ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatra

Practice article

Adaptive distribution transformation for enhanced bearing fault detection in independent cart systems

Abdul Jabbar*, Marco Cocconcelli, Gianluca D'Elia

Department of Sciences and Engineering Methods (DISMI), University of Modena and Reggio Emilia (UNIMORE), Via Amendola 2, Pad-Morselli, Reggio Emilia, 42122, Italy

HIGHLIGHTS

- This study introduces an adaptive distribution-based transformation method for detecting bearing faults in independent cart systems.
- The method effectively reshapes skewed and bimodal feature distributions to enhance class separability.
- It demonstrates robustness in the presence of broadband, narrowband, and structural noise contamination.

ARTICLE INFO

Keywords:
Bearing faults
Bimodal distribution
Gaussian merging
Independent cart systems
Condition monitoring
Predictive maintenance
Nonstationary conditions

ABSTRACT

Unlike conventional rotary systems, where bearings remain fixed around a single axis and experience only rotational motion, bearings in independent cart systems exhibit coupled translational–rotational motion and highly nonstationary dynamics. This simultaneous translational-rotational motion of the bearings along the guide rail results in time-varying vibration signatures, complicating fault diagnosis using conventional methods. Furthermore, each cart in an independent cart system is supported by multiple rolling bearings, and industrial machines involve hundreds of carts operating simultaneously, resulting in a large number of interacting bearings whose condition monitoring is both challenging and highly desirable. This study proposes an adaptive distribution-based feature transformation method to enhance the detection of bearing faults under complex operating conditions. The proposed method effectively reshapes the skewed, bimodal, and high-variance feature distributions commonly observed in vibration signals from independent cart systems, thereby improving class separability without the need for labeled data or deep learning architectures. The approach was validated on the open-access bearing dataset based on independent cart systems comprising single- and three-cart experiments and benchmarked against established transformations including Box–Cox, Yeo–Johnson, Rotation-Based Iterative Gaussianization, hyperbolic power transformation, Cumulative Density Function - Transform-and-Shift, dip test based extraction, and Dip transformation. The results demonstrate that the proposed transformation produces compact, well-separated clusters in principal component analysis and t-distributed stochastic neighbor embeddings, achieving very high F1-scores under both One-Class Support Vector Machine and isolation forest anomaly detection methods. The proposed method exhibited robustness against broadband, narrowband, and structural noise, presenting a geometry-agnostic and computationally efficient alternative for real-time bearing diagnostics in non-stationary industrial environments.

1. Introduction

Independent cart systems (ICS), enabled by advancements in linear motor technology, have emerged as modern alternatives to conventional transport mechanisms, such as conveyors, rollers, and chain-driven systems [1–3]. These systems employ linear motors with embedded coils that generate complex magnetic field patterns which interact with permanent magnets mounted on each cart to facilitate the controlled motion along a predefined track. Although magnetically driven actuation eliminates the need for mechanical transmission components and reduces friction losses, wear, and maintenance, the carts are mechanically supported and guided by bearings running along a fixed rail, which remains in constant physical contact. This interface plays a crucial role

* Corresponding author.
Email addresses: abdul.jabbar@unimore.it (A. Jabbar), marco.cocconcelli@unimore.it (M. Cocconcelli), gianluca.delia@unimore.it (G. D'Elia).

https://doi.org/10.1016/j.isatra.2026.06.005
Received 22 October 2025; Received in revised form 3 June 2026; Accepted 3 June 2026
Available online 4 June 2026

in vibration transmission and fault propagation. Consequently, despite the reduction in mechanical complexity, independent cart systems pose significant diagnostic challenges.

Fault detection in ICS presents significant challenges owing to their nonstationary dynamics, modular architecture, and operational variability [4–6]. Individual carts operate at highly variable speeds, ranging from a few millimeters to several meters per second, and can accelerate, decelerate, or rapidly reverse direction. Furthermore, the motion of each cart encompasses both rotational and translational dynamics, as the bearings roll along the guide rail, whereas the cart itself translates along the track. Unlike stationary rotary systems, these hybrid motion patterns result in time-varying vibration signatures that complicate signal interpretation. Additionally, each cart is equipped with multiple bearings, and in large-scale systems involving dozens or even hundreds of carts, this necessitates the simultaneous monitoring of hundreds of interacting components under varying mechanical loads and motion conditions. These factors significantly impede the accurate isolation, localization, and classification of emerging faults.

In recent years, fault diagnosis has advanced toward intelligent frameworks that integrate signal processing with machine learning, with the objective of automating feature extraction and minimizing dependence on expert knowledge. This transition is particularly crucial under non-stationary conditions, where the speed and load fluctuate unpredictably. Traditional vibration-based approaches, including Fourier analysis, envelope detection, and spectral methods, are often inadequate under such variability owing to their stationarity assumptions [7,8]. To address these challenges, more robust signal processing tools have been developed. Angle-domain techniques, such as order tracking and rotational resampling, provide stationary-like representations for speed-varying signals [8], whereas localized spectral methods, including the short-time Fourier transform (STFT), wavelet transform (WT), and Hilbert–Huang transform (HHT), capture transient dynamics in the time–frequency domains [9–11]. Adaptive decomposition techniques, such as empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and variational mode decomposition (VMD), further isolate nonlinear fault-relevant modes [12]. Enhancements through bandwidth division variational mode decomposition (BDVMD) and synchrosqueezing transform (SST) offer improved resolution in noisy environments [13,14].

Recent advancements in preprocessing techniques have been increasingly integrated into machine learning models. Traditional algorithms, such as k-nearest neighbors (k-NN), support vector machines (SVM), and decision trees, utilize statistical features, whereas deep learning models, notably convolutional neural networks (CNNs) and autoencoders, derive hierarchical features from raw signals or spectrograms [12,15]. Reviews by Saufi et al. [15] and Hakim et al. [16] underscore the potential of deep architectures, such as CNNs, deep belief networks (DBNs), and recurrent neural networks (RNNs), and transfer learning frameworks, such as deep adaptation networks (DANs), transfer component analysis (TCA), and adversarial domain adaptation (ADA). These reviews also highlight challenges, such as noise sensitivity, data imbalance, and interpretability. Unsupervised learning methods, including k-means, self-organizing maps, and Isolation Forests, have been applied to fault detection in unlabeled data [17,18] and are often enhanced through autoencoder-based representation learning [19].

Current trends emphasize domain adaptation and unsupervised deep transfer learning (UDTL) to facilitate fault diagnosis across different machines and operational contexts without the need for labeled data [20–22]. Models such as the unsupervised feature alignment network (UFAN) and domain-aligned CNNs aim to align features across diverse regimes [23,24], whereas domain generalization approaches strive for invariant representations without target access [25]. Hybrid models that combine time–frequency analysis with deep learning, such as Ensemble Empirical Mode Decomposition and Continuous Wavelet Transform (EEMD-CWT) [26], and SST-CNN [27]. Additional advancements include wavelet-enhanced classifiers [28], Grasshopper Optimization Algorithm (GOA)-optimized models [29], and interpretable methods such as principal component analysis (PCA), SVM, and tree ensembles for edge deployment [30,31]. Despite these advancements, many methods remain benchmark focused and face challenges in practical deployment under conditions of signal noise, data shifts, or limited labels.

Despite advancements in signal processing and learning-based diagnostic frameworks, most current methodologies focus on extracting or learning discriminative features from raw sensor data, or enhancing classifier architectures through hyperparameter tuning and domain adaptation [15,20]. However, these approaches frequently neglect the fundamental statistical structure of the input features, particularly their distribution shape, which can substantially influence the separability of fault classes. In numerous real-world scenarios, sensor signals display skewed, multimodal, or heavy-tailed distributions that contravene the assumptions of many classifiers and exacerbate decision boundary overlap [32,33]. Although deep models can approximate complex, non-Gaussian feature distributions through hierarchical nonlinear mappings, their reliability diminishes in low-data regimes and under non-stationary or noisy conditions, where overfitting, domain shift, and a lack of generalization significantly impair performance [15,20]. Consequently, there is a critical research gap in explicit distribution transformation techniques, particularly those aimed at reshaping feature distributions (e.g., bimodal or high-variance signals) into more Gaussian-like forms to enhance separability and improve downstream classification and clustering performance. This gap is particularly pronounced in the context of fault detection in complex real-world platforms, such as independent cart systems.

To address this gap, this study introduces an adaptive distribution-based feature transformation algorithm aimed at reshaping skewed, bimodal, and high-variance feature distributions, which are frequently observed in vibration signals from independent cart systems. The proposed method operates without the need for labeled data, model training, or deep learning architectures, rendering it suitable for low-resource diagnostic environments. Rather than modifying classifier architectures or embedding strategies, the approach directly acts on the statistical structure of the feature space to facilitate improved class separability for downstream fault detection and clustering. The method is validated using a newly developed real-world dataset of rolling element bearing faults collected from an industrial-scale independent cart system under non-stationary operating conditions [34]. The symbols and acronyms used throughout this study are summarized in Tables 1–3.

Motivation

To explicitly clarify the drivers behind the proposed methodology, the following factors motivate this study:

- **Economic impact:** ICS are capital-intensive investments in which downtime disrupts production and economic viability. Reliable condition monitoring prevents failures and ensures efficient operations.
- **Nonstationarity:** The ICS operates with temporal variability, including acceleration, deceleration, and shift directions. Bearing motion along the guide rail creates vibration signals that challenge the stationarity and speed invariance assumptions fundamental to conventional diagnostic methods.
- **Intrinsic distributional overlap in extracted features:** The features of ICS vibration data exhibit bimodal or dispersed distributions within specific fault categories. However, there is significant overlap in distributions across different categories, which constrains the performance of anomaly detectors and classifiers.
- **Mismatch with conventional benchmark datasets:** Bearing fault datasets are mainly obtained from stationary rotary systems operating at constant speeds. These datasets poorly represent the hybrid

Table 1
Symbols used in the proposed algorithm.

Symbol	Description
$\mathbf{X} \in \mathbb{R}^{n \times d}$	Input feature matrix with n samples and d features
$\widetilde{X}$	Output matrix of transformed features
F_j	j-th feature vector of length n
$f_{\text{KDE}}(x)$	Kernel Density Estimate of feature distribution
k_j	Number of valid peaks detected in the KDE of F_j
x_p	Peak locations in the estimated density function
x_{valley}	Minimum point between KDE peaks
$p_{\min}$	Minimum peak height threshold for bimodality detection
$z_{j,i}$	Cluster assignment of the i-th sample in feature F_j
$\mu_{j,k}$	Mean of cluster $k \in \{1,2\}$ for feature F_j
$\mu_{\text{higher}}, \mu_{\text{lower}}$	Higher and lower cluster means, respectively
d	Distance between the two cluster centers
σ_{F_j}	Standard deviation of feature F_j
γ	Upper bound for the cluster shift factor α
α	Shift factor to displace higher-valued cluster
μ_{overall}	Global mean of feature F_j
β	Adaptive scaling factor for distribution merging
δ, ϵ, η	Constants controlling scaling behavior
σ_t	Variance threshold for identifying scattered features
μ_{F_j}	Mean of unimodal feature F_j
ξ	Compression scaling factor for scattered distributions

Table 2
Symbols used in additive noise models.

Symbol	Description
$\mathbf{x}$	Original vibration signal
$\tilde{\mathbf{x}}$	Noisy vibration signal
$\mathbf{n}_{\text{AWGN}}$	Additive white Gaussian noise
$\mathbf{n}_{\text{UNI}}$	Uniform quantization noise
σ^2	AWGN variance
a	Uniform half-width (quantization scale)
λ	Proportionality constant for quantization
f_L, f_H	Band-pass filter cutoff frequencies
f_s	Sampling frequency
T	Interference repetition period
$w(t)$	White noise driving signal
$h_{[f_L, f_H]}(t)$	Band-pass filter impulse response
κ	Exponential decay rate of pulse envelope
PFRF[r]	Pseudo-Frequency Response Function shaping filter
$v_1[r], v_2[r]$	Random uniform scaling variables
ζ	Offset to prevent vanishing amplitudes
N_s	Number of sinusoidal terms in PFRF
χ	Amplitude decay rate for spectral tapering

Table 3
Acronyms.

Acronym	Meaning
ICS	Independent Cart system
CWRU	Case Western Reserve University
IMS	Intelligent Maintenance Systems
XJTU-SY	Xi'an Jiaotong University (XJTU) and Changxing Sumyoung Technology Co., Ltd. (SY)
SEU	Southeast University
PHM	Prognostics and Health Management
STFT	short Time Fourier Transform
WT	Wavelet Transfrom
HHT	Hilbert–Huang Transform
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
CEEMDAN	Complete Ensemble Empirical Mode Decomposition With Adaptive Noise
VMD	Variational Mode Decomposition
BDVMD	Bandwidth Division Variational Mode Decomposition
SST	Synchrosqueezing Tranform
k-NN	k-Nearest Neighbors
SVM	Support Vector Machine
CNN	Convolutional Neural Network
DBN	Deep Belief Network
RNN	Recurrent Neural Network
DAN	Deep Adaptation Network
TCA	Transfer Component Analysis
ADA	Adversarial Domain Adaptation
UDTL	Unsupervised Domain Transfer Learning
UFAN	Unsupervised Feature Alignment Network
CWT	Continuous Wavelet Transform
GOA	Grasshopper Optimization Algorithm
PCA	Principal Component Analysis
FPGA	Field-Programmable Gate Array
RBIG	Rotation-Based Iterative Gaussianization
CDF	cumulative distribution function
KDE	Kernel Density Estimate
t-SNE	t-distributed Stochastic Neighbor Embedding
CDF–TS	Cumulative Density Function - Transform-and-Shift
OCSVM	One-Class Support Vector Machine
IF	Isolation forest
SNR	Signal-to-noise ratio
THD	Total harmonic distortion
SINAD	Signal to noise and distortion ratio
ANOVA	Analysis of Variance
OWA	One-way ANOVA
KW	Kruskal–Wallis
LS	Laplacian Score
Var	Variance
Mono	Monotonicity
HPT	Hyperbolic Power Transformation
FRF	Frequency Response Function
AWGN	Additive White Gaussian Noise

dynamics, nonstationarity, and modularity of ICS, thereby limiting the applicability of the methods validated on these benchmarks.

- **Label scarcity and industrial deployment constraints:** In industrial settings, the limited availability of labeled fault data and confidentiality issues necessitate the development of diagnostic methods that avoid extensive training while maintaining efficiency and interpretability for real-time monitoring.
- **Need for distribution-aware transformations:** Improving fault separability in ICS requires explicit distribution-aware transformations that reshape bimodal and high-variance features into compact separable forms before anomaly detection or clustering.

Key contributions

The principal contributions of this study are as follows:

- **Distribution-centric diagnostic perspective for ICS vibration data:** This study identifies distributional overlap, bimodality, and variance as key obstacles to fault separability in vibration features from cart systems under nonstationary conditions, framing the statistical structure of the feature space as an underexplored challenge.
- **Adaptive distribution reshaping framework:** An adaptive, distribution-aware feature transformation algorithm effectively reshapes skewed, bimodal, and high variance distributions. This method does not require training and is computationally efficient, making it particularly suitable for industrial environments where labels are limited.
- **Unified treatment of bimodal and highly dispersed unimodal features:** The transformation mechanism addresses bimodal distributions and dispersed unimodal features by employing density estimation, clustering separation, adaptive shifting, and variance scaling to enhance class separability.
- **Validation under realistic industrial operating conditions:** The proposed methodology was validated using a practical dataset from an independent cart system collected under variable speed, bidirectional movement, and multi-cart operation conditions. This dataset captures the hybrid translational–rotational dynamics that are not represented in conventional bearing fault benchmarks.
- **Robustness and comparative assessment against established transformations:** The robustness of the system was assessed under various noise conditions, including broadband noise, narrowband interference, and structural resonance. Comparative analysis with

established techniques—Box-Cox, Yeo-Johnson, rotation-based Gaussianization, hyperbolic power transformation, and CDF-based homogenization—showed improved feature compactness and separability, leading to superior anomaly detection and clustering performance.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature. Section 3 presents the problem formulation and the associated challenges. The proposed adaptive distribution-based transformation methodology is described in Section 4. Section 5 describes the experimental setup and dataset configuration. The results and performance analysis are discussed in Section 6, and the main conclusions are summarized in Section 7.

2. Related work

The presence of non-Gaussian characteristics in process and vibration data poses significant challenges in control performance assessment, reliability analysis, anomaly detection, and fault diagnosis. This has led to the development of various distribution transformation and modeling strategies aimed at reshaping, Gaussianizing, and regularizing data to enhance their interpretability and analytical consistency. Initial efforts concentrated on quantile-based and power-transformation methods, which provided interpretable, training-free approaches to approximate Gaussianity. A quantile transformation framework that converts non-Gaussian system outputs into Gaussian-like representations while preserving mutual information between the inputs and outputs, thereby facilitating consistent controller benchmarking and tuning under non-Gaussian disturbances, was introduced in [35].

Subsequent advancements have resulted in more general Gaussianization techniques for high-dimensional data. The Rotation-Based Iterative Gaussianization (RBIG) algorithm proposed in [36] incrementally applies marginal Gaussianization and orthonormal rotations to achieve joint Gaussian representation. This concept was expanded to a convolutional RBIG formulation in [37], where learned convolutional rotations enabled scalable layer-wise Gaussianization suitable for large image datasets. To enhance computational efficiency, an optimized field-programmable gate array (FPGA) implementation of RBIG (ORBIG) was presented in [38] for real-time hyperspectral anomaly detection, substituting eigen-decomposition with correlation-based rotations and parallelized cumulative distribution computations. As a complementary approach, a fully unsupervised invertible Gaussian model for change and anomaly detection was introduced in [39], wherein multivariate data were transformed via differentiable mappings that allow direct likelihood estimation under a Gaussian prior. Collectively, these studies demonstrate the efficacy of distribution Gaussianization as an interpretable, model-free method for aligning data with analytical frameworks predicated on Gaussian assumptions.

In the domain of industrial fault diagnosis, power transformations such as the Box–Cox and Yeo–Johnson families remain prevalent for addressing skewed or heavy-tailed monitoring statistics. Box–Cox normalization has been employed within multi-regime canonical correlation analysis frameworks [40], enabling the use of Gaussian-based control limits for fault detection across varying operating conditions. Further improvements in classifier reliability under non-Gaussian conditions have been achieved through Box–Cox preprocessing in HVAC sensor fault detection [41], while the combination of Yeo–Johnson transformations with noise-reducing resampling has proven effective for mitigating class imbalance in transformer fault datasets [42]. Building on these power-transformation principles, an improved Box–Cox sparsity measure (IBCSM) was proposed in [43], which integrates transformation parameters directly into sparsity metrics to increase sensitivity to weak fault signatures. Related work in [44] generalized the envelope spectrum through a Box–Cox-based transformation that consolidates multiple spectra into a single product envelope spectrum (PES), effectively amplifying consistent fault frequencies, even under low signal-to-noise conditions. These studies emphasize that well-designed feature transformations can enhance the separability and monotonicity of fault indicators without relying on complex learned models.

