are held identical across every $m$ under comparison. Fixing $\theta$ across configurations is the central design constraint of this methodology: it ensures that any measured difference in risk between two configurations $m$ and $m'$ is attributable to the feature blocks $\sigma m(x)$ versus $\sigma m'(x)$, and not to unequal tuning effort or model capacity favoring one configuration over another.

This design constraint has an immediate, distribution-free consequence for the fused configuration whenever the hypothesis class $H_\theta$ is closed under coordinate projection (true of any classifier family that can learn to assign zero weight or importance to a subset of its inputs, which includes linear models with $L_1$ / $L_2$ shrinkage, tree ensembles, and any universal function approximator). If $h_m^*$ denotes a population-risk minimizer for configuration $m$, then because every single-block configuration's input space is a coordinate subspace of the fused input space,

$$R\left(h_{fused}^*\right) \leq R(h_m^*) \quad (3)$$

for every single-block configuration $m$ , where $R(h) = \mathbb{E}_{(x,y)\sim P}[\,L(h(x), y)\,]$ is the population risk under loss function $L$. Equation (3) states that adding a modality to the input space cannot increase the population risk attainable by the best classifier in $H_\theta$: the fused hypothesis class nests every single-modality class as a special case (set the added block's weights to zero), so its risk infimum is at least as low. A practical implication follows directly: an observed empirical loss for a fused configuration relative to a single-modality configuration is a finite-sample estimation effect (the extra parameters needed to exploit the added block are harder to fit precisely from $N$ trials) rather than evidence that the added modality carries no information, the two explanations make different, testable predictions as $N$ grows, and should not be conflated when interpreting a fusion result on a fixed, finite dataset.

### 3.3 Feature descriptors

Table 3.1 lists the amplitude-domain descriptors used to instantiate $\varphi$ in Section 4; any subset, or additional descriptors (e.g. spectral centroid, kurtosis, or a learned embedding), may be substituted without changing the formulation in Sections 3.1-3.2 and 3.4-3.6, provided $\varphi$ is applied identically to every channel and every modality configuration.

**Table 3.1. Representative per-channel amplitude-domain feature descriptors.**

| Descriptor | Definition (signal segment s[1..T]) | Captures |
|---|---|---|
| Root-mean-square (RMS) | $RMS(s) = \sqrt{\frac{1}{T}\sum_{t=1}^{T} s[t]^2}$ | signal energy |
| Peak-to-peak (P2P) | $P2P(s) = \max_t s[t] - \min_t s[t]$ | amplitude range |
| Crest factor (CF) | $CF(s) = \frac{\max_t \lvert s[t]\rvert}{RMS(s)}$ | impulsiveness / peaked ness |

### 3.4 In-distribution risk estimation

For a configuration $m$, the in-distribution ($IID$) risk is estimated by repeated stratified $K$-fold cross-validation over the full dataset $D$. Partitioning $D$ into $K$ stratified folds $F1, \ldots, FK$, training on the complement of fold $k$ and evaluating on fold $k$, then pooling predictions across all $K$ folds and $R$ repeats with independent fold assignments gives

$$\hat{R}_{IID}\left(h^{(m)}\right) = \frac{1}{K\cdot R}\sum_{r=1}^{R}\sum_{k=1}^{K} L(h_{-k,r}^{(m)}, F_{k,r}) \tag{4}$$

where $h_{-k,r}^{(m)}$ is a configuration-m classifier trained on all data outside fold $k$ of repeat $r$, feature-standardized using statistics computed on that training partition only. Repetition ($R > 1$) with independent fold assignments, rather than a single K-fold pass, is used so that a variance estimate over the pooled risk is available for the confidence interval in Section 3.6, rather than reporting a single split's point estimate as if it had no sampling variability.

### 3.5 Out-of-distribution risk estimation under covariate shift

Where the operating covariate $c$ is set by intervention rather than merely observed (as is typical of a controlled factorial design), the dataset supports a second, distinct evaluation: transfer to a covariate level withheld entirely from training, rather than resampling within the pooled covariate distribution. Let $P$ and

$Q$ denote two covariate levels in $C$. Training on all trials with $c_i = P$ and testing exclusively on trials with $c_i = Q$ gives a directional out-of-distribution (OOD) risk

$$\hat{R}_{P\to Q}\left(h^{(m)}\right) = L\left(h_p^{(m)}, D_Q\right), D_Q = \{(x_i, y_i): c_i = Q\} \quad (5)$$

where $h_p^{(m)}$ is trained only on trials at level $P$. Evaluating both directions ($P \to Q \; and \; Q \to P$) and pooling the held-out predictions before scoring gives a single pooled OOD risk per configuration,

$$\hat{R}_{OOD}\left(h^{(m)}\right) = L\left(\left\{h_p^{(m)} on\ D_Q\right\} \cup \left\{h_Q^{(m)} on\ D_P\right\}\right) \quad (6)$$

For designs with more than two covariate levels, Equations (5)-(6) generalize directly to a leave-one-level-out scheme: train on $C \setminus \{ Q \}$, test on level $Q$, for every $Q \in C$, and pool all held-out predictions for $\hat{R}_{OOD}$. The two-level case in Equations (5)-(6) is the minimal instance of this general scheme and is written out separately because it additionally admits the directional asymmetry statistic in Section 3.6.

**3.6 Shift-robustness and cost-efficiency metrics**

Three derived statistics connect the risk estimates of Sections 3.4-3.5 to deployment-relevant questions about robustness and instrumentation cost. Retained capacity expresses what fraction of in-distribution performance survives covariate shift:

$$\rho(m) = \hat{R}_{OOD}(h^{(m)})/\hat{R}_{IID}\left(h^{(m)}\right) \quad (7)$$

Risk spread quantifies whether OOD degradation is symmetric with respect to the direction of transfer, for the two-level case of Equation (5):

$$\Delta E(m) = \left|\hat{R}_{P\to Q}\left(h^{(m)}\right) - \hat{R}_{Q\to P}(h^{(m)})\right| \quad (8)$$

a small $\Delta E$ indicates that a configuration's OOD risk does not depend materially on which covariate level is withheld, while a large $\Delta E$ indicates that the pooled estimate $\hat{R}_{OOD}$ in Equation (6) is summarizing two substantially different regimes and should be reported alongside its directional components, not in place of them.

Instrumentation cost is modeled additively over the sensors used by a configuration. Let $S_m$ be the set of physical sensors required by configuration $m$ and $\kappa_s$ the unit cost of sensor s; total cost is

$$\kappa(m) = \sum_{s\in S_m} \kappa_s \quad (9)$$

and a scalarized accuracy-per-cost yield, the Cost-Normalized Diagnostic Yield (CNDY), useful for a single-number ranking when a full Pareto comparison (below) is not required, is

$$Y(m) = R_{OOD}(h^{(m)})/\kappa(m) \quad (10)$$

CNDY expresses out-of-distribution diagnostic accuracy per unit of instrumentation cost, so that two configurations differing in both accuracy and price can be compared on a single scale. It is reported alongside, never in place of, the Pareto-dominance analysis of Equation (11), since a scalar ratio can rank

two non-dominated configurations in either order depending on which is more expensive, information the Pareto frontier alone preserves and CNDY collapses.

Because a scalarized yield collapses two objectives into one and is therefore sensitive to the (somewhat arbitrary) choice to divide rather than, e.g., subtract a weighted cost term, the primary cost-accuracy comparison across configurations should be a Pareto-optimality analysis: configuration $m$ is Pareto-dominated if there exists an alternative configuration $m'$ that is no more expensive and no less accurate,

$$m\ is\ dominated \iff \exists m' \in \mathcal{M}: \kappa(m') \leq \kappa(m)\ and\ \hat{R}_{OOD}\left(h^{(m')}\right) \geq \hat{R}_{OOD}\left(h^{(m)}\right) \text{ (with at least one strict)} \quad (11)$$

and the non-dominated set (the Pareto frontier) is reported directly, with $Y(m)$ from Equation (10) used only as a secondary, ranking-convenience statistic once the frontier has been established.

