Mechanical Systems and Signal Processing 250 (2026) 114191

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Full length article

Intelligent time–frequency feature embedding and reshaping for bearing fault diagnosis in motion control applications

Abdul Jabbar *, Marco Cocconcelli, Gianluca D'Elia

The Department of Sciences and Engineering Methods (DISMI), University of Modena and Reggio Emilia (UNIMORE), Via G. Amendola, pad-Morselli, Reggio Emilia, 42122, Italy

ARTICLE INFO

Communicated by Y. Lei

Keywords:
Ball bearings
Fault diagnosis
Condition monitoring
Feature transformation
Independent cart systems

ABSTRACT

Accurate diagnosis of bearing faults under nonstationary operating conditions presents significant challenges, particularly in independent cart systems where variable speeds, coupled translational–rotational motion, and transient dynamics substantially influence vibration signals. In such contexts, traditional time–frequency representations and unsupervised learning methods often yield inadequate class separation and unreliable anomaly detection results. This study presents a fully unsupervised diagnostic framework that overcomes existing limitations through two complementary innovations. First, an intelligent strategy automatically selects the lengths of the spectrogram windows using Theil index-based inequality measures. Second, a technique for reshaping distributions transforms the extracted features into compact, uniformly distributed representations without the need for class labels. Experimental validation using the MOIRA–UNIMORE bearing dataset demonstrated significant enhancements in terms of feature compactness and class separability. Further robustness and cross-dataset validation on the Case Western Reserve University and Politecnico di Torino bearing datasets corroborated the stability and generalizability of the proposed framework. These findings suggest that integrating adaptive time–frequency analysis with principled distribution reshaping provides an effective, computationally efficient solution for unsupervised bearing-fault diagnosis in nonstationary industrial environments.

1. Introduction

Recent bearing fault diagnosis research has increasingly adopted intelligent frameworks that integrate signal processing with machine learning to automate feature extraction under nonstationary operating conditions with variable speed and load [1,2]. Localized time–frequency methods (STFT, WT, and HHT) and adaptive decomposition techniques (EMD, EEMD, CEEMDAN, and VMD) are widely used to capture transient dynamics and extract fault-relevant modes [3–6]. In parallel, deep learning models, including convolutional neural networks (CNNs), autoencoders, and more recently transformer-based and self-supervised architectures, have been employed to learn hierarchical representations directly from raw signals or time–frequency images [6–11], while unlabeled diagnostic scenarios often rely on unsupervised pipelines combining k-means, self-organizing maps, or Isolation Forests with autoencoder-based representation learning [12–15].

To address cross-condition variability, domain adaptation and unsupervised deep transfer learning (UDTL) have emerged as key paradigms for fault diagnosis without labeled target data [16–19]. Prominent approaches include feature alignment models, such as UFAN and domain-aligned CNNs, as highlighted in recent studies [20,21]. Additionally, there are domain generalization strategies

* Corresponding author.
E-mail 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.ymssp.2026.114191
Received 19 November 2025; Received in revised form 13 February 2026; Accepted 22 March 2026

Available online 25 March 2026
0888-3270/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

aimed at achieving invariant representations [22]. Other notable methods include hybrid time–frequency deep learning pipelines, such as EEMD-CWT attention, SST-CNN, and flexible generalized demodulation. Furthermore, few-shot and meta-learning models have been developed specifically for data-scarce environments [23–28].

Despite these advances, including recent resilient deep learning strategies designed for complex load and speed variations [29], many methods remain benchmark-oriented and struggle in real-world deployment under noise, data shift, and limited labeling. These challenges are especially pronounced in industrial platforms exhibiting strong nonstationarity and hybrid translational–rotational dynamics, such as Independent Cart Systems (ICS).

The diagnosis of rolling-element bearings in Independent Cart Systems (ICS) [30–32] poses a unique challenge owing to the inherently nonstationary operating environment of these systems. Unlike traditional motor–shaft assemblies, the bearings in an ICS maintain continuous mechanical contact with the guide rail and simultaneously undergo *both translational and rotational motions* as the cart moves along the track. This coupled kinematic behavior, driven by the mover's variable-speed profile, bidirectional motion cycles, and fluctuating electromagnetic forces, induces nonlinear time-varying dynamics in the measured vibration signals [33–35].