In addition to marginal transformations, unsupervised and one-class methodologies have investigated nonparametric density modeling to characterize nominal system behavior in non-Gaussian environments. The ICA–DW–SVDD framework described in [45] extracts independent components, selects decorrelated features using the Durbin–Watson criterion, and encapsulates nominal data through one-class Support Vector Data Description. Kernel density estimation (KDE) was applied in [46] to model the distribution of healthy motor current signals and identify anomalies via divergence measures against the nominal density. The lightweight KDE-based anomaly detector proposed in [47] extends this concept to multivariate time-series monitoring, learning exclusively from normal process data while adapting to gradual concept drift. These methodologies share the common objective of constructing or transforming distributions to approximate stable Gaussian-like structures, from which deviations indicate anomalies or faults.

Efforts to generalize these models across domains have incorporated distributional alignment principles. The Variance Discrepancy Representation (VDR) metric introduced in [48] aligns both the mean and variance of the feature distributions between the source and target domains using a robust kernel function. This approach demonstrates greater efficacy than conventional mean matching techniques under heavy-tailed and asymmetric distributions, underscoring that controlling higher-order distributional properties is essential for consistent fault diagnosis under variable operating conditions.

Overall, these contributions provide the conceptual foundation for the present study, which develops an adaptive, interpretable transformation for reshaping nonGaussian feature distributions in vibration-based diagnostics.

3. Problem formulation and diagnostic challenges in ICS

To underscore the necessity for advanced data-driven learning frameworks in Independent Cart Systems (ICS), it is imperative to first elucidate the inherent limitations of direct signal-level interpretation under realistic operational conditions. In contrast to conventional rotating machinery, where fault signatures are typically referenced to a dominant rotational order and can often be rendered quasi-stationary through speed-synchronous analysis, vibration signals acquired from ICS are influenced by coupled translational-rotational motion, time-varying speed profiles, repeated acceleration-deceleration phases, and spatial interactions with guide rails. These factors significantly complicate the identification of discriminative fault characteristics from the raw measurements.

Fig. 1 presents representative raw vibration signals acquired from a healthy bearing and an inner-race fault of 0.25mm during a bidirectional traversal along a curved motor module. Impulsive events are intermittently observed in *both* healthy and faulty conditions, primarily arising from repeated acceleration-deceleration phases and transient interactions between the cart and guide rail. However, the overall waveforms are dominated by the large-amplitude modulation effects induced by the motion dynamics. Consequently, fault-related components are neither temporally isolated nor consistently repeatable across cycles, rendering a reliable diagnosis based solely on direct time-domain inspection infeasible. To further assess the frequency-domain separability, Fig. 2 depicts the power spectral density (PSD) estimates for multiple bearing conditions. Although broad energy redistribution is observable, the dominant spectral content is largely governed by structural resonances and motion-induced excitation. Fault-related variations appear as subtle deviations rather than distinct spectral lines, and a strong overlap persists across healthy and faulty conditions, even when averaged spectra are considered.

Envelope analysis is commonly adopted to enhance impulsive fault signatures; however, Fig. 3 demonstrates that under nonstationary ICS

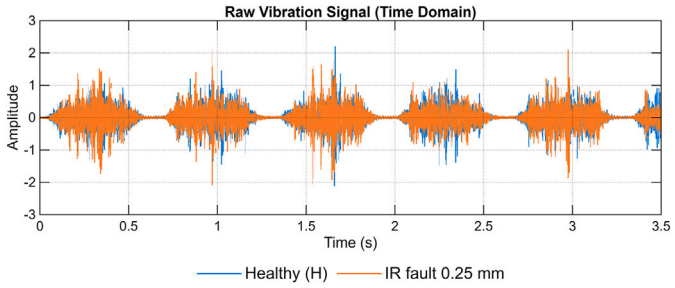

Fig. 1. Vibration signals were obtained from an accelerometer affixed to the guide rail during a single bidirectional traversal of a curved module under both healthy (H) and inner-race fault (IR 0.25 mm) conditions. These signals demonstrated pronounced amplitude modulation due to the acceleration, steady motion, and deceleration phases. The impulses associated with faults are obscured by variability induced by motion and are not consistently discernible in the time domain.

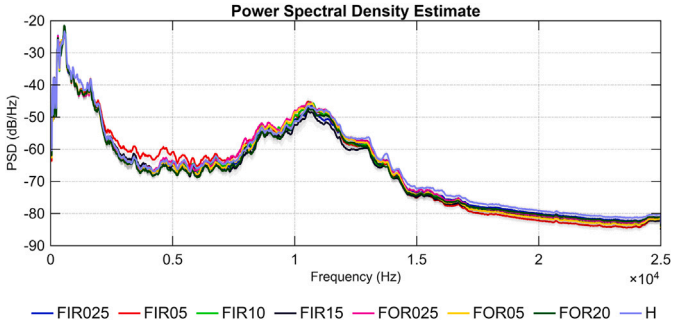

Fig. 2. Power spectral density (PSD) estimates for both healthy and faulty bearing conditions under nonstationary ICS operation reveal energy redistribution across frequency bands. However, the dominant spectral components are predominantly influenced by structural and motion-related effects of the target. The significant overlap among conditions constrains the ability to distinguish faults within the frequency domain.

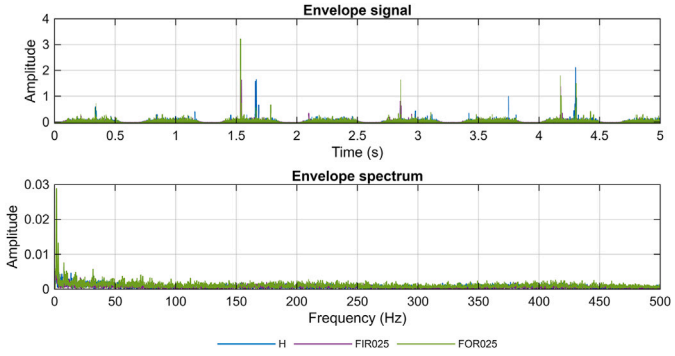

Fig. 3. The envelope signals and their corresponding envelope spectra, derived through band-limited Hilbert demodulation, are analyzed for various bearing conditions. Although envelope analysis is effective in highlighting impulsive behavior, the resultant representations are significantly influenced by nonstationary speed variations and transient impacts, leading to broadband and overlapping spectral characteristics.

operation, envelope signals and their corresponding spectra remain heavily influenced by speed fluctuations and transient impacts. Although certain amplitude differences emerge, they are intermittent, broadband, and highly sensitive to the selected demodulation band, limiting their reliability as standalone diagnostic indicators. Finally, Fig. 4 illustrates the time-frequency representations of the same signals. The spectrograms reveal pronounced nonstationarity, with energy concentrations that

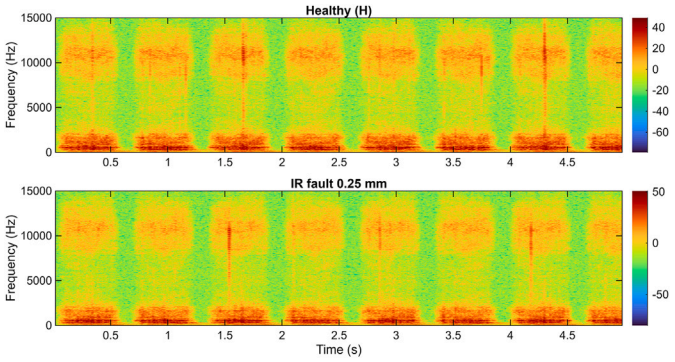

Fig. 4. The time-frequency spectrograms of vibration signals from both healthy and inner-race faulty conditions during bidirectional cart motion along a curved module reveal significant nonstationarity and dynamic energy concentrations associated with phases of acceleration, deceleration, and direction reversal. Patterns indicative of faults are subtle and exhibit substantial overlap with those of healthy behavior.

evolve continuously over time and frequency as the cart accelerates, decelerates, and reverses direction. Although localized transient patterns can be observed, these structures do not yield clear, manually separable fault signatures and exhibit significant overlap between healthy and faulty cases.

Collectively, Figs. 1–4 demonstrate that, under realistic ICS operating conditions, neither time-domain inspection nor conventional frequency- or time–frequency-domain representations produce stable or manually separable fault signatures. Consequently, a reliable diagnosis cannot be achieved solely at the signal level. This necessitates the extraction of higher-level representations through windowed feature computation, which facilitates the aggregation of information across the time, frequency, and motion phases. However, as will be discussed subsequently, the challenges posed by ICS nonstationarity persist even after feature extraction, manifesting as overlapping and poorly structured feature distributions.

While fault diagnosis for conventional rotating machinery has been extensively studied, the signal-level limitations illustrated in Figs. 1–4 necessitate a feature-based diagnostic approach for Independent Cart Systems (ICS). In contrast to fixed-axis rotating machinery, ICS bearings are subject to coupled rotational and translational motions as carts navigate guide rails under time-varying speed profiles. Furthermore, the presence of multiple carts, each equipped with several bearings, results in rapidly expanding monitoring tasks and high-dimensional feature spaces. These characteristics introduce additional complexity beyond that encountered in classical rotating systems and necessitate diagnostic frameworks capable of addressing strong nonstationarity, motion-dependent variability, and limited class separability issues.

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ represent the feature matrix, where n denotes the number of samples and d denotes the number of features extracted from various sensor modalities. Each $\mathbf{x}_i \in \mathbb{R}^d$ corresponds to a sliding window derived from the synchronized vibration and position data acquired during the cart motion cycle. The window extraction strategy is illustrated in Fig. 5, where the segmentation is aligned with the forward and reverse cart motion events.

Fig. 5 depicts the motion indicators utilized in Experiment Type 2 of the bearing dataset [34], wherein a single cart performed a bidirectional back-and-forth movement along a curved motor module. In this setup, one forward traversal corresponds precisely to the entire length of the curved module, followed by a reverse traversal along the identical path. Consequently, the cart's *speed* and *position* are employed to characterize the inherently non-stationary operating states and define physically consistent signal-extraction windows. The *sign of the speed* differentiates between forward and reverse motions, whereas

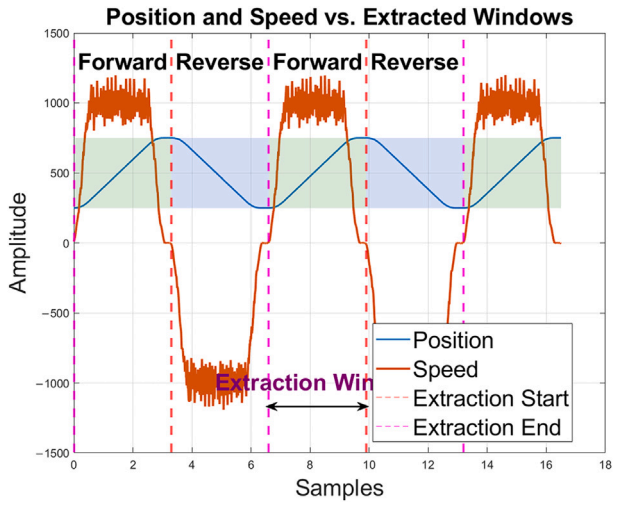

Fig. 5. Window segmentation is conducted based on the position of the cart. The motion phases are divided into forward and reverse intervals to facilitate feature extraction.

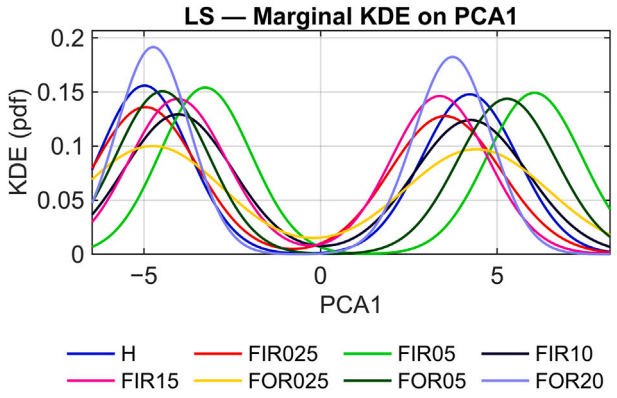

Fig. 6. The marginal kernel density estimate (KDE) along the first principal component for various fault classes. Notably, several classes demonstrate overlapping and multi-modal distributions.

its *time-varying magnitude* indicates the acceleration, steady state, and deceleration phases within each traversal. The *position* signal provides a spatial reference, ensuring that each extracted window corresponds to the same track segment across repeated cycles, despite variations in speed. Based on these indicators, the signal windows were defined over half-cycles, each representing a complete motion in a single direction (acceleration–steady–deceleration). These indicators were employed exclusively for motion-state characterization and window segmentation and were not incorporated as diagnostic features in the proposed framework.

Although dimensionality reduction techniques are employed, numerous features continue to exhibit inadequate separability due to the following reasons.

- **Bimodal distributions:** Certain features exhibit overlapping class distributions characterized by multiple peaks, which complicate the delineation of boundaries. This phenomenon is illustrated in Fig. 6 for the first principal component.
- **Scattered unimodal features:** Other features adhere to unimodal Gaussian-like distributions but exhibit substantial intra-class variance, as depicted in the t-SNE plots (Fig. 7).

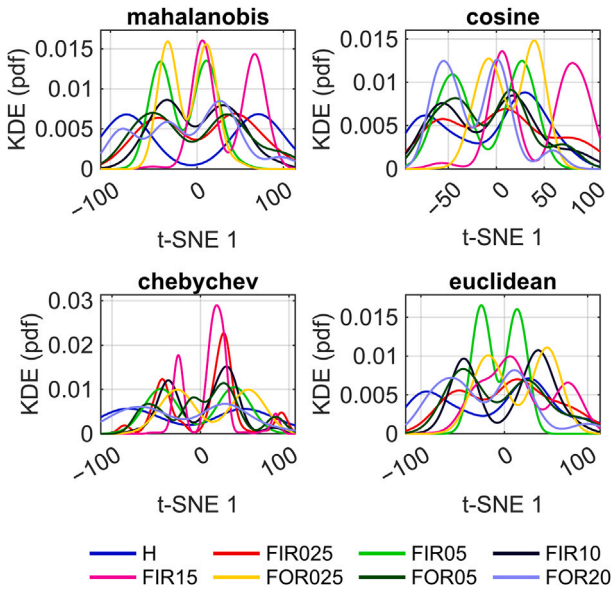

Fig. 7. Marginal KDEs along the first t-SNE dimension employing different distance metrics. Class overlap and dispersed distributions are evident even within non-linear embeddings.

Key Challenges:

- **Label scarcity:** Acquiring fault labels in industrial systems is both costly and labor-intensive, necessitating models to generalize with minimal supervision.
- **High dimensionality:** The integration of features from vibration and motion sensors results in increased redundancy and noise.
- **Overlapping classes:** As illustrated in Figs. 6 and 7, even reduced features display bimodality or scattering, which hinders classification.
- **Scalability:** The deployment of multiple carts exponentially increases diagnostic complexity.
- **Nonstationary dynamics:** The concurrent rotation and translation generate complex, time-varying signal patterns that are uncommon in fixed-rotation systems.

This study presents an innovative adaptive transformation method that explicitly reshapes feature distributions, thereby enhancing class separability under challenging diagnostic conditions. This facilitates fault classification in ICS environments, which are characterized by limited labels, significant overlap, and intricate motion dynamics.

4. Proposed methodology

This section introduces a distribution-aware diagnostic framework specifically designed for ICS, which addresses the complexities associated with high-dimensional feature spaces and overlapping class distributions. Two complementary processing pipelines are proposed. The first pipeline employs a distribution-based transformation directly within the original feature space, foregoing an explicit dimensionality reduction. Conversely, the second pipeline applies dimensionality reduction prior to the proposed transformation, facilitating low-dimensional analysis and visualization. Both pipelines incorporate identical feature extraction and ranking stages, differing solely in the application of dimensionality reduction before the transformation stage.

As depicted in Figs. 8 and 9, the initial step involves segmenting the vibration and system-level signals from the ICS, followed by the extraction and ranking of time-domain features using one-way

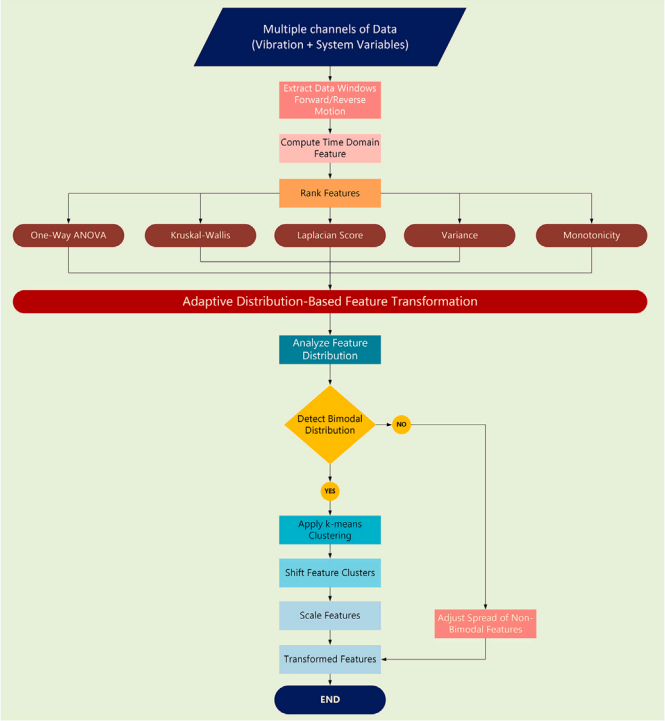

Fig. 8. An overview of the proposed adaptive distribution-based feature transformation pipeline is presented, which functions without the need for explicit dimensionality reduction. Multichannel vibration signals are segmented to compute the time-domain features. These features are subsequently ranked using multiple relevance criteria, including one-way analysis of variance (ANOVA), Kruskal–Wallis, Laplacian Score, Variance, and Monotonicity. The proposed transformation is then applied directly within the original feature space by analyzing and reshaping marginal feature distributions.