### 3.7 Uncertainty quantification

Every point estimate above ( $\hat{R}_{IID}, \hat{R}_{OOD}$, and any statistic derived from them) is accompanied by a 95% confidence interval obtained by the percentile bootstrap. For a pooled set of $B$ held-out (prediction, label) pairs $\{(\hat{y}_j, y_j)\}_{\{j=1\}^{\{B\}}}$ and a chosen risk/accuracy functional $g$, the interval is constructed by resampling the B pairs with replacement $N_{boot}$ times,

$$g_b^* = g\left(\left\{\left(\hat{y}_j, y_j\right): j \in resample_b\right\}\right), b = 1, \ldots, N_{boot} \quad (12)$$

$$CI_{95\%} = [Q_{0.025}(\{g_b^*\}), Q_{0.975}(\{g_b^*\})] \quad (13)$$

where $Q_p$ denotes the $q^{th}$empirical quantile of the bootstrap distribution. This resampling is performed independently for every configuration m and every risk estimate ($IID$, each directional $OOD$, and pooled $OOD$), so that overlapping or non-overlapping intervals between configurations can be read directly as evidence for or against a statistically reliable difference, without requiring a distributional assumption on the underlying per-trial loss.

### 3.8 General workflow

Figure 1 summarizes Sections 3.1–3.7 as a single procedure, independent of any particular dataset, sensor pair, or operating covariate. Raw multi-sensor trials are reduced to per-channel features and partitioned into modality blocks, each modality - including the fused concatenation of all blocks, is passed to one fixed classifier family sharing identical hyperparameters, and every configuration is then scored under two parallel risks: in-distribution accuracy via repeated cross-validation, and out-of-distribution accuracy via transfer to a covariate level withheld entirely from training. These two risk estimates are combined into the retained-capacity and risk-spread statistics that summarize shift robustness. An instrumentation-cost model, computed independently of any accuracy result, joins the evaluation only at the final stage, shown in the figure as the dashed connection into the cost-accuracy Pareto analysis, where it is combined with out-of-distribution risk to report the non-dominated configurations and a scalarized yield. Every arrow in Figure 1 corresponds to an equation in Sections 3.1–3.7 rather than to an implementation-specific step, so the workflow transfers to any multi-modal sensing study with a factorial or quasi-factorial operating-condition design.

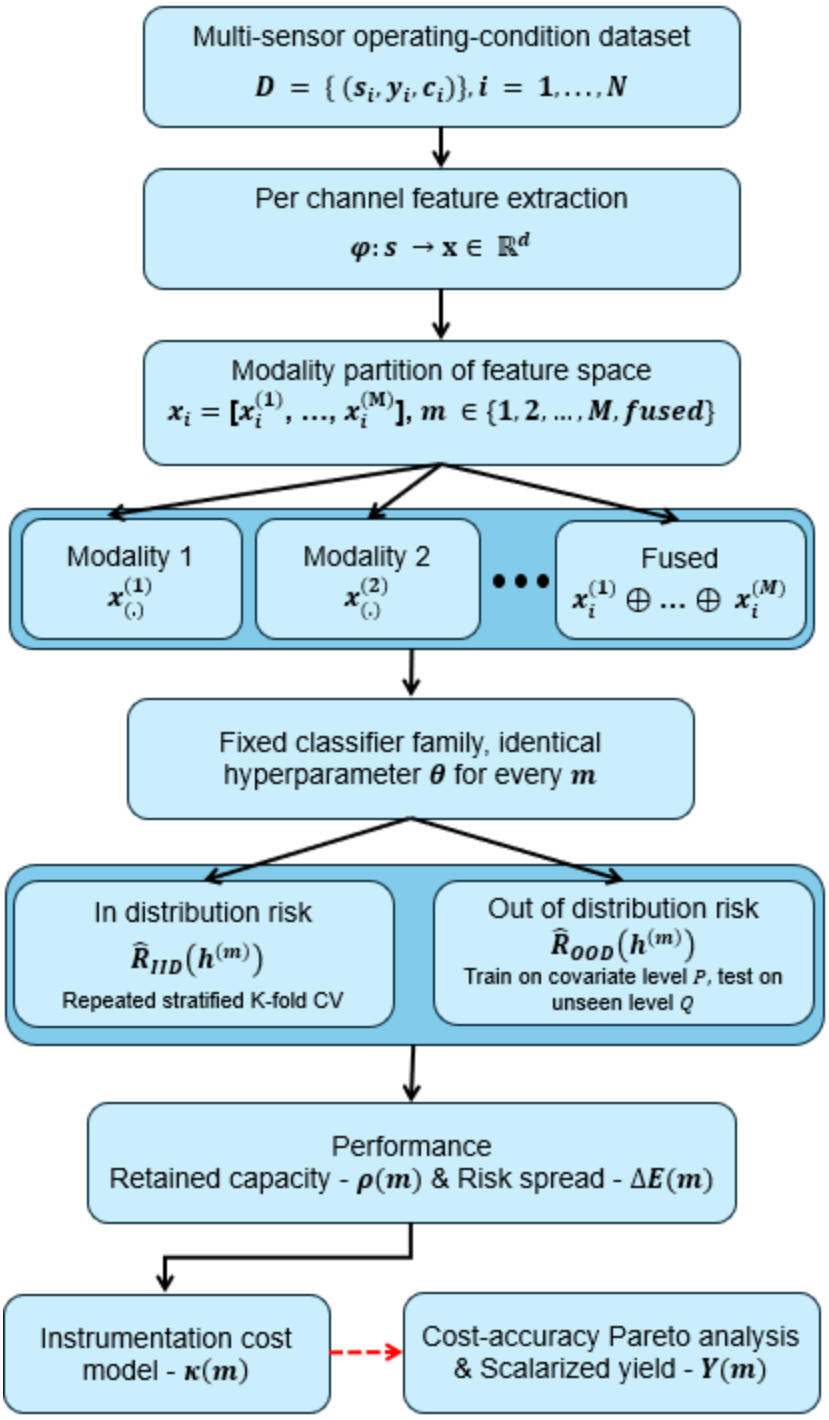


Figure 1. General workflow for cost-aware, shift-robust modality ablation

Two aspects of this structure are worth making explicit rather than leaving implicit in the diagram. First, every modality branch, including the fused one, passes through the same classifier box with the same hyperparameters; the figure has no branch where one configuration receives extra tuning or a different

validation scheme, which is what makes the accuracy differences reported in Section 4 attributable to the sensing channel rather than to unequal treatment of the classifier. Second, the cost model is drawn as a separate box reaching the Pareto analysis by its own path, distinct from the solid risk-estimation chain above it, cost is computed without reference to accuracy and only meets it at the very last stage, so a cheap configuration is never given an easier accuracy bar to clear on the way there

## 4. Experimental Setup and Results

This section validates the methodology of Section 3 on a delta-configuration parallel robot performing repeated operations under controlled variation in speed, trajectory, and payload. Section 4.1 describes the platform, the sensing modalities, and the resulting dataset. Section 4.2 reports in-distribution and out-of-distribution performance for each modality configuration, the resulting cost-accuracy Pareto comparison, feature-level attribution, and a cross-classifier check of these findings. Throughout, the same evaluation protocol defined in Section 3 is applied without modification, so that every result below is a direct instantiation of Equations (1)–(13) rather than a platform-specific variant of them.

### 4.1. Platform and data collection

Data were collected from a delta-configuration parallel robot instrumented with two co-located sensing modalities: a vibration measurement chain (two accelerometers, V1 and V2) and a pair of USB condenser microphones (A1 and A2). The two chains differ substantially in unit cost (approximately $50.00 per vibration sensor versus $27.00 per microphone) and were mounted to observe the same physical events from complementary transduction principles, allowing sensing modality to be varied as an independent experimental factor while holding the mechanical event itself fixed. Figure 2 shows the experimental platform used in this experiment.