These conditions cause the embedded feature space — whether derived from handcrafted descriptors, spectral representations, or modern nonlinear projections — to deviate significantly from the idealized assumptions of stationarity and geometric regularity. When manifold learning techniques such as t-distributed stochastic neighbor embedding (t-SNE) [36,37] or uniform manifold approximation and projection (UMAP) [38–40] are applied, the latent space commonly exhibits:

- elongated, curved, or spiral-shaped trajectories produced by simultaneous translation–rotation dynamics,
- speed- and position-dependent drift in the low-dimensional embedding,
- strong local density imbalance across different motion phases,
- partial or substantial overlap between healthy and faulty states.

These geometric distortions impede the performance of unsupervised learning algorithms, particularly clustering and one-class anomaly detection, which rely on compact intra-class structures and well-separated inter-class boundaries. Conventional variance-stabilizing or Gaussianizing transformations do not adequately correct these nonlinear effects, necessitating a principled preprocessing strategy that actively reshapes the distribution before downstream learning.

Non-Gaussian characteristics in process and vibration data have long posed challenges for control performance assessment, reliability analysis, anomaly detection, and fault diagnosis, motivating the development of distribution transformation and regularization strategies to Gaussianize or homogenize data. Early efforts focused on quantile-based and power transformations, which provide interpretable and training-free mechanisms for approximating Gaussianity. For instance, a quantile transformation preserving mutual information while mapping non-Gaussian system outputs to Gaussian-like representations was proposed in [41] to enable consistent controller benchmarking under non-Gaussian disturbances.

Subsequent research extended Gaussianization to high-dimensional settings. Rotation-Based Iterative Gaussianization (RBIG) [42,43] combines marginal Gaussianization with orthonormal rotations to achieve joint Gaussian representations, with later extensions introducing convolutional RBIG for scalable layer-wise processing [44]. Hardware-efficient implementations such as ORBIG [45] and fully unsupervised invertible Gaussian models with explicit likelihood estimation [46] further demonstrate the effectiveness of distribution Gaussianization as a model-free preprocessing paradigm.

In industrial fault diagnosis, power transformations such as Box–Cox [47] and Yeo–Johnson [48] are widely used to regularize skewed or heavy-tailed monitoring statistics. Their integration into multi-regime canonical correlation analysis enables Gaussian-based control limits under varying conditions [49], while related preprocessing strategies have improved the robustness of HVAC fault detection [50] and addressed class imbalances in transformer diagnostics [51]. Beyond marginal normalization, Box–Cox concepts have been embedded directly into fault-sensitive metrics, including sparsity-enhanced measures for weak fault amplification [52] and generalized envelope-spectrum formulations for low signal-to-noise environments [53].

Parallel studies have explored distribution-aware and sparsity-driven transformations for blind filtering and deconvolution in nonstationary vibration signals. Box–Cox-based sparse measures unify kurtosis- and entropy-driven criteria to extract impulse-dominated components under strong noise and cyclostationary interference [54,55], with extensions incorporating cyclic band sparsity and fractional lower-order statistics to improve robustness. Structural simulation and modal analysis have been combined with Box–Cox sparse-measure-based deconvolution to enhance resonance-band selection in large-scale systems, such as wind turbines [56]. Box–Cox transformations have been utilized in dynamic reliability analysis by employing transformed extreme-value distributions [57].

Unsupervised and one-class methodologies have additionally advanced nonparametric density modeling for characterizing nominal behavior in non-Gaussian environments. Representative approaches include ICA–DW–SVDD frameworks for feature decorrelation and one-class encapsulation [58], kernel density estimate (KDE)-based anomaly detection for electric motor monitoring [59], and lightweight KDE detectors capable of adapting to gradual concept drift using only healthy data [60]. Related domain-alignment strategies, such as Variance Discrepancy Representation (VDR), explicitly manage mean and variance mismatches across domains under heavy-tailed distributions [61].

Despite their demonstrated effectiveness in signal enhancement, statistical regularization, and threshold calibration, these techniques are predominantly developed at the signal, spectral, or scalar-statistic levels. Consequently, they do not explicitly address the geometry, multimodality, or density imbalance of low-dimensional feature spaces produced by nonlinear embedding. These limitations are particularly evident in clustering and unsupervised fault diagnosis under strongly nonstationary conditions, underscoring the need for preprocessing strategies that directly reshape dynamically induced feature manifolds while preserving class structure and discriminability.