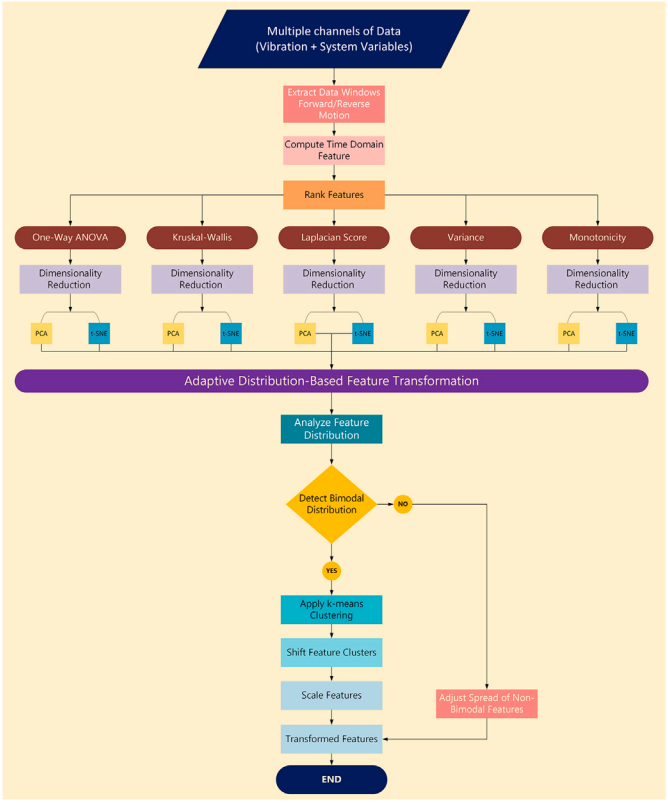

Fig. 9. The proposed transformation pipeline advances from the extraction of multi-channel features to the enhancement of class separability through distribution shaping. During the preprocessing phase, multichannel sensor signals are segmented to compute time-domain features, which are subsequently ranked based on multiple relevance criteria. Dimensionality reduction is implemented prior to the proposed transformation.

ANOVA, Kruskal–Wallis, Laplacian Score, Variance, and Monotonicity. In the pipeline illustrated in Fig. 9, dimensionality reduction techniques, such as PCA and t-SNE, are employed post-feature ranking to facilitate visual and quantitative analyses within a low-dimensional space prior to transformation. Conversely, the pipeline in Fig. 8 implements the proposed distribution-based transformation directly on the ranked feature space, bypassing the intermediate dimensionality reduction. The core of the framework is a distribution-based transformation strategy that adaptively modifies the feature distributions to enhance class separability. The comprehensive procedure is elaborated in Algorithm 1, where both bimodal and high-variance unimodal features are addressed through KDE analysis, clustering, shifting, and scaling operations.

4.1. System variables and feature extraction

The ICS platform is equipped with multiple sensor modalities, including two triaxial and three monoaxial accelerometers, along with system-level signals, such as the following error, set current, and velocity error. The raw signals were segmented using synchronized motion intervals (see Fig. 5), and time-domain statistical features—mean, standard deviation, skewness, kurtosis, and peak-to-peak—were computed per segment, excluding the cart position and actual velocity to avoid redundancy.

This yields a high-dimensional feature matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, where n is the number of motion windows, and d is the number of extracted features. To manage redundancy and improve model tractability, the feature space is ranked and reduced using unsupervised and supervised metrics.

4.2. Feature ranking and optional dimensionality reduction

To identify the most discriminative and structurally informative features, five complementary ranking strategies were employed.

- **One-Way ANOVA**—a parametric test assessing inter-class variance assuming Gaussianity;
- **Kruskal–Wallis test**—a nonparametric rank-based alternative robust to outliers and skewness;
- **Laplacian Score**—a graph-based filter method evaluating local manifold preservation via neighborhood affinity;
- **Variance Ranking**—selecting features with high dispersion under the assumption that variance correlates with information richness;
- **Monotonicity Score**—quantifying the directional consistency of feature trends with degradation progression.

4.3. Adaptive distribution-based feature transformation

To enhance the separability among fault classes in compressed feature spaces, particularly when features exhibit either bimodal or scattered unimodal distributions, an adaptive transformation strategy is proposed. The procedure comprises six sequential steps, which are outlined below.

Step 1: Density estimation. Each feature F_j is evaluated using the kernel density estimator $f_{\mathrm{KDE}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-F_{j,i}}{h}\right)$. In this framework, $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ represents the kernel smoothing function, and h denotes

Algorithm 1 Adaptive Distribution-Based Feature Transformation.

Require: $\mathbf{X} \in \mathbb{R}^{n\times d}$, $c_{\min}$, γ, δ, ϵ, η, σ_t, ξ
Ensure: $\widetilde{X}$
  Initialize $\widetilde{X} \leftarrow \emptyset$; $c_{\min}$, γ, δ, ϵ, η, σ_t, ξ
  for each feature F_j in **X do**
    Compute KDE: $f_{\mathrm{KDE}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-F_{j,i}}{h}\right)$
    Identify peaks x_p where $f'_{\mathrm{KDE}}(x_p) = 0$ and $f''_{\mathrm{KDE}}(x_p) < 0$
    Retain peaks where $f_{\mathrm{KDE}}(x_p) - f_{\mathrm{KDE}}(x_{\mathrm{valley}}) \geq p_{\min}$
    Count number of valid peaks k_j
    if $k_j > 1$ **then**
      Assign clusters $z_{j,i} = \arg\min_{k\in\{1,2\}} \|F_{j,i} - \mu_{j,k}\|^2$
      Update cluster centers: $\mu_{j,k} = \frac{1}{|\{i:z_{j,i}=k\}|}\sum_{i:z_{j,i}=k} F_{j,i}$
      $\mu_{\mathrm{higher}} \leftarrow \max(\mu_1, \mu_2)$
      $\mu_{\mathrm{lower}} \leftarrow \min(\mu_1, \mu_2)$
      $d \leftarrow \mu_{\mathrm{higher}} - \mu_{\mathrm{lower}}$
      $\alpha \leftarrow \min(\gamma, \frac{d}{\sigma_{F_j}} \cdot \gamma)$
      Shift higher cluster: $F_{j,\mathrm{shifted}} = F_{j,\mathrm{higher}} - \alpha \cdot d$
      $\beta \leftarrow \max\left(\delta, \epsilon - \frac{\eta\cdot d}{\sigma_{F_j}}\right)$
      $\mu_{\mathrm{overall}} \leftarrow \mathrm{mean}(F_j)$
      $F_{j,\mathrm{scaled}} = \mu_{\mathrm{overall}} + \beta \cdot (F_j - \mu_{\mathrm{overall}})$
      Store $F_{j,\mathrm{scaled}}$ in $\widetilde{X}$
    else if $\sigma_{F_j} > \sigma_t$ **then**
      $\mu_{F_j} \leftarrow \mathrm{mean}(F_j)$
      $F_{j,\mathrm{centered}} = \mu_{F_j} + \xi \cdot (F_j - \mu_{F_j})$
      Store $F_{j,\mathrm{centered}}$ in $\widetilde{X}$
    else
      Store original F_j in $\widetilde{X}$
    end if
  end for
  return $\widetilde{X}$

the bandwidth, which is determined using the normal-approximation rule (Silverman's rule of thumb) as implemented in MATLAB `ksdensity`. This data-driven selection circumvents the introduction of an additional tuned hyperparameter and provides an optimal smoothing level for near-Gaussian distribution features. Given that undersmoothing may result in spurious local maxima in scattered unimodal features, bimodality is not inferred solely from the peak count but is instead assessed using a minimum peak prominence threshold $p_{\min}$. Significant peaks in the KDE are identified as local maxima satisfying $\frac{d}{dx} f_{\mathrm{KDE}}(x) = 0$ and $\frac{d^2}{dx^2} f_{\mathrm{KDE}}(x) < 0$. A peak x_p is considered valid if $f_{\mathrm{KDE}}(x_p) - f_{\mathrm{KDE}}(x_{\mathrm{valley}}) \geq p_{\min}$, where x_{valley} denotes the nearest local minimum.

Step 2: Bimodality detection. A feature is classified as bimodal if more than one peak in its density exceeds the minimum threshold.

$$k_j = \sum_{x_p} \mathbb{1}\left[f_{\mathrm{KDE}}(x_p) - f_{\mathrm{KDE}}(x_{\mathrm{valley}}) \geq p_{\min}\right], \tag{1}$$

In this context, $\mathbb{1}[\cdot]$ represents the indicator function, while $p_{\min} = c_{\min} \max \hat{f}_j(x)$, where $c_{\min} \in (0, 1)$, signifies a user-defined relative density factor that regulates the stringency of peak detection for the jth feature. A feature is deemed bimodal if $k_j > 1$.

Step 3: Distribution separation via clustering. To identify and isolate distinct modes within the bimodal features, we applied k-means clustering. To mitigate the sensitivity associated with the k-means initialization, clustering was conducted in a one-dimensional context with $k = 2$ and multiple replicates, preserving the solution that exhibited the minimum within-cluster dispersion. Empirically, the resultant transformation and subsequent detection performance demonstrated insensitivity to random initializations. The objective function is given as

$$J = \sum_{i=1}^{n}\sum_{k=1}^{2} \mathbb{I}[z_{j,i} = k]\,\|F_{j,i} - \mu_{j,k}\|^2, \tag{2}$$

where $\mathbb{I}[\cdot]$ is the indicator function, $F_{j,i}$ denotes the i-th sample of feature F_j, and $z_{j,i}$ is the cluster assignment.

$$z_{j,i} = \arg\min_{k\in\{1,2\}} \|F_{j,i} - \mu_{j,k}\|^2. \tag{3}$$

Cluster means $\mu_{j,k}$ are iteratively updated as:

$$\mu_{j,k} = \frac{1}{|\{i : z_{j,i} = k\}|}\sum_{i:z_{j,i}=k} F_{j,i}. \tag{4}$$

Step 4: Adaptive cluster shifting. To reduce the inter-class overlap, we applied a directional shift to clusters with higher means. Let:

$$\mu_{\mathrm{high}} = \max(\mu_{j,1}, \mu_{j,2}), \tag{5}$$

$$\mu_{\mathrm{low}} = \min(\mu_{j,1}, \mu_{j,2}), \tag{6}$$

and compute the inter-cluster distance $d = \mu_{\mathrm{high}} - \mu_{\mathrm{low}}$. The adaptive shift factor is defined as

$$\alpha = \min\left(\gamma, \frac{d}{\sigma_{F_j}} \cdot \gamma\right), \tag{7}$$

where γ is a user-defined upper bound (e.g., $\gamma = 0.5$), and σ_{F_j} is the standard deviation of feature F_j. All samples in the higher cluster were then shifted as follows:

$$F_{j,\mathrm{shifted}} = F_{j,\mathrm{high}} - \alpha \cdot d. \tag{8}$$

Step 5: Smooth merging via scaling. To control the residual overlap and ensure continuity in the transformed distributions, a scaling factor β is applied as follows:

$$\beta = \max\left(\delta, \epsilon - \frac{\eta \cdot d}{\sigma_{F_j}}\right), \tag{9}$$

where δ, ϵ, and η are user-defined constants that control the scaling bound. The final transformed feature is then rescaled to its global mean μ_{overall}.

$$F_{j,\mathrm{scaled}} = \mu_{\mathrm{overall}} + \beta \cdot (F_j - \mu_{\mathrm{overall}}). \tag{10}$$

Step 6: Compression of scattered features. For features with no detected bimodality ($k_j \leq 1$) but with excessive spread ($\sigma_{F_j} > \sigma_t$), a variance compression operation is applied as follows:

$$F_{j,\mathrm{centered}} = \mu_{F_j} + \xi \cdot (F_j - \mu_{F_j}), \tag{11}$$

where μ_{F_j} is the mean of feature F_j, and $\xi \in (0, 1)$ is a user-defined compression factor that reduces the dispersion while preserving the overall shape of the distribution.

Existing publicly available fault datasets, such as the Case Western Reserve University (CWRU) bearing dataset, Intelligent Maintenance Systems (IMS) bearing dataset, Politecnico di Torino bearing dataset, and Xi'an Jiaotong University (XJTU) and Changxing Sumyoung Technology Co., Ltd. (SY) (XJTU-SY) rolling bearing dataset, Southeast University (SEU) gearbox fault dataset, and Prognostics and Health Management (PHM) challenge dataset were primarily designed for conventional rotary systems with stationary or uniform-speed configurations. These datasets do not adequately represent the hybrid translational–rotational dynamics inherent in Independent Cart Systems (ICS), which result in fundamentally different signal characteristics. Despite the increasing industrial significance of ICS, research on condition monitoring for these systems remains limited, primarily because of proprietary restrictions and the absence of publicly available datasets. This study addresses this gap by focusing specifically on the diagnostic challenges that are unique to this platform.

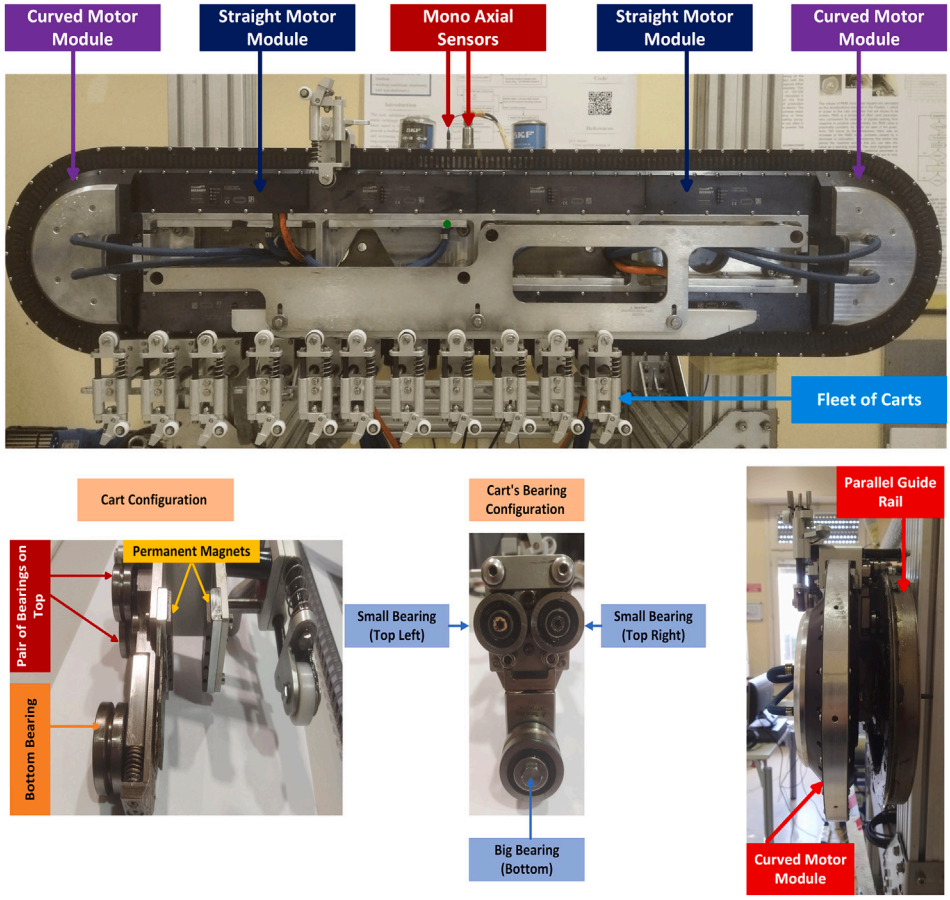

Fig. 10. An overview of the experimental setup utilizing the Beckhoff XTS platform is presented. **Top:** a comprehensive view of the closed-loop track is depicted, illustrating the modular configuration and the multiple carts affixed along the guide rail. **Bottom left:** a detailed side view of an individual cart is provided, emphasizing the pair of identical small bearings located at the top, the larger bearing at the bottom, and the integrated permanent magnets. **Bottom center:** the front view of the cart's bearing arrangement is shown. **Bottom right:** a side view of the curved motor module is displayed, along with the parallel guide rail that provides mechanical guidance.

5. Experimental setup

The experimental platform was centered on the Beckhoff eXtended Transport System (XTS), a contemporary linear motor-driven conveyance system designed for high-speed programmable motion. This system offers a modular framework for investigating vibration-based diagnostics in ICS, which is an emerging alternative to traditional rotary or belt-driven motion platforms.

5.1. System overview

ICS technologies are increasingly supplanting conventional chain and belt conveyance systems that depend on interconnected motors, pulleys, and mechanical couplings to transport materials. These older systems are prone to cumulative wear, mechanical complexity, and energy inefficiencies owing to their distributed actuation. In contrast, ICS platforms reduce mechanical complexity by employing stationary linear motor stators that generate dynamic electromagnetic fields to propel individually controlled carts along closed-loop tracks. The primary advantages of this architecture are its frictionless propulsion, compact mechanical footprint, and enhanced programmability. However, although propulsion is contactless, *structural support is still provided mechanically* via *rolling element bearings* that are mounted on each cart. These bearings traverse a precision guide rail and experience load dynamics distinct from those encountered in conventional rotary machinery.

5.2. Track and actuation configuration

As shown in Fig. 10, the experimental configuration comprises an oval track constructed from modular stator segments (Beckhoff XTS) organized into straight and curved sections. Each cart is equipped with embedded permanent magnets that interact with the electromagnetic fields generated by the stator, facilitating bidirectional high-speed movement. Fig. 11 presents a detailed annotated schematic depicting the arrangement of the linear motor modules, absolute positions along the track, and placement of sensors.

Carts can autonomously reverse or accelerate within specified spatial parameters, thereby facilitating precise positioning and executing complex trajectories. A *back-and-forth motion routine* was implemented to simulate realistic bidirectional loading and investigate the impact of force reversals on bearing dynamics.

5.3. Bearing assembly and mechanical guidance

Each cart is mechanically guided by three deep-groove rolling element bearings, which form a stable V-shaped mounting profile against the guide rail. The upper side accommodates *two identical small bearings*, whereas the lower side supports a *larger third bearing*, thereby providing lateral restraint and vertical stability.