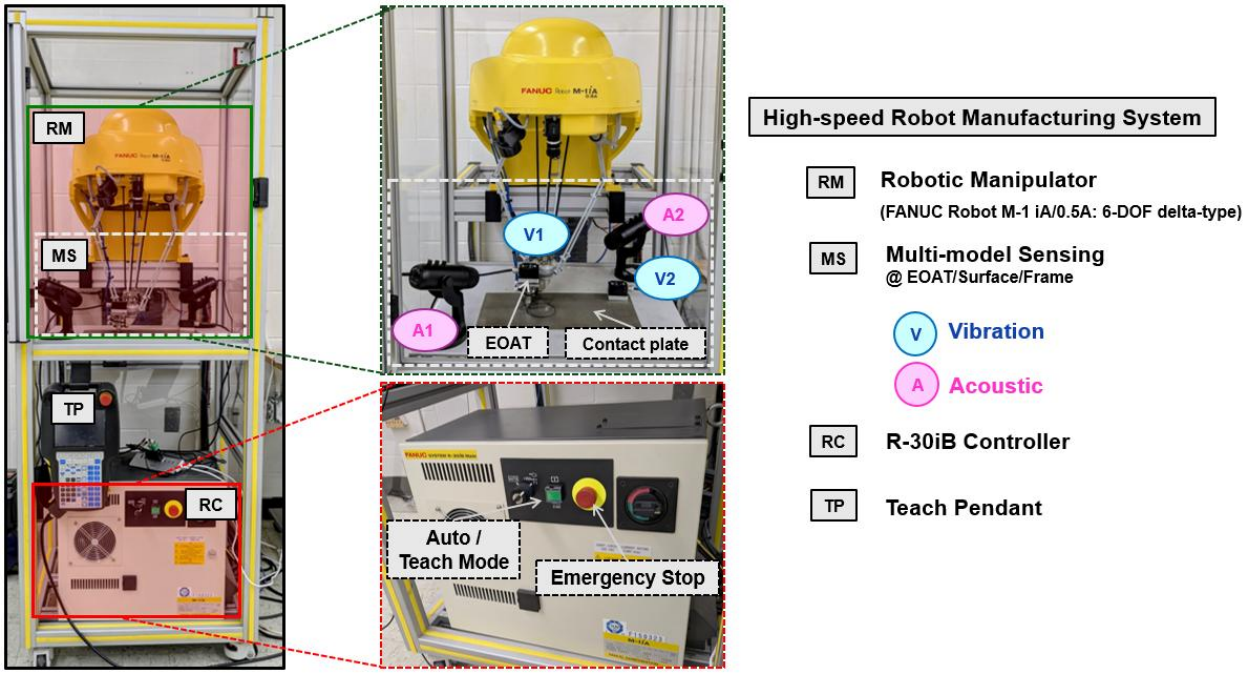


Figure 2: Experimental setup on the FANUC M-1iA/0.5A delta-robot platform, showing placement of acoustic sensors A1/A2 and tri-axial accelerometers V1/V2

The dataset spans a fully factorial operating-condition design of three end-effector speeds (50, 150, and 300 mm/s, labelled Low/Medium/High), three trajectory profiles (Linear, Circular, and a mixed Linear+Circular path), and two payload levels (108.71 g and 179.93 g), each replicated 20 times. Payload was set by intervention rather than observed, which is what makes it usable as a held-out generalization test in Section 4.3.2 rather than merely a resampling split.

### 4.2 Feature extraction and modality partitions

For every run, three standard time-domain features, root-mean-square (RMS), peak-to-peak amplitude (P2P), and crest factor (CF) were computed per channel, yielding 6 vibration features (V1/V2 × RMS/P2P/CF) and 6 acoustic features (A1/A2 × RMS/P2P/CF), 12 features total. These are partitioned into four non-overlapping evaluation configurations: **Vibration** (both accelerometers, 6 features), **Acoustic** (both microphones, 6 features), **Fused** (all 12 features concatenated), and **MinimalCross (V1+A1)**, one accelerometer paired with one microphone (6 features), the cheapest possible configuration that still combines both sensing modalities. No feature-level fusion beyond concatenation is applied, consistent with the manuscript's stated scope of an evaluative modality ablation rather than a new fusion architecture.

### 4.3 Evaluation protocol

The experimental design fixes the classifier and its hyperparameters across all four modality configurations (e.g. a Random Forest, 300 trees, unrestricted depth, identical random seed and preprocessing pipeline for every configuration) so that any measured performance difference is attributable to the sensing channel rather than to unequal modelling effort. Features were standardized (zero mean, unit variance) with the scaler fit on the training partition only in every evaluation, to avoid information leakage from test to train. Two complementary risks are reported for every modality.

#### 4.3.1 In-distribution (IID) risk

Repeated stratified 5-fold cross-validation (10 repeats, 50 train/test splits total) over the full dataset, predicting the three-level speed conditions (Low/Medium/High). Predictions from all folds and repeats are pooled before scoring, and a 95% confidence interval is obtained by percentile bootstrap (2,000 resamples) over the pooled (true, predicted) pairs.

#### 4.3.2 Out-of-distribution (OOD) risk under payload shift

A leave-one-payload-out protocol in which the model is trained on all runs at one payload level and tested exclusively on runs at the other, in both directions (train on 108.71 g / test on 179.93 g, and the reverse). Held-out predictions from both directions are pooled for one bootstrap-CI balanced accuracy per modality (Pooled OOD), alongside the two directional accuracies individually.

Balanced accuracy (the mean of per-class recall) is used as the primary metric throughout, since it is directly comparable across the IID and OOD evaluations regardless of any future class-size adjustment. Macro-averaged F1, precision, and recall are reported alongside balanced accuracy for the IID condition.

#### 4.3.3 Cost accounting and efficiency metrics

Total cost per configuration is computed from unit sensor cost × sensor count: \$100.00 for Vibration (2 × \$50), \$54.00 for Acoustic (2 × \$27), \$154.00 for Fused (all four sensors), and \$77.00 for MinimalCross (one accelerometer + one microphone). Three derived efficiency metrics connect accuracy to this cost basis: Retained Capacity, the ratio of Pooled OOD to IID balanced accuracy (as a percentage) and Risk Spread (ΔE), the absolute difference between the two directional OOD accuracies, and the Cost-Normalized

Diagnostic Yield (CNDY), Pooled OOD balanced accuracy divided by total instrumentation cost (Equation 10), a scalarized accuracy-per-dollar figure reported for ranking convenience alongside the primary cost-accuracy Pareto comparison.

### 4.4 Results

Results are organized to show the two risks defined in Section 3: in-distribution performance is reported first (Section 4.4.1), establishing each configuration's accuracy under the conditions it was trained on, followed by out-of-distribution performance under payload shift (Section 4.4.2) and its cost-adjusted interpretation (Section 4.4.3). Sections 4.4.4–4.4.6 then examine why these results occur, through feature-level attribution, a cross-classifier check, and the structure of the errors themselves. Section 4.4.7 summarizes the findings that carry forward to the Discussion.

#### 4.4.1 In-distribution performance

Table 1 summarizes in-distribution performance across all four configurations. Vibration achieves 98.47% balanced accuracy [95% CI: 98.05%, 98.85%], Acoustic achieves 89.97% [89.00%, 90.95%], Fused achieves 99.56% [99.31%, 99.75%], and MinimalCross achieves 97.06% [96.47%, 97.59%].

Table 1: In-distribution performance by sensing modality

| Modality | Features (N) | Total Cost | IID Bal. Acc. [95%CI] | Macro F1 | Macro Precision | Macro Recall |
|---|---|---|---|---|---|---|
| Vibration | 6 | $100.00 | 0.9847 [0.9805, 0.9885] | 0.9847 | 0.9847 | 0.9847 |
| Acoustic | 6 | $54.00 | 0.8997 [0.8900, 0.9095] | 0.9007 | 0.9026 | 0.8997 |
| Fused | 12 | $154.00 | 0.9956 [0.9931, 0.9975] | 0.9956 | 0.9956 | 0.9956 |
| MinimalCross (Vib_1+Mic_1) | 6 | $77.00 | 0.9706 [0.9647, 0.9759] | 0.9705 | 0.9705 | 0.9706 |

The gap between Vibration and Acoustic (8.5 points) is a reliable in-distribution difference. MinimalCross, a single sensor per channel, already reaches 97.06%, between Acoustic and Vibration and closer to the full four-sensor Fused result (99.56%) than to either full single-modality configuration alone, indicating that most of the in-distribution separability gained from fusion is available even from minimal cross-modal instrumentation.

#### 4.4.2 Out-of-distribution performance under payload shift

The model's performance evaluated on the payload shift, which was absent from the training, is shown in Table 2. All four configurations degrade sharply relative to their in-distribution figures: Pooled OOD balanced accuracy falls to 54.17% for Vibration, 54.72% for Acoustic, 63.89% for Fused, and 56.11% for MinimalCross, corresponding to Retained Capacity of 55.01%, 60.82%, 64.17%, and 57.81% respectively. Figure 3 shows the decline in accuracy on OOD data.