Table 4 presents the key specifications of these bearings, including the race diameter and ball geometry. All bearings were sourced from *HepcoMotion's GFX system*, specifically engineered for the Beckhoff

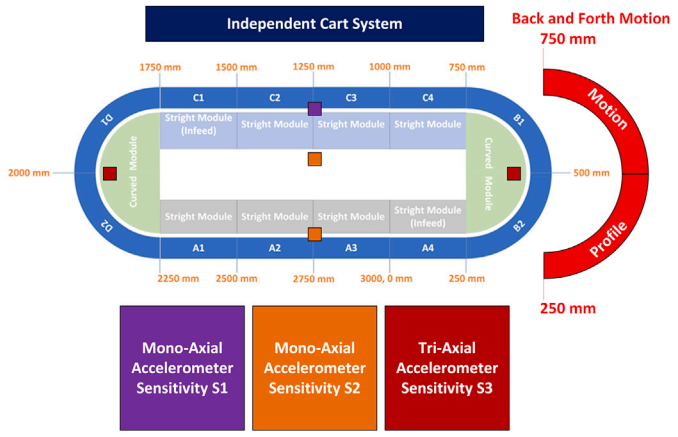

Fig. 11. The topological schematic of the experimental independent cart system (ICS) is presented, featuring annotated dimensions, sensor locations, and the motion profile. The oval-shaped track comprises straight (C1–C4 and A1–A4) and curved (D1–D2 and B1–B2) modules, which facilitate a closed loop. Colored squares indicate accelerometer placements: **S1** and **S2** represent the mono-axial sensor sensitivities on the straight sections, whereas **S3** denotes the tri-axial sensor sensitivity located on the curved module. The dimension labels specify the distances between the key module boundaries (in mm). The red arc on the right delineates the 750 mm back-and-forth reversal zone employed during testing to induce bidirectional motion. This configuration enables the analysis of sensor response under varying kinematics and track geometries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

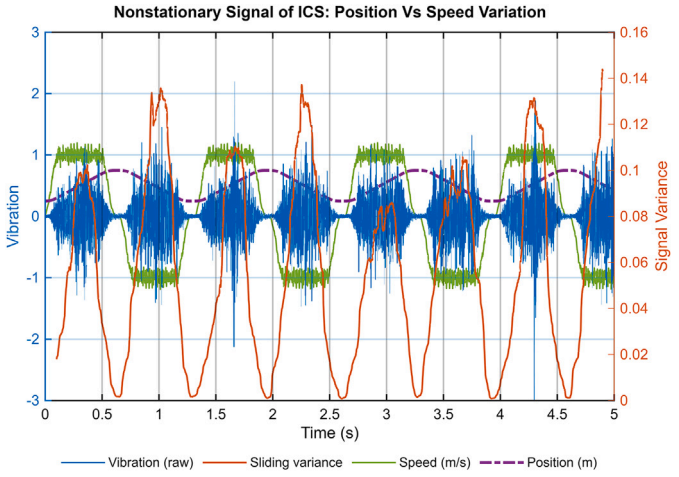

Fig. 12. A representative example of a nonstationary vibration response measured on an ICS-based experimental testbed. The **raw vibration signal** recorded at the guide rail is depicted on the left-hand side of the axis. The **variance** of the vibration signal, illustrated as a curve on the right axis, was computed using a sliding window, thereby highlighting the pronounced time-varying energy and non-stationarity. The cart's **speed** and **position** along the track were overlaid after unit normalization to provide the motion context. A clear modulation of the vibration variance was observed during the phases of acceleration, deceleration, and direction reversal, illustrating the strong coupling between the translational motion, bearing–rail interaction, and vibration characteristics of ICS platforms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Table 4
Specifications for bearings are provided herein. All measurements are expressed in millimeters (mm). The abbreviations used are as follows: OR denotes the Outer Race, IR signifies the Inner Race, and Ø represents the diameter.

Type	Rows	Balls	Ball Ø	OR Ø	IR Ø	Pitch Ø
Top Bearing	2	7, 7	3.95	25.00	10.75	14.70
Bottom Bearing	2	7, 7	5.55	34.00	14.60	20.15

XTS platform, featuring *steel-on-steel contact* surfaces to ensure wear resistance under dynamic loading conditions.

To minimize chatter and vibration artifacts, a preload mechanism uniformly pressed the bearings against the rail throughout the motion cycles. This preload ensured rigid alignment, particularly during rapid acceleration, deceleration, and directional changes.

5.4. Sensor configuration and data logging

The testbed is equipped with five accelerometers: two monoaxial and three triaxial sensors, strategically positioned across the straight and curved track segments. Additionally, *system-level variables* such as cart position, velocity, drive current, following error, and controller status were recorded.

All data channels were synchronized using Beckhoff's TwinCAT control platform to enable time-aligned signal analyses. Sampling was conducted at a high resolution to capture both low-frequency load transitions and high-frequency bearing signatures.

5.5. Distinctive characteristics of the testbed and vibration signals

The experimental test platform utilized in this study constituted a controlled yet representative model of an industrial Independent Cart System (ICS), specifically engineered to isolate vibration phenomena resulting from bearing–rail interaction under nonstationary motion. Unlike full-scale industrial ICS installations, which may involve mechanically coupled carts, robotic end-effectors, or payload-dependent loading conditions, the proposed testbed emphasizes single/multi-cart operation without external payloads. This design choice facilitates a systematic analysis of the vibration behavior while maintaining the fundamental kinematic and mechanical characteristics pertinent to bearing diagnostics.

Despite its simplified configuration, the test platform exhibited several distinctive features that set it apart from conventional rotating machinery test rigs. First, cart motion is inherently nonstationary and non-synchronous, involving repeated acceleration, deceleration, and direction reversal within a confined spatial region. Second, the bearings experience continuously varying contact forces as the cart transitions between straight and curved track segments, leading to time-varying load transfers at the bearing–rail interface. Third, propulsion is electromagnetic and contactless, whereas mechanical guidance relies entirely on rolling element bearings, resulting in vibration signatures that are strongly coupled to motion control dynamics rather than a steady rotational speed.

A representative example of the vibration response generated by this platform is shown in Fig. 12. Although the raw vibration signal alone appears highly irregular, the sliding variance reveals pronounced time-varying energy modulation that is synchronized with the cart speed profile and position along the track. Elevated variance levels consistently coincide with the phases of acceleration, deceleration, and direction reversal, whereas reduced variance is observed during the quasi-steady motion segments. This behavior demonstrates that nonstationarity in the measured signals is a structural characteristic of the test platform, rather than a transient disturbance or measurement artifact. These properties render the proposed testbed particularly suitable for investigating diagnostic methods under realistic non-stationary conditions while avoiding confounding effects introduced by payload variability or multi-degree-of-freedom robotic assemblies. Consequently, the platform provides a reproducible and interpretable environment for studying distributional changes in vibration features, which are central to the diagnostic challenges addressed in this study.

5.6. Fault injection procedure

To simulate bearing damage in a controlled, repeatable, and diagnostically significant manner, artificial faults were introduced using a precision picosecond laser ablation technique. This method was selected due to both the inherent characteristics of ICS and the specific requirements of data-driven diagnostic studies.

In industrial ICS applications, bearings function under highly nonstationary translational motion characterized by frequent speed variations and direction reversals. Consequently, natural run-to-failure experiments would necessitate prohibitively long testing durations and would result in fault geometries that are challenging to quantify, reproduce, or compare across experiments. Furthermore, intentionally inducing faults in operational industrial systems is economically impractical because of the significant costs associated with equipment downtime, component replacement, and production disruption. For newly deployed ICS platforms, historical fault data are typically unavailable, and the size and progression rates of real faults remain unknown. These combined constraints necessitate an artificial yet physically consistent fault injection strategy.

Initial evaluations of traditional artificial faulting techniques, such as drilling, revealed their unsuitability for the current application. The high hardness of the bearing materials and restricted geometry of the raceways result in irregular and poorly controlled defects, leading to weak or inconsistent vibration signatures. Conversely, ultrafast laser ablation offers precise control over fault geometry, width, and depth, irrespective of material hardness, while maintaining the integrity of the surrounding surface.

In this study, a picosecond laser system (EKSPLA Atlantic 50, 1064 nm) with a focal spot diameter of approximately 10 μm was utilized to engrave controlled grooves on the bearing raceways. A multipass ablation strategy was implemented to achieve the specified fault dimensions while accommodating the limited depth of field characteristic of ultrashort-pulse laser systems. This methodology facilitated the systematic creation of faults with nominal widths of 0.25, 0.5, 1.0, 1.5, and 2.0 mm, encompassing a spectrum from incipient to severe damage levels. Optical microscopy was employed to verify the resulting fault geometry and ensure dimensional consistency.

Crucially, this methodology ensures a high level of repeatability across various bearings and fault severities, which is vital for the comparative evaluation of diagnostic algorithms. Although laser-induced faults do not aim to replicate the full complexity of naturally occurring damage, they offer a controlled and interpretable approximation that facilitates a rigorous assessment of the sensitivity of algorithms to fault size under realistic nonstationary operating conditions. A comprehensive description of the laser ablation process, including the optical setup, depth-of-field considerations, and microscopic validation, is provided in [4] and the associated dataset documentation [34].

5.7. Dataset availability and reproducibility

Existing publicly available bearing fault datasets, such as those from Case Western Reserve University (CWRU), Intelligent Maintenance Systems (IMS), Politecnico di Torino, Xi'an Jiaotong University, and Changxing Sumyoung Technology Co., Ltd. (XJTU–SY), Southeast University (SEU), and the PHM challenge were primarily developed for conventional rotary systems operating under stationary or near-stationary conditions. These datasets typically assume fixed rotational kinematics, steady operating regimes, and weak couplings between machine motion and structural excitation. However, the vibration signals generated in the ICS significantly deviate from these assumptions. To address this gap, the present study utilizes the publicly released MOIRA–UNIMORE bearing dataset [34], which was specifically acquired to capture the signal characteristics induced by realistic ICS operation. This dataset includes synchronized vibration measurements, motion-related variables, and controlled fault annotations and is made openly available to support reproducibility and benchmarking under non-stationary industrial conditions.

6. Results and discussion

This section delineates the diagnostic outcomes derived from both single-cart and three-cart experimental configurations, as documented in the bearing dataset [34]. In the single-cart configuration, a cart traverses bidirectionally along a curved motor module at a nominal velocity of 1000 mm/s (Fig. 11). This reciprocal motion facilitates a distinct analysis of the vibration characteristics during forward and reverse traversals. Signal windows were independently extracted for each direction, and statistical features were computed for each windowed segment of the data. The corresponding speed profile is illustrated in Fig. 5, highlighting the acceleration, deceleration, and steady-state phases during both forward and reverse movements.

In the single-cart experiments, the dataset encompassed 14 channels, comprising nine vibration channels and five system variables pertinent to the cart motion, specifically, position, following error, actual velocity, velocity error, and set current. Vibration data were acquired using two tri-axial and three mono-axial accelerometers strategically positioned along the track to capture the dynamic interactions between the cart bearings and guide rail (see Fig. 11). Statistical features (enumerated in Table 5) were extracted from all the sensor channels, except for the cart position and speed, resulting in a high-dimensional-feature space. To address feature redundancy and enhance class separability, five feature ranking techniques were employed, followed by Principal Component Analysis (PCA) for dimensionality reduction and t-distributed stochastic neighbor embedding (t-SNE).

The results of the three-cart experiments are also presented. In this configuration, three carts spaced 250 mm apart move continuously around the full track in a closed-loop arrangement. The analysis concentrated on data windows extracted when the middle cart

Table 5
Summary of classical time-domain statistical descriptors. Closed-form expressions for features computed per analysis window.

Parameter	Expression
Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
RMS	$x_{\text{rms}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Std. deviation	$x_{\text{std}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (x_i-\mu)^2}$
Shape factor	$x_{\text{shape}} = \frac{x_{\text{rms}}}{\frac{1}{N}\sum_{i=1}^{N} \lvert x_i\rvert}$
Kurtosis	$x_{\text{kurt}} = \frac{\frac{1}{N}\sum_{i=1}^{N} (x_i-\mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N} (x_i-\mu)^2\right)^2}$
Skewness	$x_{\text{skew}} = \frac{\frac{1}{N}\sum_{i=1}^{N} (x_i-\mu)^3}{\left(\frac{1}{N}\sum_{i=1}^{N} (x_i-\mu)^2\right)^{3/2}}$
Peak value	$x_p = \max_i \lvert x_i\rvert$
Impulse factor	$x_{\text{IF}} = \frac{x_p}{\frac{1}{N}\sum_{i=1}^{N} \lvert x_i\rvert}$
Crest factor	$x_{\text{crest}} = \frac{x_p}{x_{\text{rms}}}$
Clearance factor	$x_{\text{clear}} = \frac{x_p}{\left(\frac{1}{N}\sum_{i=1}^{N} \sqrt{\lvert x_i\rvert}\right)^2}$
SNR (dB)	$\text{SNR} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right)$
THD (dB)	$\text{THD} = 10\log_{10}\left(\frac{\sum_{h=2}^{H} P_h}{P_1}\right)$
SINAD (dB)	$\text{SINAD} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}} + P_{\text{dist}}}\right)$

Notes: x_i: ith sample in a length-N window. x_p: peak magnitude. Spectral powers: P_1 (fundamental), P_h (hth harmonic, $h \geq 2$), H (highest harmonic index), $P_{\text{dist}} = \sum_{h=2}^{H} P_h$, and P_{noise} is the in-band noise excluding harmonic bins. The log factors use $10\log_{10}$ because they are power ratios.

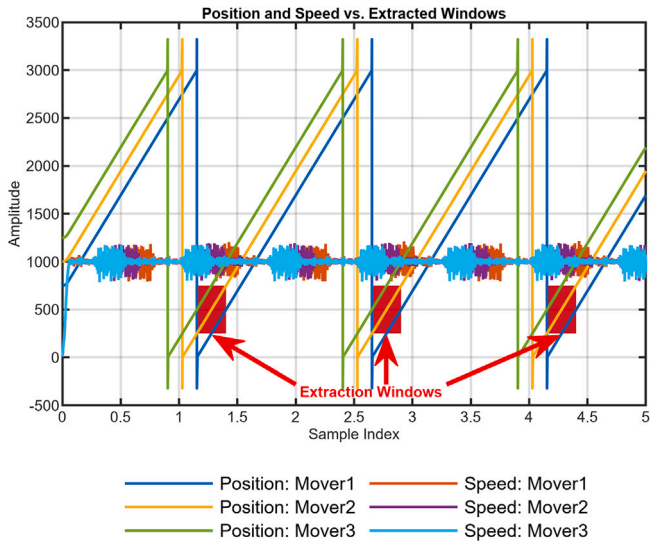

Fig. 13. The position and velocity trajectories of three carts moving continuously along the XTS track are examined. Extraction windows, highlighted in red, were specified for the middle cart when its absolute position ranged from 250 to 750 mm. The carts are uniformly spaced at intervals of 250 mm and operate in a cyclic motion pattern, facilitating consistent data segmentation across repetitive cycles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

containing a faulty bearing traversed the right-side curved module between absolute positions of 250 mm and 750 mm, as depicted by the shaded region in Fig. 13. While the vibration channels remained unchanged (nine channels), the system variables included cart-specific measurements for each of the three carts, resulting in 15 system variable channels. This enriched context increases the complexity of the dataset, reflecting the diagnostic challenges associated with multi-cart industrial deployment.

Owing to the interaction of all carts with a common guide rail, the recorded vibration signals inherently represent a superposition of excitations originating from multiple bearings and carts, rendering the complete isolation of individual bearing contributions unattainable in this setup. To mitigate the confounding effects, faults were introduced solely in the middle cart, and signal windows were extracted exclusively when this cart traversed the curved module equipped with a triaxial accelerometer. This approach ensures that the predominant fault-related excitation is attributable to the faulty bearing while acknowledging the presence of residual interference from healthy bearings and adjacent carts.

The primary aim of this study was to detect anomalies under realistic multi-source excitation, rather than to achieve precise fault localization or fault-type classification. The inclusion of interference reflects actual industrial conditions and presents a deliberate challenge to the proposed framework. Future research directions include explicit source separation, cart-level isolation, and fault localization. A detailed description of the experimental setup, configurations, and fault injection methodology is available in [34].

6.1. Fault classes and labeling

In this study, eight health conditions were examined, as detailed in Table 6. These conditions comprise one healthy class (*H*), four types of inner race faults (*FIR025*, *FIR05*, *FIR10*, and *FIR15*), and three types of outer race faults (*FOR025*, *FOR05*, and *FOR20*). Within this labeling framework, the prefix `F` signifies a faulty condition, whereas `IR` and `OR` denote the defect location on the inner or outer race, respectively. The numerical suffix indicates the notch width in millimeters (e.g., `025` = 0.25 mm, `05` = 0.5 mm). Although the initial inspection

Table 6
Class labels and fault specifications. Codes denote bearing defect type and notch width in millimeters; IR = inner race, OR = outer race.

Label	Location	Width (mm)	Description
H	–	–	Healthy (no induced defect)
FIR025	Inner race	0.25	Narrow inner-race notch; depth not specified
FIR05	Inner race	0.5	Inner-race notch; moderate width
FIR10	Inner race	1.0	Inner-race notch; increased width
FIR15	Inner race	1.5	Inner-race notch; severe width
FOR025	Outer race	0.25	Narrow outer-race notch; depth not specified
FOR05	Outer race	0.5	Outer-race notch; moderate width
FOR20	Outer race	2.0	Outer-race notch; severe width

Notes: Class codes adhere to the format `F{IR/OR}{width}`. Although the initial assessment of defect depth was conducted using optical microscopy, accurate quantification of the depth was not achievable. This limitation was due to the curvature of the deep-groove raceway and the nonplanar geometry of the defects, which led to optically unresolved and potentially nonuniform depth profiles. Consequently, the severity of the defects was characterized by the notch width, which was precisely controlled and consistently reproducible across the experiments.

of the defect depth was conducted using optical microscopy, the accurate quantification of the depth proved challenging. This difficulty arose because the laser-induced defects conformed to the curved deep-groove raceway, rendering them optically unresolved and potentially nonuniform along their length. The severity of the faults was evaluated based on the notch width, which was meticulously controlled during fabrication and offered a consistent metric for assessing the extent of the defect.

6.2. Rationale for hyperparameter selection

The proposed distribution-based transformation is characterized by a concise set of hyperparameters explicitly linked to the structural attributes of the feature distributions. These parameters do not serve as dataset-specific tuning variables; rather, they function as *distributional gates* that determine the conditions and methods for reshaping the data. The framework specifically utilizes two complementary decision-making mechanisms: a bimodality gate and a high-spread gate.

Bimodality gate and peak prominence threshold. The detection of bimodality is primarily governed by the minimum peak prominence parameter, $c_{\min}$, which is defined in relation to the maximum value of the estimated marginal density, as follows: By design, $c_{\min} \in (0, 1)$ specifies the minimum prominence necessary for a density peak to be deemed structurally significant. Interpreting $c_{\min}$ as a percentage of the maximum density ensures scale invariance across different features and data sets. The values of $c_{\min}$ ranging from approximately 10% to 20% of the maximum density are adequate for reliably identifying genuine bimodal structures while mitigating minor fluctuations attributable to sampling noise or finite sample effects. Selecting excessively low values of $c_{\min}$ increases the sensitivity to spurious local maxima, resulting in false bimodality detection and unwarranted clustering.