Table 2: Out-of-distribution shift and efficiency metrics (held-out payload).

| Modality | Held-out 108.71g | Held-out 179.93g | Pooled OOD [95% CI] | Retained Capacity | Risk Spread (ΔE) | Yield/USD |
|---|---|---|---|---|---|---|
| Vibration | 0.4611 | 0.6222 | 0.5417 [0.4916, 0.5923] | 55.01% | 0.1611 | 0.0054 |
| Acoustic | 0.5889 | 0.5056 | 0.5472 [0.4958, 0.5970] | 60.82% | 0.0833 | 0.0101 |
| Fused | 0.5556 | 0.7222 | 0.6389 [0.5904, 0.6878] | 64.17% | 0.1667 | 0.0041 |
| MinimalCross (V1+A1) | 0.5444 | 0.5778 | 0.5611 [0.5119, 0.6070] | 57.81% | 0.0333 | 0.0073 |

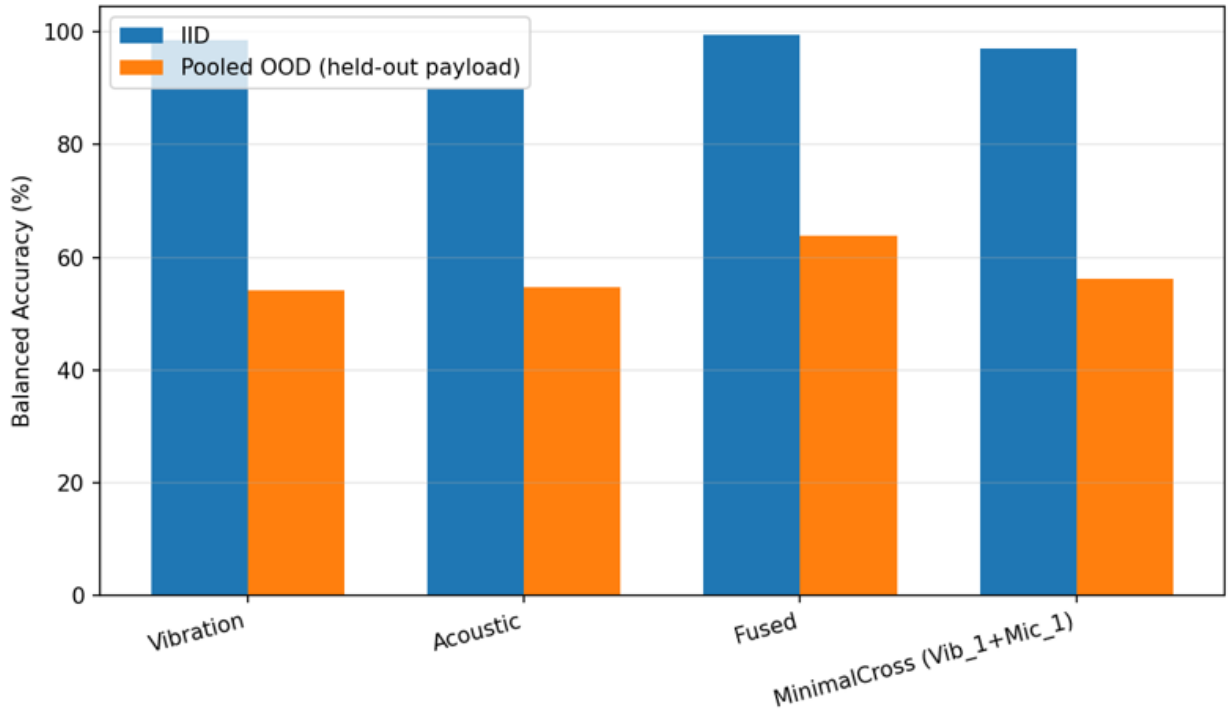


Figure 3: Balanced accuracy for different modalities: IID vs OOD (held-out payload)

Two findings stand out. First, the in-distribution ranking (Fused > Vibration > Acoustic) does not survive payload shift. The vibration's Pooled OOD balanced accuracy (0.5417 [0.4916, 0.5923]) and acoustic's (0.5472 [0.4958, 0.5970]) have almost completely overlapping confidence intervals, so the 8.5-point in-distribution gap between them is not reliable under shift. Second, MinimalCross, the cheapest cross-modal configuration, outperforms both full single-modality configurations under shift (0.5611 [0.5119, 0.6070] vs. 0.5417 for Vibration and 0.5472 for Acoustic), while costing less than vibration alone ($77 vs. $100). MinimalCross also has the lowest risk spread of any configuration (ΔE = 0.0333, versus 0.1611 for Vibration, 0.0833 for acoustic, and 0.1667 for fused), meaning its accuracy under shift is noticeably less sensitive to which payload direction is withheld (0.5444 held out at 108.71 g vs. 0.5778 held out at 179.93 g) than any other configuration tested.

### 4.4.3 Cost-adjusted comparison

The values presented in Table 2 indicate that vibration-only configurations are comparably improved by MinimalCross, which not only achieves lower costs but also higher OOD accuracy. Among the configurations, acoustic offers the highest yield, providing 0.0101 balanced accuracy points per dollar spent. MinimalCross follows with 0.0073, vibration comes next at 0.0054, and the fused configuration has the lowest yield at 0.0041. Figure 3 plot all four configurations on cost versus Pooled OOD balanced accuracy. Three configurations now occupy the Pareto frontier such as acoustic (lowest cost, $54), MinimalCross (intermediate cost, $77, and higher OOD accuracy than either full single-modality configuration), and fused (highest cost, $154, highest accuracy).

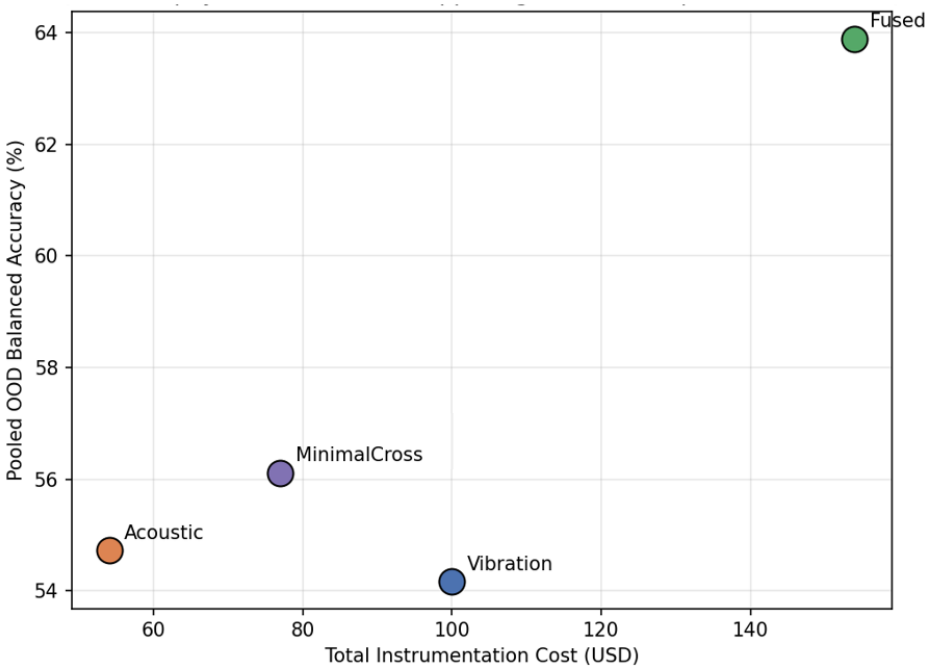


Figure 4: Cost versus Pooled OOD balanced accuracy

While the fully fused system delivers the best overall accuracy, it has the lowest accuracy per dollar among all tested configurations. Whether the additional cost of the fused or MinimalCross configurations compared to the acoustic-only option is justified depends on the accuracy requirements of the specific application.