Upon detection of bimodality, the transformation parameters within the framework are conditionally activated. The shift coefficient γ regulates the maximum fraction of inter-cluster separation that can be applied during cluster realignment. To prevent excessive displacement, the effective shift is adaptively constrained as $\min\left(\gamma,\ \gamma\frac{\Delta\mu}{\sigma}\right)$, where $\Delta\mu$ represents the distance between cluster centers and σ the feature standard deviation. This normalization ensures that the intensity of shifting is governed by relative rather than absolute separation, rendering γ applicable across datasets. The subsequent scaling is controlled by the parameters δ, ϵ, and η, which collectively define a bounded contraction range. The scale factor is calculated as $\max\left(\delta,\ \epsilon - \eta\frac{\Delta\mu}{\sigma}\right)$, ensuring that contraction strength decreases smoothly as cluster separation increases, while remaining strictly lower-bounded by δ to prevent distribution

collapse. Typical values of $\delta \in [0.01, 0.2]$, $\epsilon \in [0.2, 0.4]$, and $\eta \in [0.3, 0.6]$ have been found to provide stable reshaping across the feature types. Notably, these parameters regulate the extent of reshaping once bimodality is confirmed but do not influence the detection decision itself.

Unimodal high-spread gate and dispersion threshold. For features that do not display bimodality, the transformation is selectively applied via a unimodal high-spread gate, which is regulated by the dispersion threshold σ_t. This parameter determines whether a unimodal feature exhibits excessive spread relative to its mean, indicating poor compactness despite the absence of multiple modes in the distribution. Moderate values of σ_t, typically ranging from 0.2 to 0.5 after normalization, facilitate the effective contraction of overly dispersed features without affecting naturally compact distributions. Selecting excessively large values of σ_t suppresses the activation of this gate, resulting in high-variance unimodal features remaining unmodified. Values exceeding unity are not meaningful in this context, as they would correspond to unrealistically large dispersion levels and would effectively disable the corrective mechanism. Once activated, the contraction strength is governed by the parameter ξ, which scales deviations from the mean according to $x \leftarrow \mu + \xi(x - \mu)$. Values of ξ in the range of approximately 0.1 to 0.3 have been found to reduce excessive variance while preserving the relative feature ordering and informative spread of the data. Smaller values risk over-compression and loss of discriminative information, whereas values approaching unity yield negligible transformations. Overall, the hyperparameters were structured to operate on normalized, dimensionless quantities and were activated conditionally through explicit gating mechanisms. This design ensures that the transformation adapts to the distributional structure rather than the dataset-specific characteristics, thereby supporting stable behavior and transferability across fault types, sensor configurations, and operating conditions.

6.3. Transferability and stability of hyperparameter

Although the hyperparameters of the proposed transformation were selected based on distributional considerations, their functionality is not contingent upon dataset-specific characteristics such as fault type, sensor placement, or operating regime. Instead, each parameter governs a distinct structural property of the feature distributions, namely modality, dispersion, or scaling intensity, rendering the transformation inherently data-agnostic. The minimum prominence threshold $c_{\min}$ serves as a relative criterion for normalized kernel density estimates, thereby adapting naturally across datasets with varying feature magnitudes or sample sizes. Its primary function is to suppress spurious modes rather than finely tune the cluster boundaries, ensuring that the effective operating range remains stable across applications. Similarly, the standard deviation threshold σ_t is activated only for unimodal features exhibiting excessive spread, ensuring that compact distributions remain unaltered while highly dispersed features are selectively reshaped. Notably, the transformation employs gated activation mechanisms: bimodal reshaping is applied only when statistically meaningful secondary modes are detected, whereas the compression of the unimodal variance is triggered only when the dispersion exceeds a predefined threshold. This conditional structure prevents unnecessary modifications of well-behaved features and limits sensitivity to hyperparameter variations. In practice, only minor adjustments within the recommended parameter ranges are required when transferring the framework to datasets with substantially different noise characteristics or feature normalization schemes. Such adjustments correspond to coarse calibration rather than dataset-specific retuning and do not alter the underlying transformation. Consequently, the selected hyperparameter values serve as conservative defaults that regulate distributional behavior rather than optimize dataset-specific performance. Observations across multiple datasets indicate that moderate perturbations around the reported values do not lead to qualitative changes in the transformed feature space, thereby supporting the robustness and transferability of the proposed framework.

6.4. Hyperparameter sensitivity analysis

A one-at-a-time sensitivity analysis was conducted to assess the sensitivity of the hyperparameter values. Each transformation parameter was varied over a broad interval (0.01–0.9), while all other parameters were maintained at their nominal values, and the entire detection pipeline was re-evaluated using the F1-score as the performance metric. The analysis was performed on two distinct feature families, classical statistical descriptors (Table 5) and Hjorth parameters (Table 8), to determine whether the parameter sensitivity was consistent across heterogeneous feature representations. In the reported sensitivity plots, the horizontal axis represents the varied parameter, the vertical axis indicates the mean F1-score, and each curve corresponds to a different baseline regime defined by either the unimodal dispersion threshold σ_t or the unimodal contraction factor ξ.

Across both feature families, most parameters exhibited wide stability regions (Figs. 14 and 15), where performance varied only marginally. Notably, the parameters associated with the bimodality gate ($c_{\min}$, γ, δ, ϵ, and η) exhibit relatively flat response curves over broad intervals, with performance degradation occurring only at the extreme values. This behavior suggests that these parameters are not sharply tuned and that moderate deviations from the nominal values do not result in qualitative changes in the detection performance.

In contrast, the sensitivity curves for σ_t and ξ exhibit more pronounced trends, reflecting their direct influence on the reshaping of the unimodal variance. For σ_t, the plots reveal a transition beyond which the performance deteriorates rapidly, corresponding to the suppression of the unimodal variance compression when the threshold becomes 0excessively large. Similarly, increasing ξ toward unity results in a monotonic reduction in performance because the contraction effect becomes negligible. These trends explain why σ_t and ξ are treated as regime-defining parameters in the analysis and motivate the selection of

Table 7
Hyperparameters for the adaptive distribution-based transformation are applied directly within the original feature space, foregoing any dimensionality reduction.

Parameter	$c_{\min}$	γ	δ	ϵ	η	σ_t	ξ
Value	0.1	0.95	0.05	0.3	0.5	0.3	0.1

Notes: $c_{\min}$ is the minimum prominence threshold coefficient for the KDE peak detection. γ and α control the shift magnitude of bimodal distributions. δ, ϵ, and η regulate the adaptive scaling of the separated clusters. σ_t is the standard deviation threshold for identifying dispersed and unimodal features. ξ governs the variance compression strength of the non-bimodal distributions.

Table 8
Hjorth parameters. Closed-form expressions for activity, mobility, and complexity computed per analysis window.

Activity	Mobility	Complexity
$\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$	$\sqrt{\frac{\mathrm{Var}(\dot{x})}{\mathrm{Var}(x)}}$	$\sqrt{\frac{\mathrm{Var}(\ddot{x})}{\mathrm{Var}(\dot{x})}} \Big/ \sqrt{\frac{\mathrm{Var}(\dot{x})}{\mathrm{Var}(x)}}$

Notes: x_i is the ith sample in a length-N window and $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$. The first and second discrete differences are $\dot{x}_i := x_i - x_{i-1}$ for $i = 2, \ldots, N$ and $\ddot{x}_i := \dot{x}_i - \dot{x}_{i-1}$ for $i = 3, \ldots, N$. $\mathrm{Var}(\cdot)$ denotes the sample variance over the corresponding sequence (e.g., $\mathrm{Var}(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$, $\mathrm{Var}(\dot{x}) = \frac{1}{N-1}\sum_{i=2}^{N}(\dot{x}_i-\overline{\dot{x}})^2$, $\mathrm{Var}(\ddot{x}) = \frac{1}{N-2}\sum_{i=3}^{N}(\ddot{x}_i-\overline{\ddot{x}})^2$). The mean $\overline{\dot{x}}$ and $\overline{\ddot{x}}$ values are obtained from the respective sequences.

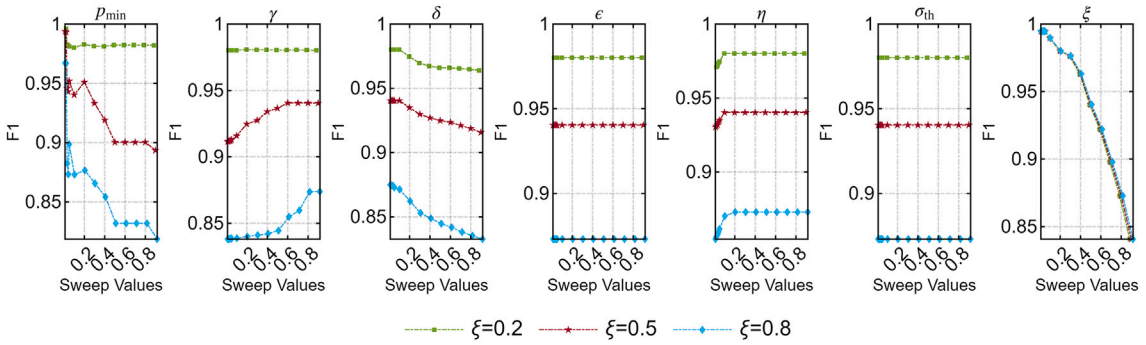

Fig. 14. Sensitivity analysis of adaptive transformation hyperparameters on **classical statistical descriptors** is conducted. Each panel examines the variation of a single parameter while maintaining the others at their nominal values. The curves represent different baseline regimes of the unimodal-compression gate, with performance evaluated using the mean F1-score.

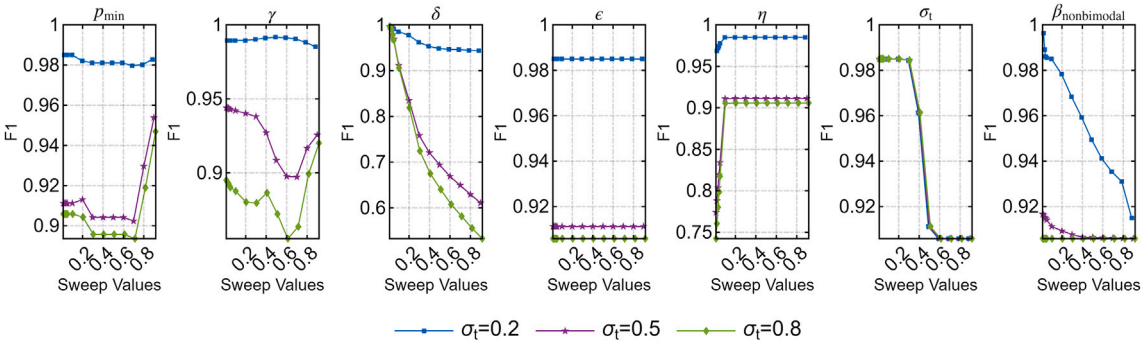

Fig. 15. The sensitivity analysis of adaptive transformation hyperparameters on **Hjorth parameters** is presented. Each panel examines the variation of a single parameter while maintaining the constancy of the others. The findings indicate extensive stability regions and affirm the function of gating parameters as activation thresholds, rather than variables requiring dataset-specific tuning.

moderate values that balance the variance reduction against information preservation.

Overall, the sensitivity analysis confirmed that the hyperparameter values adopted in Table 7 lie well within the stable operating regions for both feature families. The observed trends demonstrate that the proposed transformation is robust to moderate parameter perturbations and that performance degradation occurs only under extreme, practically irrelevant settings. These results provide empirical support for fixing the hyperparameters in all experiments.

6.5. Additive noise categories

In practical industrial environments, vibration measurements are inevitably influenced by various noise sources arising from environmental conditions and sensor limitations. To assess the robustness of the proposed feature transformation framework, we introduced three categories of synthetic noise into the vibration channels while maintaining the integrity of the system variables, such as the control inputs and the encoder feedback. Following noise injection, the entire pipeline, which comprised feature transformation, reduction, and clustering, was executed again to evaluate any performance degradation. The three types of noise were designed to represent the typical challenges encountered in condition monitoring scenarios.

6.5.1. Broadband measurement and quantization noise

To simulate ambient sensor noise and digitization errors, each vibration channel was perturbed using zero-mean Additive White Gaussian Noise (AWGN) at a target signal-to-noise ratio (SNR) of 10 dB. A small uniformly distributed jitter was added to emulate quantization effects:

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{n}_{\mathrm{AWGN}} + \mathbf{n}_{\mathrm{UNI}}, \quad \mathbf{n}_{\mathrm{AWGN}} \sim \mathcal{N}(0, \sigma^2), \quad \mathbf{n}_{\mathrm{UNI}} \sim \mathcal{U}(-a, a),$$

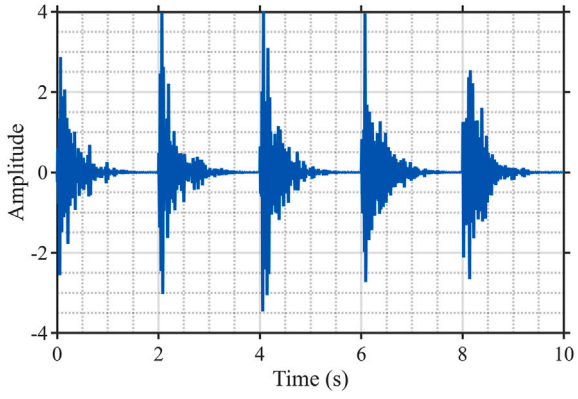

Fig. 16. Example of the amplitude-modulated narrowband interference used in Category 2 (f_s=50 kHz, band-pass [20, 40] Hz, pulse repetition f_p=0.5 Hz, exponential decay δ). The decaying bursts mimic intermittent tonal contamination that is aligned with the mechanical cycle. The vertical axis represents the signal amplitude in arbitrary units (a.u.), as the signal is synthetically generated and normalized for controlled noise injection rather than representing a physically measured vibration level.

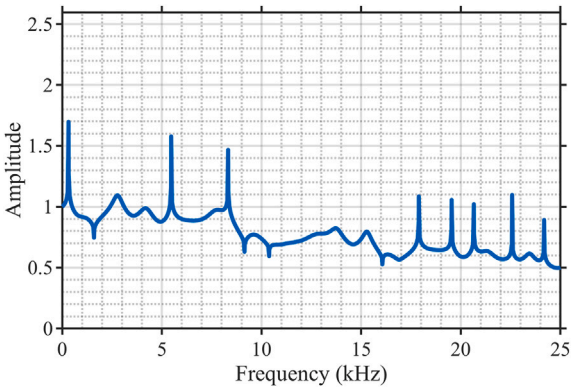

Fig. 17. Pseudo-FRF profile (Category 3) generated using N_p=20, peak_{RES}=2, $\text{peak}_{\text{ARES}}$=0.75, and ADR=0.5. The vertical axis denotes the relative magnitude in arbitrary units (a.u.), which defines the target spectral envelope used for noise synthesis via inverse FFT.

where $a = \lambda \cdot \text{RMS}(\mathbf{x})$ and $\lambda = 0.02$. The SNR was computed on a per-channel and per-window basis.

6.5.2. Narrowband intermittent interference

To simulate periodic tonal interference arising from mechanical resonances or rotating machinery, we employed band-limited pulses generated through amplitude-modulated noise, which was shaped by a Butterworth band-pass filter:

$$n_{\text{NB}}(t) = A\, e^{\kappa t}\left(w(t) * h_{[f_L, f_H]}(t)\right),$$

where $w(t)$ represents white noise, $h_{[f_L,f_H]}(t)$ denotes a 4th-order Butterworth filter with cutoff frequencies $f_L = 20\,\text{Hz}$ and $f_H = 40\,\text{Hz}$, A is the amplitude scale, and $\kappa < 0$ dictates the exponential decay. The pulses recur every $T = 2\text{s}$ with a sampling rate of $f_s = 50$ Hz.

6.5.3. Synthetic structural noise via pseudo-FRF shaping

To simulate broadband excitation shaped by mechanical resonance profiles, a pseudo-frequency response function (PFRF) was generated, comprising sinusoidal peaks with randomized phases:

$$|H[r]| \propto \text{PFRF}[r] \cdot (\nu_1[r]\,\nu_2[r] + \zeta), \quad \angle H[r] \sim \mathcal{U}[0, 2\pi),$$

where $\nu_1[r], \nu_2[r] \sim \mathcal{U}(0, 1)$ introduce fine amplitude variation, and $\zeta > 0$ serves as a bias to ensure a minimum magnitude. Hermitian symmetry guarantees a real-valued output, which is obtained as

$$\text{SN}(t) = \Re\{\text{IFFT}(H)\},$$

followed by peak normalization. A total of $N_s = 20$ sinusoidal terms were employed, with peak magnitudes ranging from $[0.75, 2.0]$ and a decay rate of $\chi = 0.5$.

The numerical values selected for the synthetic noise parameters were determined based on the kinematic and mechanical characteristics of the ICS platform, rather than through arbitrary tuning. Specifically, the AWGN level represents a conservative lower-bound signal-to-noise ratio observed during high-speed cart motion. The narrowband interference band corresponds to the fundamental rolling frequencies of the cart bearings at the nominal speed, and the pseudo-FRF parameters emulate the resonance-shaped excitation induced by unknown rail–cart transfer paths. These selections ensure physical relevance while facilitating a controlled robustness assessment (Figs. 16 and 17).