### 4.4.4 Feature-level attribution

Table 3 reports Gini feature importances from the Random Forest fit on the full dataset. The top five features are V1_CF (0.1555), V2_RMS (0.1449), A2_P2P (0.1364), A1_P2P (0.1197), and V1_RMS (0.1149), together accounting for 63.0% of total importance mass. Notably, both features used by the MinimalCross configuration's vibration channel (V1_CF, rank 1; V1_RMS, rank 5) and its acoustic channel (A1_P2P, rank 4; A1_RMS, rank 7) rank among the top half of all twelve features, MinimalCross was not constructed with this ranking in mind, but its two constituent sensors happen to carry substantial fused-model importance, consistent with its strong showing in Table 2 relative to its low cost.

Table 3. Random Forest feature importances

| Feature | Sensor Type | Sub-Metric | Gini Importance | Rank |
|---|---|---|---|---|
| V1_CF | Vibration | Crest Factor | 0.1555 | 1 |
| V2_RMS | Vibration | Root Mean Square | 0.1449 | 2 |
| A2_P2P | Acoustic | Peak-to-Peak | 0.1364 | 3 |
| A1_P2P | Acoustic | Peak-to-Peak | 0.1197 | 4 |
| V1_RMS | Vibration | Root Mean Square | 0.1149 | 5 |
| V2_CF | Vibration | Crest Factor | 0.0650 | 6 |
| A1_RMS | Acoustic | Root Mean Square | 0.0616 | 7 |
| V2_P2P | Vibration | Peak-to-Peak | 0.0616 | 8 |
| V1_P2P | Vibration | Peak-to-Peak | 0.0463 | 9 |
| A2_RMS | Acoustic | Root Mean Square | 0.0451 | 10 |
| A2_CF | Acoustic | Crest Factor | 0.0317 | 11 |
| A1_CF | Acoustic | Crest Factor | 0.0172 | 12 |

#### 4.4.5 Cross-category classifier comparison

As previously discussed in Table 1-3, we used Random Forest as the classifier to avoid confounding the modality ablation with classifier choice. To check that the modality-level findings are not an artifact of that one classifier, eight classifiers spanning 6 algorithmic categories, such as Random Forest [59] and Extra Trees [60] (bagging ensembles, differing in split-selection randomization), Histogram Gradient Boosting (boosting ensemble) [61], an RBF-kernel Support Vector Machine (kernel method) [62], Logistic Regression (linear) [63], a shallow multilayer perceptron with two hidden layers of 64 and 32 units (neural network) [64], k-Nearest Neighbors with k=5 (instance-based) [65], and Gaussian Naive Bayes (probabilistic) [66] were evaluated on the fused feature set under the identical IID/OOD protocol as Tables 1–2 (Table 4).

Table 4. Classifier comparison, Fused modality, identical IID/OOD protocol as Tables 1–2.

| Classifier | IID Bal. Acc. [95% CI] | Pooled OOD Bal. Acc. [95% CI] |
|---|---|---|
| Logistic Regression (linear) | 0.9422 [0.9349, 0.9495] | **0.7278** [0.6885, 0.7665] |
| Extra Trees (bagging, extra-randomized) | 0.9972 [0.9954, 0.9989] | 0.7000 [0.6572, 0.7440] |
| MLP (shallow neural network) | 0.9922 [0.9893, 0.9950] | 0.6472 [0.6057, 0.6921] |
| Random Forest (bagging ensemble) | 0.9956 [0.9931, 0.9975] | 0.6389 [0.5904, 0.6878] |
| k-NN (instance-based) | 0.9808 [0.9764, 0.9851] | 0.5833 [0.5439, 0.6262] |
| SVM-RBF (kernel) | 0.9856 [0.9814, 0.9893] | 0.5833 [0.5436, 0.6257] |
| Hist Gradient Boosting (boosting ensemble) | 0.9914 [0.9881, 0.9944] | 0.5444 [0.4960, 0.5900] |
| Gaussian Naive Bayes (probabilistic) | 0.7344 [0.7210, 0.7478] | 0.4583 [0.4110, 0.5068] |

Every classifier except Gaussian Naive Bayes exceeds 94% in-distribution balanced accuracy, with Extra Trees highest at 99.72%, confirming that near-ceiling in-distribution separability is a property of the Fused feature set rather than of any one classifier. Under payload shift, however, Logistic Regression retains the highest pooled OOD balanced accuracy of any classifier tested, 0.7278 [0.6885, 0.7665]. Extra Trees is the closest competitor at 0.7000 [0.6572, 0.7440] - its interval overlaps Logistic Regression's, so the two are not statistically distinguishable at this sample size, but its point estimate still trails despite its additional split-level randomization. The shallow neural network generalizes worse still (0.6472), below even Random Forest (0.6389). SVM-RBF and k-NN post nearly identical pooled OOD balanced accuracy (0.5833 each), and Hist Gradient Boosting and Gaussian Naive Bayes generalize worst under shift (0.5444 and 0.4583 respectively). No classifier tested matches the shift-robustness of Logistic Regression, which reflects the same in-distribution/out-of-distribution divergence seen at the modality level in Section 4.4.2.

### 4.4.6 Error structure under shift

Figure 5 shows the confusion matrices from pooled held-out-payload directions, for all four sensing modalities under Random Forest.