6.6. Results without additive noise

6.6.1. Effect of the transformation in PCA space

Figs. 18 and 19 illustrate the class distributions prior to (top rows) and after (bottom rows) the proposed transformation across the five ranking strategies (OWA, KW, LS, Var, and Mono). In the *pre-transform* maps, the clusters are elongated and exhibit significant overlap, particularly among the healthy (H) and small-fault classes (FIR025 and FOR025). After transformation, the clusters become compact and well separated, indicating density regularization and enhanced inter-class margins. This trend persisted when transitioning from the single-cart

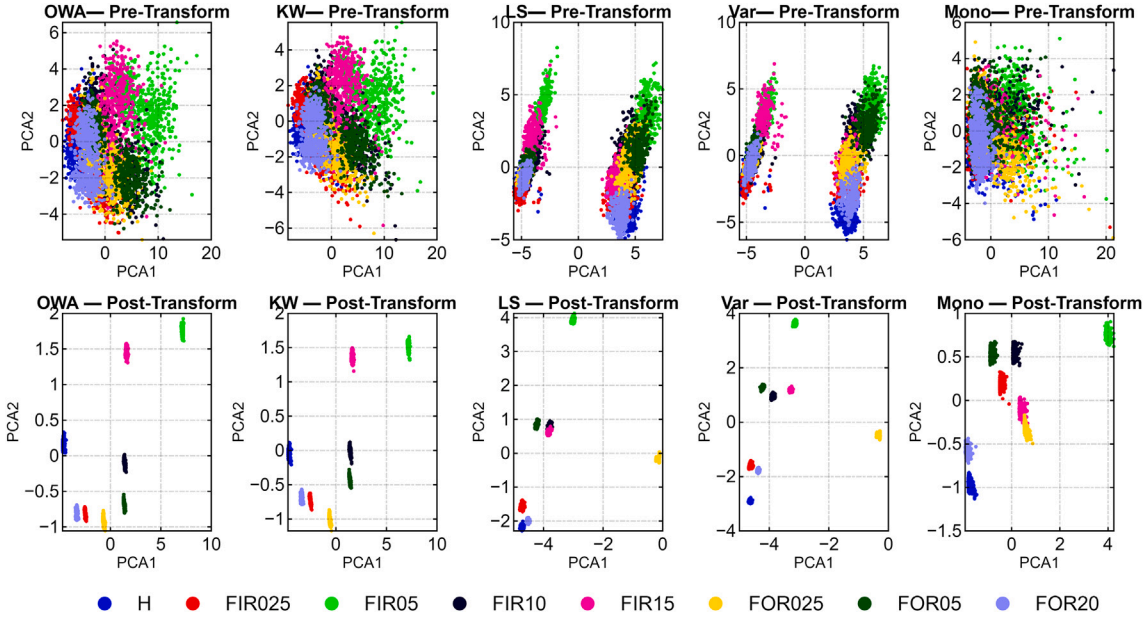

Fig. 18. PCA embeddings (**single-cart**): The pre-transformation results (top row) display overlapping, elongated clusters, whereas the post-transformation results (bottom row) exhibit compact and distinct class clusters across all five ranking strategies.

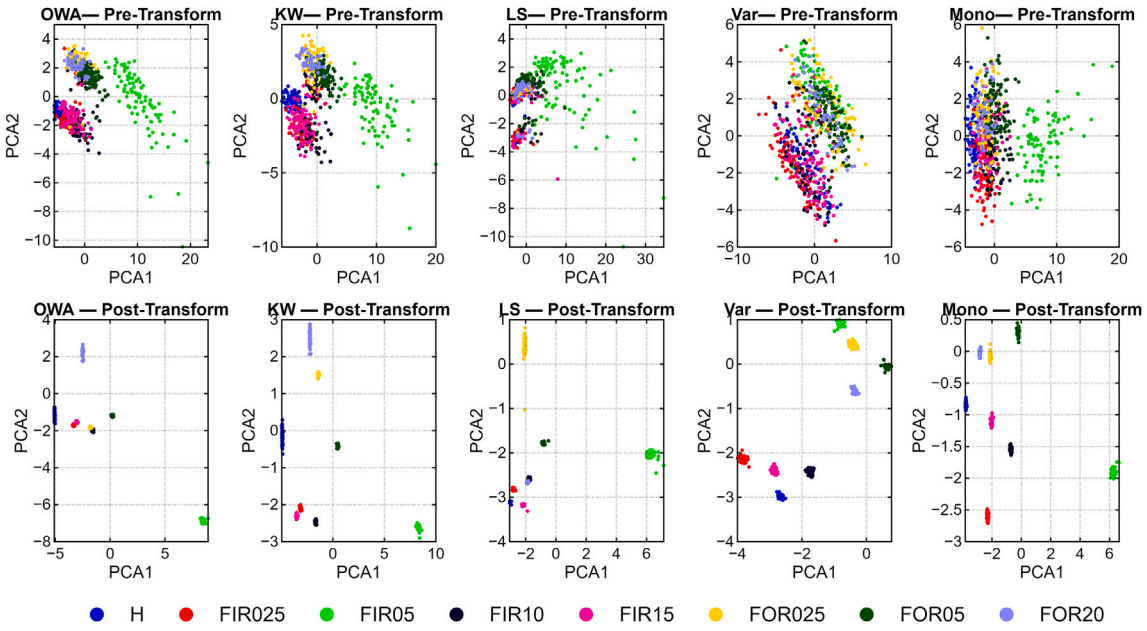

Fig. 19. The proposed transformation of PCA embeddings (**three-cart**) results in compact, non-overlapping clusters, even with increased system complexity, thereby confirming its robustness.

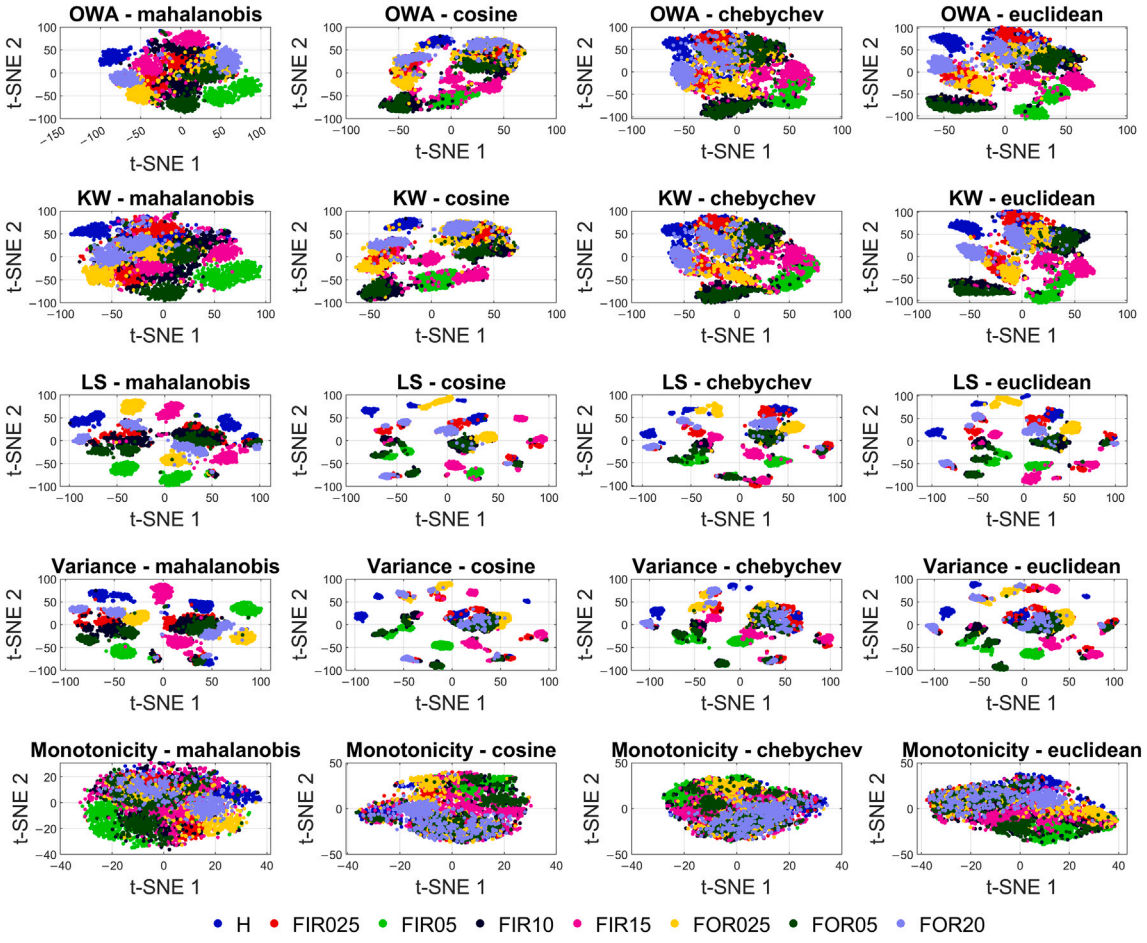

Fig. 20. T–SNE (**single-cart**, pre-transform) exhibits scattered and overlapping clusters, with a noticeable imbalance in density across various metrics, including Mahalanobis, Cosine, Chebychev, and Euclidean.

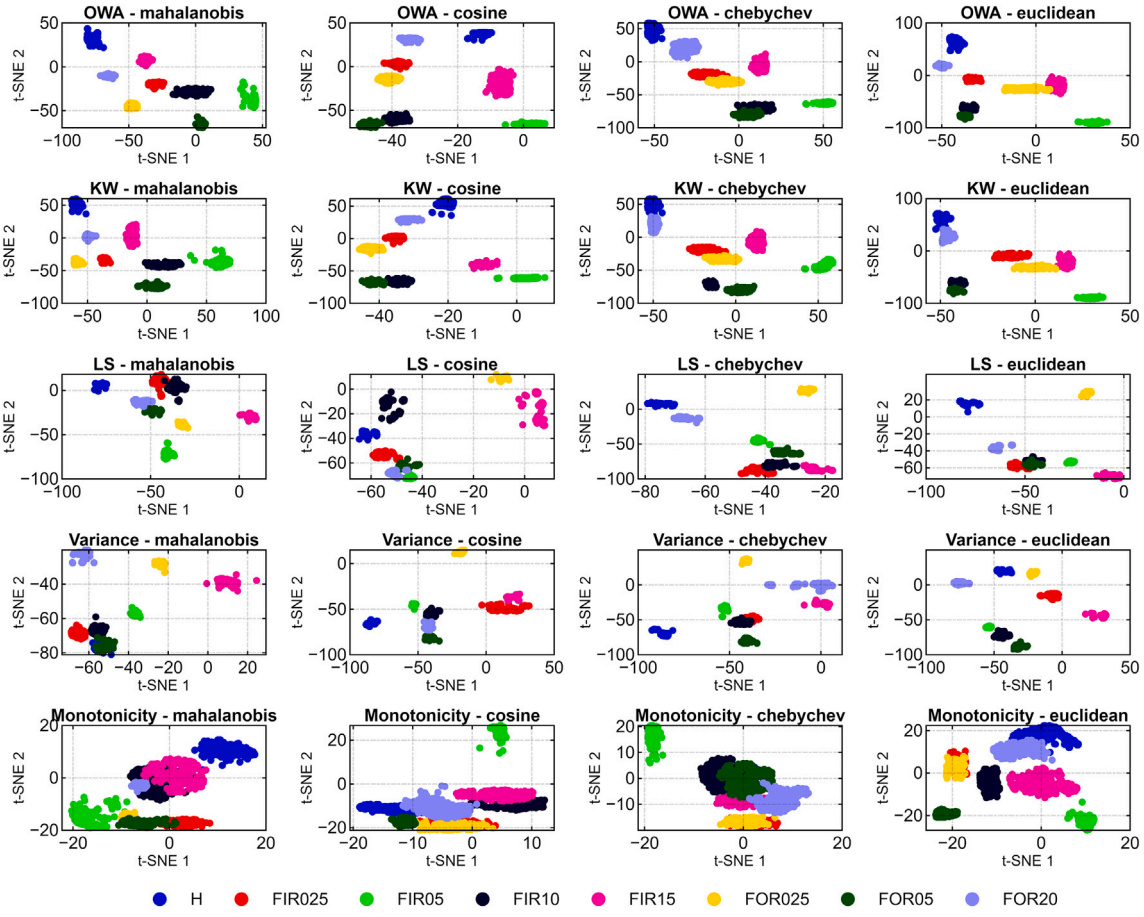

Fig. 21. t–SNE (**single-cart**, post-transform): The method yields compact and well-separated clusters with consistent spacing across all distance metrics, demonstrating effective density regularization by the proposed approach.

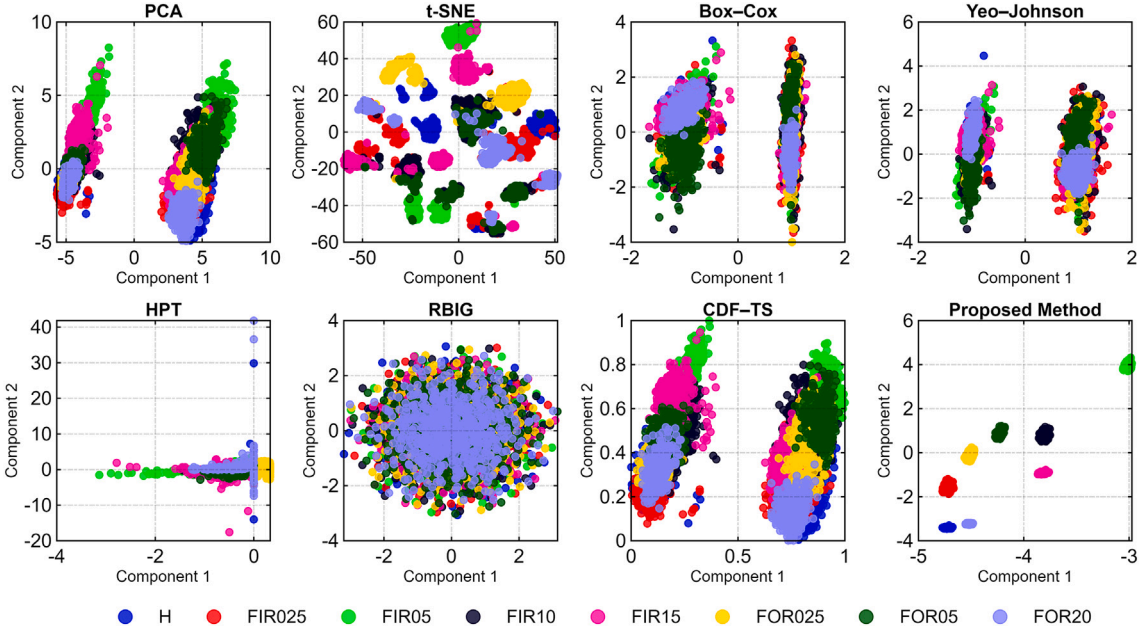

Fig. 22. Comparative 2D embeddings across methods (single-cart experiment). Top row: PCA and t–SNE baselines; Box–Cox and Yeo–Johnson marginal transforms. Bottom row: HPT and RBIG density transforms, CDF–TS, and the proposed method. Each color denotes a class (H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, FOR20). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

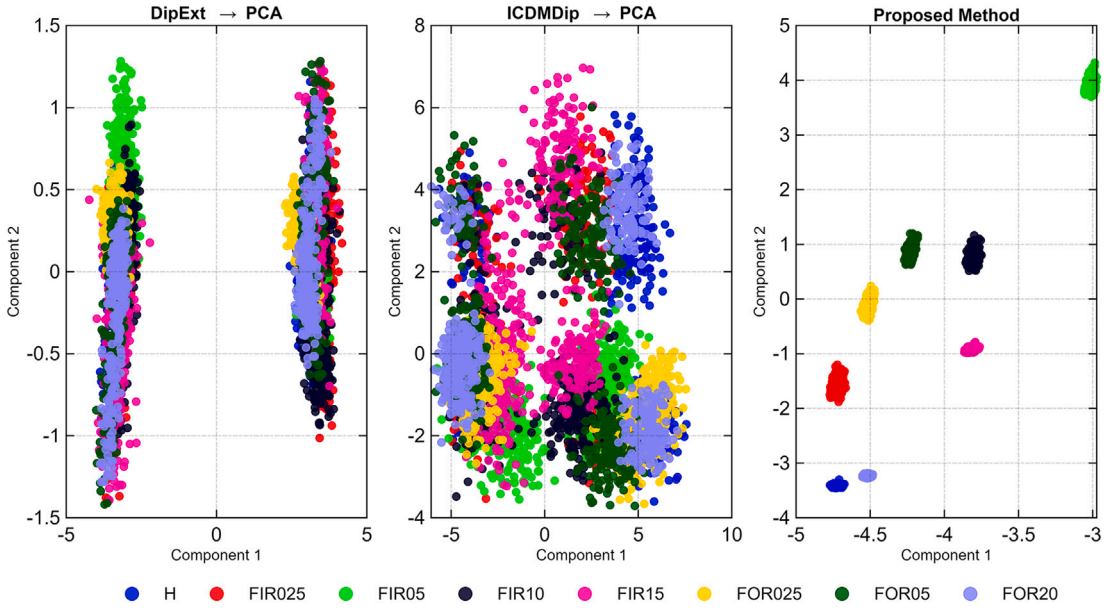

Fig. 23. Embedding comparison (single-cart): (left) DipExt projected by PCA, (middle) ICDMDip projected by PCA, and (right) proposed method (native 2D image). The proposed method yields compact, well-separated clusters for all classes.

to the more complex three-cart setup, demonstrating robustness against increased coupling.

6.6.2. Effect of the transformation in t–SNE space

Figs. 20 and 21 illustrate the t–SNE embeddings generated using four distance metrics: Mahalanobis, Cosine, Chebychev, and Euclidean. Prior to transformation, the feature points exhibited scattered and overlapping clusters with uneven densities. Small-fault cases are challenging to distinguish because their samples are intermingled with other classes. Following the application of the proposed transformation, each class forms a compact and distinctly separated cluster. The spacing between classes becomes larger and more uniform across all metrics. This balanced distribution of points facilitates the reliable identification of faults by anomaly detection models such as the OCSVM and Isolation Forest.

In both embedding families, PCA and t–SNE, the transformation effectively converts skewed and overlapping distributions into compact and well-separated clusters while maintaining the order of the classes.

6.6.3. Comparative analysis

Fig. 22 presents a comparative analysis of various transformation methodologies for feature embedding within the single cart dataset. Traditional embedding techniques, such as PCA and t–SNE, which perform linear variance maximization and nonlinear neighborhood-preserving projection, respectively, exhibit elongated overlapping clusters. In these embeddings, minor faults (FIR025 and FOR025) are interspersed with healthy samples, indicating uneven data densities and inadequate discrimination between closely related operating states.

Marginal power-based transformations, including Box–Cox and Yeo–Johnson, aim to reduce skewness and stabilize variance through monotonic nonlinear mappings applied independently to each feature. Although these methods improve marginal regularity, they do not substantially enhance global separability. The orientation and shape of the clusters remain largely unchanged, and inter-class mixing persists, highlighting the limitation of purely marginal normalization in reshaping multivariate feature geometry. Gaussianization-based methods, such as the Hyperbolic Power Transform (HPT) and Rotation-Based Iterative Gaussianization (RBIG), further enforce statistical regularity by mapping data toward Gaussian marginals, with RBIG additionally applying orthonormal rotations to decorrelate dimensions. However, HPT excessively compresses one dimension, whereas RBIG produces an almost isotropic scatter that obscures the underlying fault structures. Although these methods achieve distributional homogenization, they fail to preserve the class geometry that is essential for diagnostic interpretation. The density-homogenizing transformation CDF–TS, which equalizes local point densities through cumulative distribution-based scaling, provides more uniform neighborhood densities; however, it hardly achieves class separation. Fig. 23 further compares the two multimodality-aware transformations, DipExt and ICDMDip, projected into the PCA domain alongside the proposed method. DipExt, which enhances bimodality by amplifying dip statistics in marginal distributions, yields vertically elongated structures with limited differences between fault levels. ICDMDip, which applies iterative clustering and dip-based density modification, generates scattered and partially overlapping clouds that obscure the class boundaries. In contrast, the proposed transformation produces compact and well-separated clusters. By jointly regulating the local density and global geometry it yields stable and diagnostically meaningful feature embeddings under variable operating conditions.