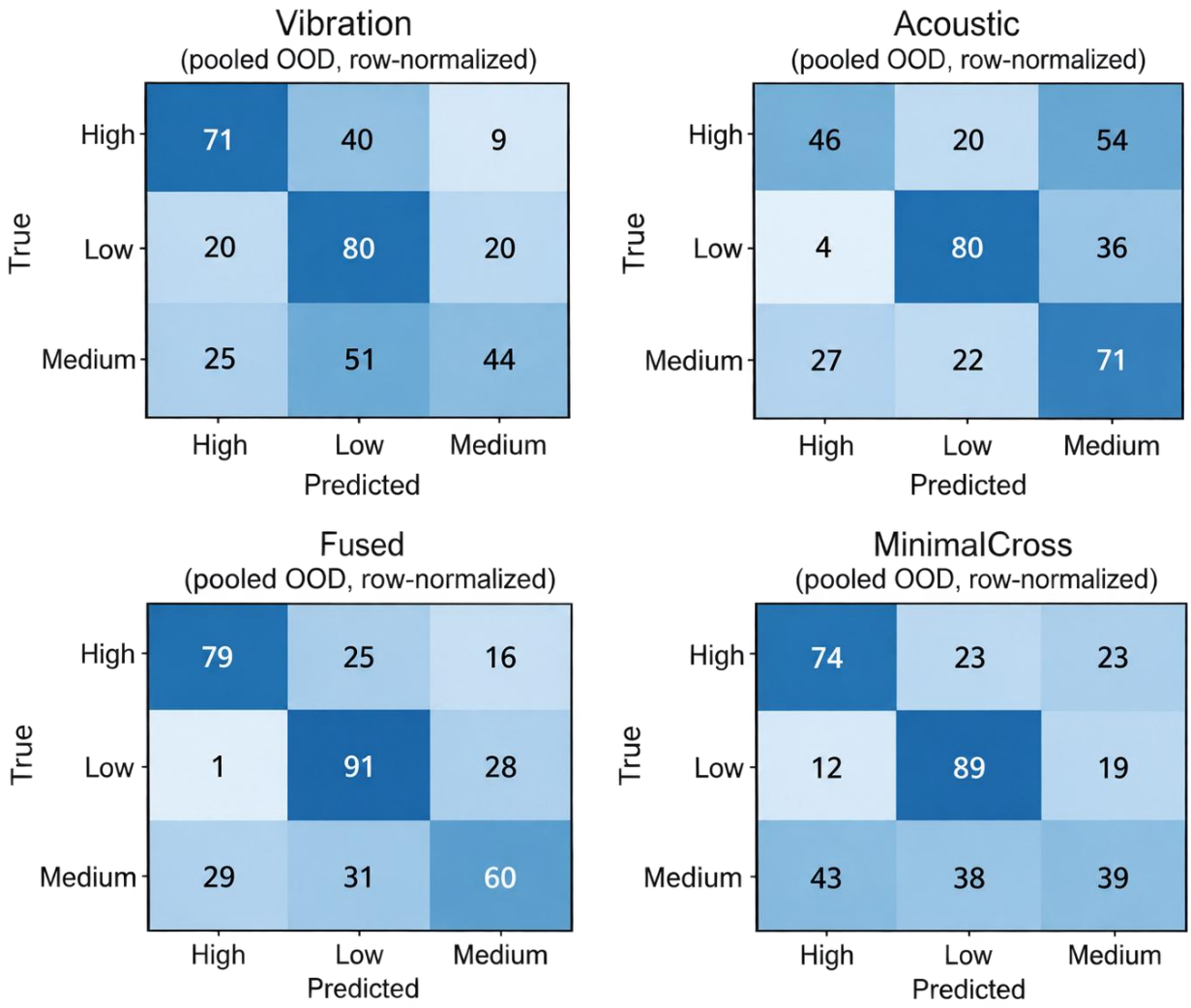


Figure 5: Confusion matrices from held-out-payload

Low-speed is recognized most reliably under shift across every configuration (80/120 for Vibration, 80/120 for Acoustic, 91/120 for Fused, 89/120 for MinimalCross), while Medium and High are confused with each other far more often than either is confused with Low - consistent with speed classes that are ordinally, not categorically, separated in the underlying feature space. MinimalCross's error pattern is notably more balanced across the Medium/High boundary (38 Medium on Low, 43 Medium on High, roughly symmetric) than Vibration's (51 Medium on Low vs. 25 Medium on High, a 2:1 skew), which is consistent with MinimalCross's noticeably lower Risk Spread in Table 2 - its errors are not just fewer, but more evenly distributed regardless of transfer direction.

**4.4.7 Summary of results**

Five findings carry forward to the Discussion (Section 5). First, the modality ranking that holds in-distribution (Fused > Vibration > Acoustic) does not hold under payload shift, where vibration and acoustic become statistically indistinguishable. Second, a minimal cross-modal configuration - one vibration sensor plus one acoustic sensor, MinimalCross - outperforms both full single-modality configurations under shift (56.11% vs. 54.17% and 54.72% pooled OOD balanced accuracy) at a cost between them ($77 vs. $100 and $54), and with substantially lower directional asymmetry ($\Delta E = 0.033$ vs. 0.161 and 0.083). Third, this

reshapes the cost-accuracy Pareto frontier from two points to three: Acoustic, MinimalCross, and Fused are all non-dominated, while vibration-only is dominated by MinimalCross. Fourth, fusion still delivers the best absolute OOD accuracy (63.89%) but the worst accuracy-per-dollar yield (0.0041) of any configuration tested, making it justifiable only where the accuracy floor requires it. Fifth, the same in-distribution/OOD divergence repeats at the classifier level: across eight classifiers spanning six algorithmic categories, Logistic Regression remains the most shift-robust despite not being the in-distribution leader, and this ranking is unchanged by two further high-capacity classifiers (Extra Trees, MLP) specifically chosen to test whether it would be overturned. Together these results indicate that evaluating sensing modality or classifier choice only in-distribution would have supported a materially different, and specifically a more expensive, instrumentation and modelling recommendation than the shift-aware, cost-aware evaluation performed here.

## 5. Discussion

We established four results: a reversal between vibration- and acoustic-only accuracy under payload shift, a minimal cross-modal configuration that outperforms both single-modality rigs at intermediate cost, a matching reversal at the classifier level, and a theoretical guarantee bounding fusion's population risk. Section 5.1 interprets what these results mean together, and what the theoretical result does and does not establish about them. Section 5.2 states the limitations that bound how far these conclusions should be generalized.

### 5.1 Interpretation of the modality and classifier reversals

Section 4.4.1 showed vibration-only reliably ahead of acoustic-only in-distribution (98.47% vs. 89.97%, non-overlapping confidence intervals). Section 4.4.2 showed that gap disappear under payload shift (54.17% vs. 54.72%, heavily overlapping intervals). These are not two independent facts, they are the same measurement under two different assumptions about the deployment environment, and they support opposite instrumentation recommendations. A practitioner relying only on the in-distribution test would install the vibration sensors, at $100, on the reasonable belief that it is the more capable sensor. The belief answers a different question ("which sensor works better under conditions already tested?") than the one that determines deployment cost-effectiveness ("which sensor works better under conditions not yet tested?"). The MinimalCross result sharpens this further: a single accelerometer paired with a single microphone ($77) outperforms both full single-modality rigs under shift (56.11% vs. 54.17% and 54.72%) with the lowest directional asymmetry of any configuration tested ($\Delta E = 0.033$) - not simply "cheaper is better," since acoustic-only is cheaper still and MinimalCross beats it too, but a genuine cross-modal effect obtained from minimal instrumentation.

The same pattern repeats at the classifier level. Extra Trees is the strongest in-distribution classifier (99.72%), but Logistic Regression is the most shift-robust (72.78% pooled OOD vs. Extra Trees' 70.00%). This is a second, structurally independent test of the same claim that the capacity that helps fit the training distribution does not reliably transfer to unseen data. Extra Trees serves as a meaningful test case because its additional split-threshold randomization acts as a regularizing mechanism that could prevent the reversal. It did not, though the gap to Logistic Regression is not statistically distinguishable at this sample size.

The nesting result (Equation 3) bears on these findings in a specific and limited way. The results clarify that Fused's population risk cannot exceed that of any single modality it includes. This rules out a common misunderstanding of Section 4.4.2. If a single-modality configuration ever outperforms Fused in empirical tests, the theorem indicates that this can only be due to a finite-sample estimation effect. It does not imply that the additional modality lacks information. Instead, it makes a falsifiable claim predicting that such performance gaps will decrease over time. However, the theorem does not address any cost-related questions since population risk is defined independently of cost. Fused's absence from the Pareto frontier

mentioned in Section 4.4.3 is a distinct empirical fact established through the cost-accuracy analysis, rather than being derived from Equation (3), and the two should not be confused. When taken together, these findings support a specific recommendation rather than a vague "it depends" statement. If the accuracy floor for an application is close to 55–58% for OOD balanced accuracy, MinimalCross is the recommended choice. It costs less than the vibration-only setup that a simplistic in-distribution evaluation might suggest, while still outperforming it in the actual shifts the application will face. If the accuracy floor genuinely requires Fused's 63.89%, that level of performance is only achievable at $154, yielding the lowest accuracy-per-dollar of any tested. This is justified by the required accuracy rather than by any claims of efficiency. The vibration-only configuration, which an in-distribution-only evaluation would recommend, is dominated.

### 5.2 Limitations

Several limitations bound how far these conclusions generalize. The out-of-distribution evaluation withholds one of exactly two payload levels at a time, so each directional OOD estimate is a single train/test partition rather than a resampled quantity, and risk spread is estimated from only two points. A design with three or more levels would allow genuine replication. Only payload was varied as the held-out covariate, speed and trajectory were pooled across training and test in every OOD evaluation, so whether the same reversals hold under speed or trajectory shift is not established here. The fixed-hyperparameter design in Section 3.2 deliberately isolates the sensing-channel effect from unequal tuning effort, but no classifier in Table 4 was tuned for this dataset, so the reversal should be read as holding under equal, untuned treatment rather than as a claim that no amount of tuning could close it. All results come from a single delta-configuration parallel robot with a specific sensor pairing and cost basis ($50/$27 per sensor). The methodology outlined in Section 3 is designed to analyze the effects of differences in sensor modalities under payload shifts. The numerical results and the specific Pareto frontier presented here represent one example, rather than a universal ranking of vibration versus acoustic sensing.

## 6. Conclusion and Future Work

This paper treated sensing modality as an experimental variable, evaluated jointly against in-distribution accuracy, out-of-distribution accuracy under payload shift, and instrumentation cost, on a delta-configuration parallel robot performing operating-condition classification. The central finding is a reversal. Vibration sensors that reliably outperforms a consumer-grade acoustic sensors in-distribution (98.47% vs. 89.97% balanced accuracy) becomes statistically indistinguishable from it under shift (54.17% vs. 54.72%), at roughly twice the cost, placing vibration-only outside the cost-accuracy Pareto frontier. A minimal cross-modal configuration, with one sensor per channel at an intermediate cost, outperforms both full single-modality rigs under shift, while also exhibiting the lowest directional asymmetry of any configuration tested. This refines the Pareto comparison rather than merely supplementing it. The same reversal appears at the classifier level. Across eight classifiers spanning six algorithmic categories, the strongest in-distribution classifier is not the most shift-robust one, and this held even against a classifier (Extra Trees) whose additional randomization made it, a priori, the most likely candidate to overturn the pattern. A distribution-free result for classifier families that are closed under coordinate projection shows that the population risk of a fusion method cannot exceed that of any individual modality it includes. This finding provides a theoretical baseline for empirical fusion results, rather than leaving them as an unexplained finite-sample observation. The results suggest that relying solely on in-distribution accuracy to evaluate sensing modalities and classifier choices, which is a common practice in the current literature, can lead to selecting more costly sensors and classifiers. In contrast, the shift-aware evaluation presented here supports a different and potentially more cost-effective choice.

This work can be extended in three key directions. First, the two-level payload design presented here should be expanded to include three or more levels. This will allow for leave-one-level-out replication of the out-

of-distribution estimates and provide a genuine variance estimate for the risk spread statistic, rather than just the single-point directional estimates discussed. Second, the shift dimension investigated in this study (payload) should be broadened to include speed and trajectory. This will help determine whether the modality and classifier reversals observed are a general feature of this platform under any operating condition shifts or if they are specific to load-induced distribution changes. Lastly, the MinimalCross result indicates that using a single sensor per channel captures much of the shift robustness advantage of full fusion at a fraction of the cost. This finding motivates a systematic search for the optimal single-sensor pairing, rather than relying on the arbitrary pairing (V1+A1) evaluated here. A comprehensive combinatorial ablation study across all sensor subsets, which was excluded from the current study for reasons of statistical power and focus, would directly address this issue. Conducting this research will be a natural next step once a larger replicate count provides sufficient statistical power for the comparison.

**Acknowledgements**

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

**Declarations**

**Author Contributions**
Md Omar Al Javed: Conceptualization, Methodology, Formal analysis, Investigation, Experimental setups, Data curation, Visualization, Writing - original draft. 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. Abdollahi-Mamoudan F, Ibarra-Castanedo C, Maldague XPV (2025) Non-Destructive Testing and Evaluation of Hybrid and Advanced Structures: A Comprehensive Review of Methods, Applications, and Emerging Trends. Sensors 25
2. Ghoreyshi M, Bergeron K, Seidel J (2024) High-Fidelity Simulations of Flight Dynamics and Trajectory of a Parachute–Payload System Leaving the C-17 Aircraft †. Aerospace 11:. https://doi.org/10.3390/aerospace11100827

3. Pană CF, Maican CA, Vrăjitoru NR, Pătrașcu-Pană DM, Rădulescu VM (2026) From Signals to Remaining Useful Life: Multimodal Sensor Fusion for Fault Diagnosis and Prognostics—Methods, Pitfalls, and Reporting Standards. Sensors 26
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. Liu Q, Senanayaka A, Lee N, Mun S, Amirlatifi A, Jabour J, Arnold T, Seale M (2026) Applications, challenges, recommendations, and future directions of additive manufacturing with transfer learning. Advanced Engineering Informatics 76:104962. https://doi.org/https://doi.org/10.1016/j.aei.2026.104962
6. 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
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. Kibrete F, Engida Woldemichael D, Shimels Gebremedhen H (2024) Multi-Sensor data fusion in intelligent fault diagnosis of rotating machines: A comprehensive review. Measurement 232:114658. https://doi.org/https://doi.org/10.1016/j.measurement.2024.114658
9. Lee N, Senanayaka A, Storey J, Murphy M, Hwang J, Lim H, Cai B, Netchaev A, Stone T (2025) Enhancing Defect Detection Accuracy In Wire-Arc Additive Manufactured Samples Through Fusion Of Multi-Directional Scanning With Phased Array Ultrasound Testing (Paut)
10. Xiao X, Li C, He H, Huang J, Yu T (2025) Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. Information Fusion 113:102621. https://doi.org/https://doi.org/10.1016/j.inffus.2024.102621
11. Zeng Y, Tang L, Xiao S, Dong Z, Li X, Wang T, Liu X, Ma Y, Chu F (2025) A Multi-Sensor Information Fusion Fault Diagnosis Method for Hydropower Turbine Bearings Based on Multi-Scale Spatiotemporal Graph Neural Networks. Engineered Science 38:. https://doi.org/10.30919/es1882
12. Yan X, Yan W-J, Xu Y, Yuen K-V (2023) Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network. Mech Syst Signal Process 202:110664. https://doi.org/https://doi.org/10.1016/j.ymssp.2023.110664
13. Jiang X, Li X, Wang Q, Song Q, Liu J, Zhu Z (2024) Multi-sensor data fusion-enabled semi-supervised optimal temperature-guided PCL framework for machinery fault diagnosis. Information Fusion 101:102005. https://doi.org/https://doi.org/10.1016/j.inffus.2023.102005
14. Zhang B, Shin YC (2018) A multimodal intelligent monitoring system for turning processes. J Manuf Process 35:547–558. https://doi.org/https://doi.org/10.1016/j.jmapro.2018.08.021
15. Petrich J, Snow Z, Corbin D, Reutzel EW (2021) Multi-modal sensor fusion with machine learning for data-driven process monitoring for additive manufacturing. Addit Manuf 48:102364. https://doi.org/https://doi.org/10.1016/j.addma.2021.102364
16. Li X, Zhang L, Tan T, Wang X, Zhao X, Xu Y (2025) Multi-Sensor Data Fusion and Vibro-Acoustic Feature Engineering for Health Monitoring and Remaining Useful Life Prediction of Hydraulic Valves. Sensors 25:. https://doi.org/10.3390/s25206294
17. YAO X, FENG Z, KONG X, ZHOU Z, LIU H, HU G (2025) Multi-source information fusion based fault diagnosis for complex electromechanical equipment considering replacement parts. Chinese Journal of Aeronautics 38:103420. https://doi.org/https://doi.org/10.1016/j.cja.2025.103420

18. Lei Y, Liu H, Li N, Cao J, Qiao Y, Wang H (2024) Condition monitoring and fault diagnosis of industrial robots: A review. Sci China Technol Sci 68:1110301. https://doi.org/10.1007/s11431-024-2810-2
19. Urrea C, Domínguez C (2024) Fault Diagnosis in a Four-Arm Delta Robot Based on Wavelet Scattering Networks and Artificial Intelligence Techniques. Technologies (Basel) 12:. https://doi.org/10.3390/technologies12110225
20. Domínguez C, Urrea C (2025) Intelligent Fault-Tolerant Control of Delta Robots: A Hybrid Optimization Approach for Enhanced Trajectory Tracking. Sensors 25:. https://doi.org/10.3390/s25061940
21. Al Javed MO, Senanayaka A (2026) Multichannel Acoustic Feature Fusion for Predictive Maintenance in High-Speed Robotic Manufacturing. In: Proceedings of the International Conference on Industrial Engineering and Operations Management. IEOM Society International, Michigan, USA
22. Al Mamun A, Bappy MM, Mudiyanselage AS, Li J, Jiang Z, Tian Z, Fuller S, Falls TC, Bian L, Tian W (2023) Multi-channel sensor fusion for real-time bearing fault diagnosis by frequency-domain multilinear principal component analysis. The International Journal of Advanced Manufacturing Technology 124:1321–1334. https://doi.org/10.1007/s00170-022-10525-4
23. Zhao C, Zio E, Shen W (2024) Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. Reliab Eng Syst Saf 245:109964. https://doi.org/https://doi.org/10.1016/j.ress.2024.109964
24. Shi Y, Deng A, Deng M, Xu M, Liu Y, Ding X, Bian W (2023) Domain augmentation generalization network for real-time fault diagnosis under unseen working conditions. Reliab Eng Syst Saf 235:109188. https://doi.org/https://doi.org/10.1016/j.ress.2023.109188
25. Ren L, Mo T, Cheng X (2024) Meta-learning Based Domain Generalization Framework for Fault Diagnosis With Gradient Aligning and Semantic Matching. IEEE Trans Industr Inform 20:754–764. https://doi.org/10.1109/TII.2023.3264111
26. Li Q, Chen L, Kong L, Wang D, Xia M, Shen C (2023) Cross-domain augmentation diagnosis: An adversarial domain-augmented generalization method for fault diagnosis under unseen working conditions. Reliab Eng Syst Saf 234:109171. https://doi.org/https://doi.org/10.1016/j.ress.2023.109171
27. Xiao Q, Yang M, Yan J, Shi W (2024) Feature decoupling integrated domain generalization network for bearing fault diagnosis under unknown operating conditions. Sci Rep 14:30848. https://doi.org/10.1038/s41598-024-81489-6
28. Qian Q, Zhou J, Qin Y (2023) Relationship Transfer Domain Generalization Network for Rotating Machinery Fault Diagnosis Under Different Working Conditions. IEEE Trans Industr Inform 19:9898–9908. https://doi.org/10.1109/TII.2022.3232842
29. Xia P, Huang Y, Qin C, Liu C (2026) Multi-modal cross-domain mixed fusion model with dual disentanglement for fault diagnosis under unseen working conditions. https://doi.org/10.1016/j.ymssp.2026.114693
30. Zhu J, Long S, Yuan Y (2026) Physics-Informed Graph Learning with Uncertainty Awareness for Open-Set Domain Generalization in Fault Diagnosis
31. Sun H, Fink O (2025) TARD: Test-time Domain Adaptation for Robust Fault Detection under Evolving Operating Conditions
32. Jakobsen MO (2024) Low cost MEMS accelerometer and microphone based condition monitoring sensor, with LoRa and Bluetooth Low Energy radio. HardwareX 18:e00525. https://doi.org/https://doi.org/10.1016/j.ohx.2024.e00525
33. Papananias M (2025) Bayesian monitoring of machining processes using non-intrusive sensing and on-machine comparator measurement. The International Journal of Advanced Manufacturing Technology 137:1929–1942. https://doi.org/10.1007/s00170-025-15174-x