The proposed transformation results in compact clusters with consistent interclass spacing. This balanced reshaping of the feature space mitigates local density imbalances, enhances separability, and facilitates downstream anomaly detection and classification by providing clear and uniform decision regions.

From a computational standpoint, Table 9 presents the asymptotic time complexities associated with the evaluated distribution-shaping transformations. While iterative Gaussianization and density-equalization methods demonstrate higher-order scaling concerning the number of samples or feature dimensions, the proposed method exhibits approximately linear complexity in both N and D. This computational efficiency arises from the transformation's independent operation on each feature, utilizing fixed-resolution kernel density estimation and a limited number of clustering iterations, thereby circumventing pairwise distance computations and iterative global density redistribution.

6.6.4. Quantitative performance analysis

To quantitatively assess the enhancements in cluster compactness and separability, Table 10 presents the F1-scores obtained using one-class SVM (OCSVM) and Isolation Forest (IF), both trained exclusively on healthy (H) samples, across all transformation methods and feature

Table 9
Asymptotic computational complexity of distribution-shaping transforms compared in this study.

Method	Time complexity
Box–Cox	$O(N\,D)$
Yeo–Johnson	$O(N\,D)$
HPT	$O(N\,D)$
RBIG	$O\big(L\,(D\,N\log N + N\,D^2) + L\,D^3\big)$
CDF–TS	$O(T\,N^2\,D)$
DipExt	$O(N\log N) + O(D\log D)$
ICDMDip	$O(\ell\,D^2\,N\log N)$
Proposed Method	$O\big(D\,N\,G + D\,R\,I\,N\big) \approx O(N\,D)$

Notes: N: number of samples; D: feature dimension; G: KDE grid size (evaluation points); R: k-means replicates; L: RBIG layers; I: k-means iterations; T: CDF–TS iterations; ℓ: ICDMDip iterations.

ranking schemes under PCA- and t-SNE-reduced embeddings. Fault samples were used solely during testing, adhering to a strict one-class anomaly detection framework. The table compares the performance across five ranking criteria: one-way ANOVA (OWA), Kruskal–Wallis (KW), Laplacian Score (LS), variance (Var), and monotonicity index (Mono).

Across both embeddings, the proposed method consistently achieved the highest F1-scores for all ranking strategies and detectors, corroborating the visual separation observed in the corresponding low-dimensional projections. Under PCA, the F1-scores exceeded 0.97 for OCSVM and approached unity for IF, indicating a highly compact and density-balanced representation of the healthy class. A similar trend was observed for t–SNE, where the proposed transformation yielded near-perfect separability across all evaluated configurations.

In contrast, marginal transforms (Box–Cox and Yeo–Johnson) provide only limited improvement, with F1-scores rarely exceeding 0.7. Gaussianization-based methods (HPT and RBIG) exhibited unstable or degraded performance, which was attributable to distortions in the local geometric structure. The density-homogenizing CDF–TS transform yields moderate gains but does not achieve comparable compactness. Multimodality-aware methods (DipExt and ICDMDip) offer partial improvement but remain inferior to the proposed approach, whose adaptive distribution reshaping preserves locality and interclass continuity. Overall, the quantitative evaluation reinforces the geometric interpretation: the proposed method produces compact, well-separated, and uniformly distributed clusters, leading to substantial gains in fault discrimination under both linear and nonlinear embeddings.

6.7. Robustness under noisy conditions

To evaluate the robustness against environmental and instrumentation noise, three representative noise families were exclusively introduced into the vibration channels: (i) additive white Gaussian noise combined with uniform background noise (AWGN + U), (ii) narrowband interference (NB), and (iii) sinusoidal structural noise (SN). The drive and command signals remained unaltered to isolate the effect of the vibration corruption. All noise sources were introduced directly into the original continuous vibration signals *prior* to the window segmentation and feature extraction. Consequently, a KDE-based distribution estimation was conducted on window-level features, the statistics of which inherently reflect noise-induced variability, thereby ensuring a physically consistent robustness evaluation of the entire processing pipeline.

In all cases, PCA projections of the seven classes—Healthy (H), FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20—are shown before and after transformation. In the *pre-transform* plots, the presence of noise results in significant overlap, shape distortion, and loss of separability between fault categories. The proposed Adaptive Distribution-Based Transformation (IWCCT) restores compact, linearly separable

Table 10
F1 scores of different transformations and feature ranking methods under PCA and t–SNE embeddings using OCSVM and IF.

Embedding	Transformation	Detector	Classical descriptors				
			OWA	KW	LS	Var	Mono
PCA	Original	OCSVM	0.82	0.77	0.73	0.81	0.50
		IF	0.81	0.77	0.78	0.83	0.52
	Box–Cox	OCSVM	0.07	0.07	0.46	0.61	0.08
		IF	0.10	0.09	0.54	0.68	0.13
	Yeo–Johnson	OCSVM	0.07	0.08	0.48	0.60	0.10
		IF	0.11	0.10	0.57	0.70	0.15
	HPT	OCSVM	0.14	0.11	0.56	0.28	0.12
		IF	0.15	0.14	0.63	0.47	0.23
	RBIG	OCSVM	0.09	0.10	0.12	0.09	0.08
		IF	0.11	0.09	0.09	0.10	0.12
	CDF–TS	OCSVM	0.83	0.80	0.76	0.82	0.58
		IF	0.82	0.79	0.79	0.84	0.60
	Proposed Method	OCSVM	0.88	0.49	0.63	0.92	0.96
		IF	0.99	0.99	0.99	0.99	0.94
t-SNE	Original	OCSVM	0.90	0.88	0.58	0.84	0.69
		IF	0.90	0.87	0.99	0.98	0.71
	Box–Cox	OCSVM	0.33	0.45	0.33	0.58	0.31
		IF	0.41	0.62	0.87	0.90	0.33
	Yeo–Johnson	OCSVM	0.27	0.40	0.46	0.61	0.28
		IF	0.27	0.56	0.85	0.91	0.29
	HPT	OCSVM	0.38	0.03	0.14	0.56	0.07
		IF	0.47	0.11	0.22	0.92	0.09
	RBIG	OCSVM	0.10	0.10	0.09	0.10	0.08
		IF	0.11	0.11	0.06	0.07	0.10
	CDF–TS	OCSVM	0.87	0.76	0.87	0.77	0.71
		IF	0.89	0.86	0.99	0.99	0.73
	Proposed Method	OCSVM	0.99	0.99	0.99	0.99	0.91
		IF	0.99	0.99	0.99	0.99	0.99
	ICDMDip	OCSVM	0.89	0.88	0.54	0.67	0.62
		IF	0.89	0.86	0.78	0.81	0.52
	DipExt	OCSVM	0.86	0.87	0.38	0.33	0.23
		IF	0.86	0.83	0.77	0.77	0.20

Notes: OWA, one-way ANOVA; KW, Kruskal–Wallis; LS, Laplacian Score; Var, variance; Mono, monotonicity index. Detectors: OCSVM = one-class SVM; IF = Isolation Forest. Embeddings: PCA = Principal Component Analysis; t-SNE = t-distributed Stochastic Neighbor Embedding. The values are the F1 scores (0–1).

structures across all noise conditions, confirming its robustness to both stochastic and deterministic disturbances.

Broadband noise (Fig. 24). The combination of AWGN and uniform jitter increases both the angular and radial spreads, causing the healthy and mild-fault classes to merge. Following the proposed transformation, all classes consolidate into compact, well-separated clusters with uniform spacing and reduced intra-class variance. The method effectively suppresses stochastic noise while maintaining the intrinsic geometry of the fault manifolds, demonstrating its resilience to the additive random perturbations that are common in industrial measurements.

Narrowband interference (Fig. 25). The injected narrowband amplitude-modulated disturbance introduced anisotropic deformation and ring structures, particularly in the LS- and Var-based features. These coherent distortions expand the intra-class variance and obscure fault boundaries. After the transformation, the rings were eliminated, and each fault class reverted to a circular, compact form. A slight residual overlap appeared only in the transition regions between the mild and moderate faults. The transformation adaptively suppresses structured and frequency-localized noise, revealing stable manifolds in the principal component analysis (PCA) domain.

Sinusoidal structural noise (Fig. 26). This represents the most challenging case, indicative of structural resonance-type contamination. The

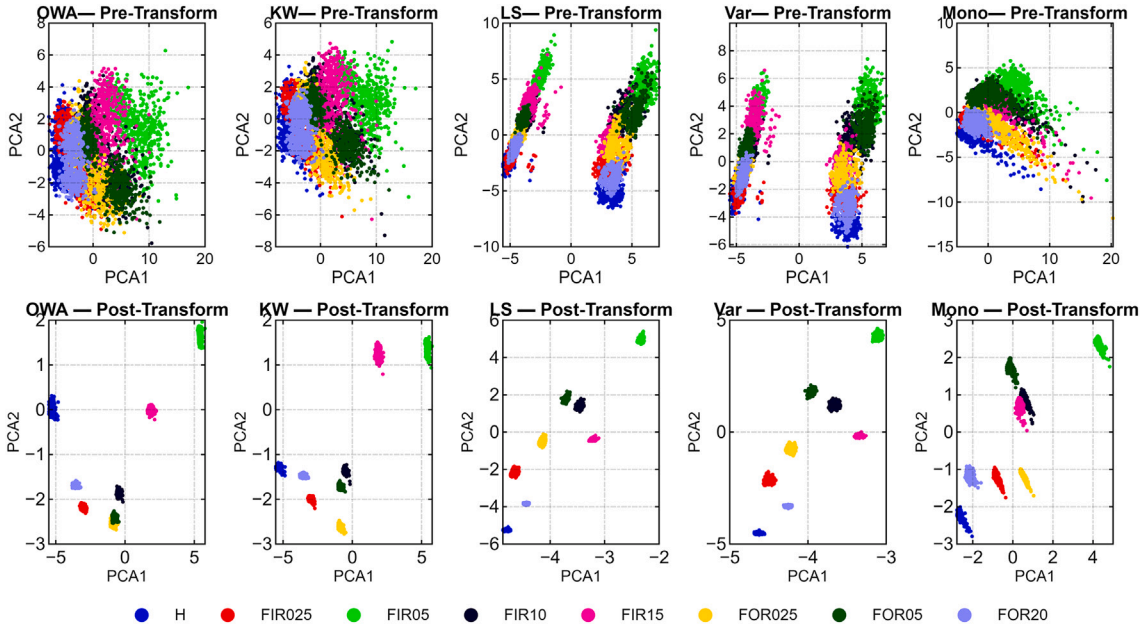

Fig. 24. The impact of broadband noise, specifically AWGN at 10 dB combined with 2% uniform jitter, on polar diagnostics is illustrated before (top row) and after (bottom row) the application of IWCCT. The columns represent OWA, KW, LS, Variance, and Monotonicity. Prior to transformation, noise increases the angular and radial dispersion; however, the proposed transformation effectively consolidates each class into a compact cluster, thereby restoring class separability.

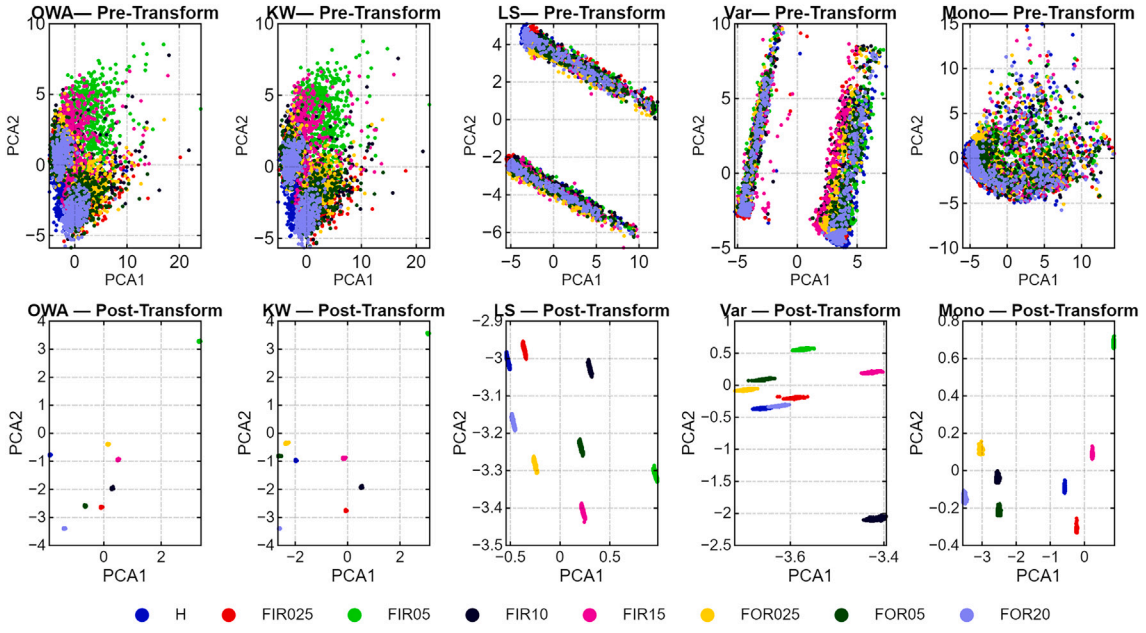

Fig. 25. The effect of narrowband, amplitude-modulated interference (band-pass $[20, 40]$ Hz, pulse rate f_p=0.5 Hz) was examined. Prior to the application of the proposed transformation (top row), anisotropic smearing and ring-like structures were observed in these images. However, following the implementation of the proposed transformation (bottom row), the clusters became compact with minimal leakage, demonstrating robustness against coherent tonal noise.

pseudo-FRF noise produces curved and entangled manifolds that obscure class boundaries. After the proposed transformation, each fault class formed a compact cluster with a distinct centroid and controlled variance. Although the cluster radii were slightly larger than those in the broadband case, the discriminative geometry was fully retained, confirming the robustness of the method against deterministic-resonance effects.

Across all three noise families, the pre-transform embeddings highlighted the high sensitivity of raw statistical descriptors to noise, exhibiting anisotropic spreading, overlap, and manifold distortion. In contrast, the proposed transformation adaptively reshapes these corrupted densities into uniform, compact, and linearly separable clusters. This confirms its capability to maintain the essential fault structure while mitigating both random and coherent disturbances, making it suitable for real-world bearing diagnostics in noisy industrial environments.

6.8. Anomaly detection in the original feature space

Although PCA and t–SNE were utilized in this study to enable low-dimensional analysis and visualization, the proposed transformation does not inherently rely on dimensionality reduction. To evaluate its

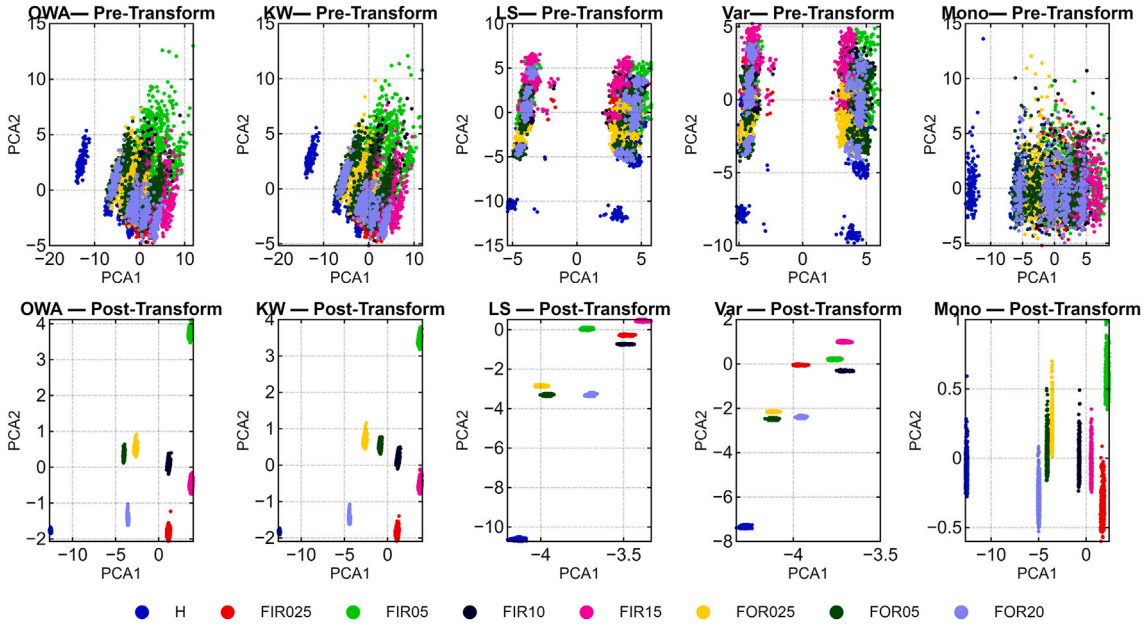

Fig. 26. The impact of structural noise synthesized from a pseudo-FRF profile is examined. The most significant pre-transform distortion is observed in this context (top row); however, the proposed transformation continues to compact each class into distinct clusters(bottom row).

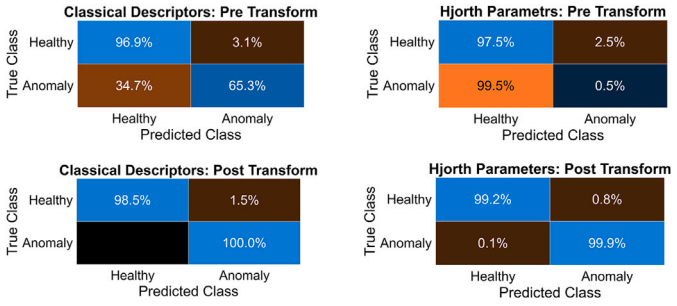

Fig. 27. The performance of anomaly detection in the *original feature space* without the application of dimensionality reduction is evaluated. Confusion matrices are presented for both classical statistical descriptors and Hjorth parameters, which were assessed before and after the implementation of the proposed distribution-based transformation. The findings indicate a significant enhancement in anomaly detection accuracy post-transformation, thereby affirming the efficacy of the proposed method in operating directly within high-dimensional feature spaces.

efficacy within the original feature space, anomaly detection was assessed by applying the transformation directly to the ranked feature set without any intermediate projection. Fig. 27 presents the anomaly detection results obtained using a one-class SVM in the original feature space for both classical statistical descriptors and Hjorth parameters. Compared to the pre-transformation scenario, the proposed distribution-based transformation results in a significant reduction in false negatives and false positives, thereby markedly enhancing the detection performance. These findings confirm that the proposed method effectively improves class separability and anomaly detectability directly in high-dimensional feature spaces, with dimensionality reduction serving solely as an auxiliary tool for visualization rather than as a prerequisite for performance improvement.