34. Bagajewicz M, Cabrera E (2003) Pareto Optimal Solutions Visualization Techniques for Multiobjective Design and Upgrade of Instrumentation Networks. Ind Eng Chem Res 42:5195–5203. https://doi.org/10.1021/ie020865g
35. Mehrjoo A, Song M, Moaveni B, Papadimitriou C, Hines E (2022) Optimal sensor placement for parameter estimation and virtual sensing of strains on an offshore wind turbine considering sensor installation cost. Mech Syst Signal Process 169:108787. https://doi.org/https://doi.org/10.1016/j.ymssp.2021.108787
36. Shen Y, Jiang Z, Yang D, You S, Luo Y (2026) Multi-objective sensor placement optimization for spatial grid structures based on strain energy analysis. Measurement 264:120252. https://doi.org/https://doi.org/10.1016/j.measurement.2025.120252
37. Lin T, Ren Z, Zhu L, Zhu Y, Feng K, Ding W, Yan K, Beer M (2025) A Systematic Review of Multisensor Information Fusion for Equipment Fault Diagnosis. IEEE Trans Instrum Meas 74:1–48. https://doi.org/10.1109/TIM.2025.3529577
38. Yan X, Shi Z, Sun Z, Zhang C-A (2024) Multisensor Fusion on Hypergraph for Fault Diagnosis. IEEE Trans Industr Inform 20:10008–10018. https://doi.org/10.1109/TII.2024.3393137
39. Yan X, Yan W-J, Xu Y, Yuen K-V (2023) Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network. Mech Syst Signal Process 202:110664. https://doi.org/https://doi.org/10.1016/j.ymssp.2023.110664
40. 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
41. Fedorova D, Tlach V, Kuric I, Dodok T, Zajačko I, Tucki K (2025) Technical Diagnostics of Industrial Robots Using Vibration Signals: Case Study on Detecting Base Unfastening. Applied Sciences (Switzerland) 15:. https://doi.org/10.3390/app15010270
42. Demšar J (2006) Statistical Comparisons of Classifiers over Multiple Data Sets
43. Dietterich TG (1998) Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Comput 10:1895–1923. https://doi.org/10.1162/089976698300017197
44. Papananias M (2025) Bayesian monitoring of machining processes using non-intrusive sensing and on-machine comparator measurement. The International Journal of Advanced Manufacturing Technology 137:1929–1942. https://doi.org/10.1007/s00170-025-15174-x
45. Clark E, Brunton SL, Kutz JN (2020) Multi-fidelity sensor selection: Greedy algorithms to place cheap and expensive sensors with cost constraints
46. Nakai K, Sasaki Y, Nagata T, Yamada K, Saito Y, Nonomura T (2023) Nondominated-Solution-based Multi-objective Greedy Sensor Selection for Optimal Design of Experiments. https://doi.org/10.1109/TSP.2022.3224643
47. Suslu B, Ali F, Jennions IK (2023) Understanding the Role of Sensor Optimisation in Complex Systems. Sensors 23
48. Bagajewicz M, Cabrera E (2003) Pareto Optimal Solutions Visualization Techniques for Multiobjective Design and Upgrade of Instrumentation Networks. Ind Eng Chem Res 42:5195–5203. https://doi.org/10.1021/ie020865g
49. Aldrich C, Auret L (2010) Fault detection and diagnosis with random forest feature extraction and variable importance methods. IFAC Proceedings Volumes 43:79–86. https://doi.org/https://doi.org/10.3182/20100802-3-ZA-2014.00020
50. Siddique AB, Alvee AMM, Das Ana P, Jahan LN, Hashan M (2025) Tree-based, boosting, and stacked models for accurate prediction of total organic carbon from conventional well logs. Energy Reports 13:4491–4513. https://doi.org/https://doi.org/10.1016/j.egyr.2025.04.021
51. Rezvani S, Pourpanah F, Lim CP, Wu QMJ (2024) Methods for Class-Imbalanced Learning with Support Vector Machines: A Review and an Empirical Evaluation. https://doi.org/10.1007/s00500-024-09931-5

52. Elshenawy LM, Chakour C, Mahmoud TA (2022) Fault detection and diagnosis strategy based on k-nearest neighbors and fuzzy C-means clustering algorithm for industrial processes. J Franklin Inst 359:7115–7139. https://doi.org/https://doi.org/10.1016/j.jfranklin.2022.06.022
53. Lundberg S, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions
54. Brusa E, Cibrario L, Delprete C, Di Maggio LG (2023) Explainable AI for Machine Fault Diagnosis: Understanding Features' Contribution in Machine Learning Models for Industrial Condition Monitoring. Applied Sciences (Switzerland) 13:. https://doi.org/10.3390/app13042038
55. Zereen AN, Das A, Uddin J (2024) Machine Fault Diagnosis Using Audio Sensors Data and Explainable AI Techniques-LIME and SHAP. Computers, Materials and Continua 80:3463–3484. https://doi.org/10.32604/cmc.2024.054886
56. Arika AM, Nazmul Huda AS, Ahmad S, Islam A (2026) Vibration-based multi-class fault severity classification of wind turbine gearboxes using explainable machine learning. Energy Conversion and Management: X 30:101621. https://doi.org/https://doi.org/10.1016/j.ecmx.2026.101621
57. Cox D, Hinkley D, Reid N, Rubin D, Silverman B (1960) An Introduction to the Bootstrap. W. Silverman
58. Sujon KM, Hassan R, Choi K, Samad MA (2025) Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. J Big Data 12:268. https://doi.org/10.1186/s40537-025-01313-4
59. Breiman L (2001) Random Forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
60. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63:3–42. https://doi.org/10.1007/s10994-006-6226-1
61. Guryanov A (2019) Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees. In: van der Aalst WMP, Batagelj V, Ignatov DI, Khachay M, Kuskova V, Kutuzov A, Kuznetsov SO, Lomazova IA, Loukachevitch N, Napoli A, Pardalos PM, Pelillo M, Savchenko A V, Tutubalina E (eds) Analysis of Images, Social Networks and Texts. Springer International Publishing, Cham, pp 39–50
62. 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
63. Sperandei S (2014) Understanding logistic regression analysis. Biochem Med (Zagreb) 24:12–18. https://doi.org/10.11613/BM.2014.003
64. Han S-H, Kim KW, Kim S, Youn YC (2018) Artificial Neural Network: Understanding the Basic Concepts without Mathematics. Dement Neurocogn Disord 17:83. https://doi.org/10.12779/dnd.2018.17.3.83
65. Zhang Z (2016) Introduction to machine learning: K-nearest neighbors. Ann Transl Med 4:. https://doi.org/10.21037/atm.2016.03.37
66. Peretz O, Koren M, Koren O (2024) Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. Eng Appl Artif Intell 136:108972. https://doi.org/https://doi.org/10.1016/j.engappai.2024.108972