6.9. *Effect of symmetric cluster shifting*

ForestGreenTo determine whether the asymmetric cluster-shifting operation introduces a distributional bias, we examined a symmetric alternative in which *both* clusters are shifted toward their midpoints. In this analysis, adaptive transformation with symmetric shifting was initially applied directly in the original feature space, and the resulting transformed features were subsequently projected onto two principal components for visualization. Fig. 28 illustrates the PCA-reduced representations obtained before and after applying the adaptive transformation with symmetric shifting. Across all ranked feature sets, the post-transform embeddings exhibited improved class separation relative to the pre-transform space, indicating that the transformation primarily reshaped within-class dispersion and inter-class structure rather than introducing qualitative distortions in the embedding geometry.

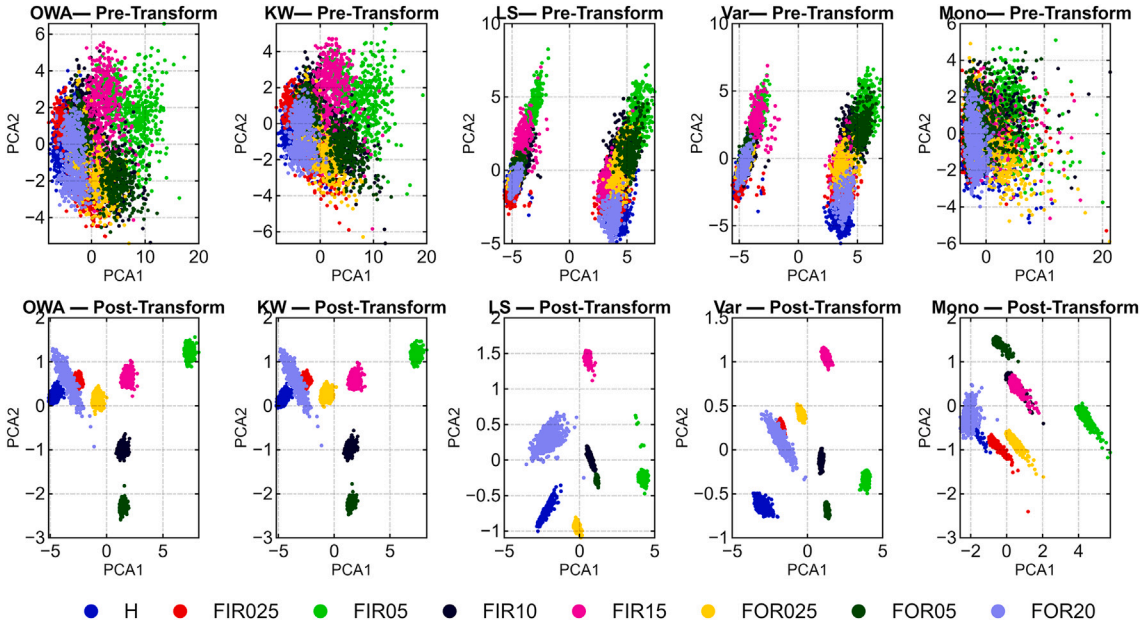

Fig. 28. PCA projections of both the original (Pre-Transform) and transformed (Post-Transform) feature spaces for the MOIRA–UNIMORE dataset are presented across five ranked feature sets. The post-transform embeddings depicted here are derived using the *symmetric* cluster-shifting variant, wherein both modes are adjusted toward their midpoint.

7. Conclusion

This study introduces an adaptive distribution-based feature transformation algorithm aimed at enhancing bearing fault detection in independent cart systems (ICS). The proposed method effectively reshapes the skewed, bimodal, and high-variance feature distributions commonly observed in vibration signals from ICS, thereby improving class separability without the need for labeled data or deep-learning architectures. The findings are as follows: 1. The transformation significantly enhanced cluster compactness and separation in both PCA and t-SNE embeddings across various feature ranking strategies. 2. Quantitative analysis using one-class SVM and Isolation Forest classifiers demonstrated the superior performance of the proposed method compared to traditional transformations and other multimodality-aware approaches. 3. The method exhibited robustness against different noise types, including broadband measurement noise, narrowband interference, and sinusoidal structural noise, and maintained clear class separation even under challenging conditions. 4. The approach was validated using both single-cart and three-cart experimental configurations, demonstrating its scalability to more complex system interactions. These results address the critical challenges in ICS diagnostics, including label scarcity, high dimensionality, overlapping classes, and non-stationary dynamics. The proposed method offers a promising solution for enhancing fault detection in complex industrial systems, particularly in cases where traditional rotating machinery approaches may prove inadequate. Future work could explore the integration of this transformation technique with advanced machine learning models and its application to other types of mechatronic systems with complex and nonstationary dynamics.

CRediT authorship contribution statement

Abdul Jabbar: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marco Cocconcelli:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Gianluca D'Elia:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered potential competing interests:

Abdul Jabbar reports that financial support was provided by the European Commission Marie Sklodowska-Curie Actions. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie Program through the H2020 ETN MOIRA Project (GA 955681).

References

[1] Jabbar A, D'Elia G, Cocconcelli M. Experimental setup for non-stationary condition monitoring of independent cart systems. In: Kumar U, Karim R, Galar D, Kour R, editors. International congress and workshop on industrial AI and eMaintenance. Springer Nature Switzerland; 2024. p. 517–30. https://doi.org/10.1007/978-3-031-39619-9-38

[2] Jabbar A, Cocconcelli M, d'Elia G, Strozzi M, Rubini R. Results on experimental data analysis of independent cart systems in non-stationary conditions. In: Surveillance, vibrations, shock and noise. Toulouse, France; 2023. https://hal.science/hal-04165905., hal-04165905.

[3] Jabbar A, Fonte C, D'Elia G, Cocconcelli M. Ball-bearings fault detection for an independent cart system: experimental campaign and preliminary results. In: Proceedings of ISMA 2024 - international conference on noise and vibration engineering and USD 2024 - international conference on uncertainty in structural dynamics. Leuven, Belgium: ISMA; 2024. p. 1698–711.

[4] Jabbar A, Mazzonetto M, Orazi L, Cocconcelli M. Ultrafast laser damaging of ball bearings for the condition monitoring of a Fleet of linear motors. In: PHM Society European Conference, 8; 2024. p. 10. https://doi.org/10.36001/phme.2024.v8i1.4136

[5] Jabbar A, D'Elia G, Cocconcelli M. Distribution reshaping transformation for bearing fault diagnosis in independent cart systems. IEEE Access 2025;13:200403–30. https://doi.org/10.1109/ACCESS.2025.3636190

[6] Jabbar A, Cocconcelli M, D'Elia G. Bimodal distribution detection and transformation for Gaussian merging for bearing fault classification of independent cart systems under nonstationary conditions. Measurement 2026;262:119900. https://doi.org/10.1016/j.measurement.2025.119900

[7] Tayyab SM, Chatterton S, Pennacchi P. Intelligent defect diagnosis of rolling element bearings under variable operating conditions using convolutional neural network and order maps. Sensors 2022;22:2026. https://doi.org/10.3390/s22052026
[8] Farhat MH, Chiementin X, Chaari F, Bolaers F, Haddar M. Order-based identification of bearing defects under variable speed condition. Appl Sci 2021;11. https://doi.org/10.3390/app11093962. https://www.mdpi.com/2076-3417/11/9/3962.
[9] Liu Z, Peng D, Zuo MJ, Xia J, Qin Y. Improved Hilbert–huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. ISA Trans 2022;125:426–44. https://doi.org/10.1016/j.isatra.2021.07.011
[10] Wu T, Chen J, Wang C. Characterization of gear faults in variable rotating speed using Hilbert-huang transform and instantaneous dimensionless frequency normalization. Mech Syst Signal Process 2012;30:103–22. https://doi.org/10.1016/j.ymssp.2012.01.022
[11] Jiang Y, Zhou J, Wu X, Liu T, Liu X. Vision-based bearing fault diagnosis under non-stationary conditions using optimized short-time concentrated transform method. Reliab Eng Syst Saf 2025;262:111183. https://doi.org/10.1016/j.ress.2025.111183
[12] Peng B, Bi Y, Xue B, Zhang M, Wan S. A survey on fault diagnosis of rolling bearings. Algorithms 2022;15. https://doi.org/10.3390/a15100347. https://www.mdpi.com/1999-4893/15/10/347.
[13] Ma J, Wei J, Li Q, Xia L. Variable-speed bearing fault diagnosis based on BDVMD, FRTSMFrBSIE, and parameter-optimized GRU-MHSA. Processes 2025;13. https://doi.org/10.3390/pr13020498. https://www.mdpi.com/2227-9717/13/2/498.
[14] Huang W, Gao G, Li N, Jiang X, Zhu Z. Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis. IEEE Trans Instrum Meas 2019;68:2819–29. https://doi.org/10.1109/TIM.2018.2868519
[15] Saufi SR, Ahmad ZAB, Leong MS, Lim MH. Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: a review. IEEE Access 2019;7:122644–62. https://doi.org/10.1109/ACCESS.2019.2938227
[16] Hakim M, Omran AAB, Ahmed AN, Al-Waily M, Abdellatif A. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: taxonomy, overview, application, open challenges, weaknesses and recommendations. Ain Shams Eng J 2023;14:101945. https://doi.org/10.1016/j.asej.2022.101945
[17] Islam MR, Kim Y-H, Kim J-Y, Kim J-M. Detecting and learning unknown fault states by automatically finding the optimal number of clusters for online bearing fault diagnosis. Appl Sci 2019;9. https://doi.org/10.3390/app9112326. https://www.mdpi.com/2076-3417/9/11/2326.
[18] Brito LC, Susto GA, Brito JN, Duarte MAV. Fault detection of bearing: an unsupervised machine learning approach exploiting feature extraction and dimensionality reduction. Informatics 2021;8. https://doi.org/10.3390/informatics8040085. https://www.mdpi.com/2227-9709/8/4/85.
[19] Spina DE, de O. Campos LF, de Arruda WF, Melo A, de S. Alves MF, Rabello GL, Anzai TK, Pinto JC. Comparison of autoencoder architectures for fault detection in industrial processes. Digit Chem Eng 2024;12:100162. https://doi.org/10.1016/j.dche.2024.100162
[20] Zhao Z, Zhang Q, Yu X, Sun C, Wang S, Yan R, Chen X. Applications of unsupervised deep transfer learning to intelligent fault diagnosis: a survey and comparative study. IEEE Trans Instrum Meas 2021;70:3525828. https://doi.org/10.1109/TIM.2021.3116309
[21] Xiao Q, Yang M, Yan J, et al. Feature decoupling integrated domain generalization network for bearing fault diagnosis under unknown operating conditions. Sci Rep 2024;14:30848. https://doi.org/10.1038/s41598-024-81489-6
[22] Kim I, Wook Kim S, Kim J, Huh H, Jeong I, Choi T, Kim J, Lee S. Single domain generalizable and physically interpretable bearing fault diagnosis for unseen working conditions. Expert Syst Appl 2024;241:122455. https://doi.org/10.1016/j.eswa.2023.122455
[23] Yan X, She D, Xu Y. Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions. Expert Syst Appl 2023;216:119479. https://doi.org/10.1016/j.eswa.2022.119479
[24] Li T, Peng Z, Xu H, He Q. Parameterized domain mapping for order tracking of rotating machinery. IEEE Trans Ind Electron 2023;70:7406–16. https://doi.org/10.1109/TIE.2022.3201311
[25] Sa'd MA, Jalonen T, Kiranyaz S, Gabbouj M. Quadratic time-frequency analysis of vibration signals for diagnosing bearing faults, ArXiv arXiv:2401.01172. 2024.
[26] Li Y, Xia H. Bearing fault diagnosis for variable operating conditions based on deep learning and mutual attention feature fusion. In: Proceedings of the 43rd Chinese control conference. IEEE; 2024. p. 4925–30.
[27] Zhang L, Lv Y, Huang W, Yi C. Bearing fault diagnosis under various operation conditions using synchrosqueezing transform and improved two-dimensional convolutional neural network. Meas Sci Technol 2022;33:085002. https://doi.org/10.1088/1361-6501/ac69b1
[28] Fu S, Wu Y, Wang R, Mao M. A bearing fault diagnosis method based on wavelet denoising and machine learning. Appl Sci 2023;13:5936. https://doi.org/10.3390/app13105936
[29] Bristi SD, Tatha MJ, Ali MF, Bhatti UA, Sarker SK, Masud M, Ghadi YY, Algarni A, Saha DK. A meta-heuristic sustainable intelligent internet of things framework for bearing fault diagnosis of electric motor under variable load conditions. Sustainability 2023;15:16722. https://doi.org/10.3390/su152416722
[30] Pule M, Matsebe O, Samikannu R. Application of PCA and SVM in fault detection and diagnosis of bearings with varying speed. Math Probl Eng 2022;2022:1–12. https://doi.org/10.1155/2022/5266054
[31] Alhams A, Abdelhadi A, Badri Y, Sassi S, Renno J. Enhanced bearing fault diagnosis through trees ensemble method and feature importance analysis. J Vib Eng Technol 2024;12:109–25. https://doi.org/10.1007/s42417-024-01405-0
[32] An Y, Zhang K, Chai Y, Liu Q, Huang X. Bearing fault diagnosis under variable working conditions base on contrastive domain adaptation method. IEEE Trans Instrum Meas 2022;71:1–11. https://doi.org/10.1109/TIM.2022.3200106
[33] Chen Y, Zhang D, Yan R, Xie M. Applications of domain generalization to machine fault diagnosis: a survey. IEEE/CAA J Autom Sin 2025;12:1–15. https://doi.org/10.1109/JAS.2025.125120, Early Access.
[34] Jabbar A, Cocconcelli M, D'Elia G, Borghi D, Capelli L, Cavalaglio Camargo Molano J, Strozzi M, Rubini R. MOIRA-UNIMORE bearing data set for independent cart systems. Appl Sci 2025;15. https://doi.org/10.3390/app15073691. https://www.mdpi.com/2076-3417/15/7/3691.
[35] Meng Y, Zhou J, Lei F, Li D, Liu R. A novel class of non-Gaussian system performance assessment and controller parameter tuning methods. ISA Trans 2024;154:199–212. https://doi.org/10.1016/j.isatra.2024.08.031
[36] Johnson JE, Laparra V, Piles M, Camps-Valls G. Gaussianizing the earth: multidimensional information measures for earth data analysis. IEEE Geosci Remote Sens Mag 2021;9:191–208. https://doi.org/10.1109/MGRS.2021.3066260
[37] Laparra V, Hepburn A, Johnson JE, Malo J. Orthonormal convolutions for the rotation based iterative gaussianization. In: 2022 IEEE international conference on image processing (ICIP); 2022. p. 4018–22. https://doi.org/10.1109/ICIP46576.2022.9897849
[38] Yu K, Wu Z, Sun J, Zhang Y, Xu Y, Wei Z, Zheng S. Accelerating hyperspectral anomaly detection with enhanced multivariate gaussianization based on FPGA. IEEE Trans Geosci Remote Sens 2024;62:1–12. https://doi.org/10.1109/TGRS.2024.3476152
[39] Padrón-Hidalgo JA, Laparra V, Camps-Valls G. Unsupervised anomaly and change detection with multivariate gaussianization. IEEE Trans Geosci Remote Sens 2022;60:1–10. https://doi.org/10.1109/TGRS.2021.3116186
[40] Wang P, Long Z, Lv Z, Wang Z. Fault detection for Non-Gaussian processes using multiple canonical correlation analysis models and Box-Cox transformation. IEEE Access 2019;7:68707–17. https://doi.org/10.1109/ACCESS.2019.2914960
[41] Tang J, You Y, Zhao Y, Guo C, Li Z, Yang B. Fault diagnosis of HVAC system sensors: a method based on Box-Cox transformation and multi-model fusion. Energy Rep 2025;13:3489–503. https://doi.org/10.1016/j.egyr.2025.03.012
[42] Yusoff M, Mahmud Y, Azmi PAR, Sallehud-Din MTM. The improvement of SMOTE-ENN-XGBoost through yeo Johnson strategy on dissolved gas analysis dataset. Energy Rep 2025;13:6281–90. https://doi.org/10.1016/j.egyr.2025.05.013
[43] Chen T, Guo L, Gao H, Wang D, Feng T, Yu Y. Investigations on improved Box-Cox sparsity measures for machine condition monitoring. ISA Trans 2025;157:466–80. https://doi.org/10.1016/j.isatra.2024.12.010
[44] Chen B, Zhang W, Xi Gu J, Song D, Cheng Y, Zhou Z, Gu F, Ball AD. Product envelope spectrum optimization-gram: an enhanced envelope analysis for rolling bearing fault diagnosis. Mech Syst Signal Process 2023;193:110270. https://doi.org/10.1016/j.ymssp.2023.110270
[45] Chen M-C, Hsu C-C, Malhotra B, Tiwari MK. An efficient ICA-DW-SVDD fault detection and diagnosis method for non-Gaussian processes. Int J Prod Res 2016;54:5208–18. https://doi.org/10.1080/00207543.2016.1161250
[46] Giantomassi A, Ferracuti F, Iarlori S, Ippoliti G, Longhi S. Electric motor fault detection and diagnosis by kernel density estimation and Kullback–Leibler divergence based on stator current measurements. IEEE Trans Ind Electron 2015;62:1770–80. https://doi.org/10.1109/TIE.2014.2370936
[47] Lang CI, Sun F-K, Lawler B, Dillon J, Dujaili AA, Ruth J, Cardillo P, Alfred P, Bowers A, Mckiernan A, Boning DS. One class process anomaly detection using kernel density estimation methods. IEEE Trans Semicond Manuf 2022;35:457–69. https://doi.org/10.1109/TSM.2022.3181468
[48] Qian Q, Pu H, Tu T, Qin Y. Variance discrepancy representation: a vibration characteristic-guided distribution alignment metric for fault transfer diagnosis. Mech Syst Signal Process 2024;217:111544. https://doi.org/10.1016/j.ymssp.2024.111544