The limitations inherent in the current distribution normalization and density homogenization techniques highlight a broader issue: effective vibration-based diagnosis under non-stationary conditions cannot depend solely on distribution reshaping. In practical applications, the efficacy of subsequent clustering and anomaly detection is limited by the time–frequency representation used to extract features. Nonstationary vibration signals often require windowed transforms; however, fixed or manually adjusted parameters yield inconsistent spectrogram quality across different operating conditions. Consequently, even the most sophisticated reshaping methods are inadequate when the foundational representation is suboptimal. These findings underscore the need for an integrated framework that combines adaptive time–frequency analysis with systematic distribution reshaping, ensuring that both representation and statistical geometry are optimized for unsupervised diagnosis.

Beyond distribution normalization, extensive research has addressed nonstationarity at the representation and feature-extraction level through time–frequency (TF) analysis, manifold learning, and data-driven models. Classical and adaptive TF methods, including generalized and optimized S-transform variants, optimized weight TF analysis, and adaptive STFT formulations, aim to enhance fault-related spectral concentration via resolution adjustment, convex optimization, or energy redistribution [62,63]. While effective in highlighting time-varying fault frequencies and suppressing noise, these approaches primarily focus on TF representation enhancement or signal reconstruction, without explicitly considering the impact of nonstationarity-induced distortions on downstream feature spaces.

To capture nonlinear and transient structures in TF representations, time–frequency manifold (TFM) learning frameworks have been proposed, leveraging nonlinear dimensionality reduction techniques such as locally linear embedding, local tangent space alignment, and Laplacian eigenmaps to generate compact descriptors [64]. Subsequent developments introduced fast TFM learning and TF manifold image synthesis by combining TF analysis with wavelet-based compression, histogram matching, and manifold-guided reconstruction to enhance transient features at a reduced computational cost [65,66]. In parallel, distribution-aware spectral preprocessing techniques, including ECDF-based statistical spectral analysis [67] and cumulative distribution sharpness profiling [68], have enabled training-free fault diagnosis under variable-speed and tacholess conditions. However, they remain limited to marginal or spectral statistics and do not explicitly address feature-space geometry or multimodality.

More recently, bearing fault diagnosis has increasingly integrated TF representations with supervised deep feature learning, where vibration signals are transformed into TF images via STFT, CWT, or related transforms and processed using CNN–LSTM architectures, attention-enhanced hybrid models, YOLO-based detectors, and transfer-learning frameworks [69–71]. Extensions addressing limited data quality and cross-condition variability incorporate robust network designs, pseudo-labeling, and knowledge transfer strategies [72–74], with comparative studies confirming the critical influence of TF representation on diagnostic performance in noisy conditions [75]. In these pipelines, feature alignment and transformation are implicitly learned via supervised training. At the same time, dimensionality reduction techniques such as t-SNE are typically used only for visualization rather than as functional diagnostic components.

In contrast to existing TF-based and TF–DL approaches, which primarily emphasize spectral concentration or supervised representation learning, the proposed TAIW strategy explicitly optimizes the statistical concentration of TF representations using the Theil index, yielding spectrograms that are inherently more consistent across operating regimes. Furthermore, unlike TF manifold learning and deep models that rely on supervised or implicit transformations, the proposed DRBT directly reshapes low-dimensional embeddings in a fully unsupervised and deterministic manner, explicitly addressing multimodal and density-imbalance issues. As a result, the integrated TAIW–DRBT framework simultaneously optimizes representation quality and feature-space geometry without requiring labeled data, network training, or retraining across datasets.

Unlike conventional time–frequency optimization strategies that focus on energy concentration or reconstruction fidelity, the proposed approach introduces a statistical optimality criterion based on inequality theory, providing a theoretically grounded link between time–frequency representation quality and downstream feature-space geometry. This explicit coupling between representation-level statistical concentration and embedding-level distribution reshaping constitutes a key methodological departure from existing bearing fault diagnosis frameworks.

Modern industrial systems increasingly require lightweight automated fault diagnosis frameworks that operate reliably under real-time constraints. Although deep learning models offer strong representational power, their reliance on large labeled datasets and high computational costs limits their applicability in several industrial scenarios. In such settings, signal transforms remain indispensable, as they reveal fault signatures hidden within raw, nonstationary vibration signals by projecting them into joint time–frequency domains. However, conventional transforms rely on fixed parameters and fail to adapt to varying speeds and transients, resulting in inconsistent representations and information loss issues. Moreover, the resulting statistical distributions of the extracted features are frequently irregular and multimodal, with significant overlap between fault classes, thereby increasing false alarms and misclassifications in unsupervised learning.

The challenge associated with Independent Cart Systems (ICS) does not inherently arise from the use of fixed window lengths *per se*, but rather from the dependence on manually selected or heuristically adjusted windows in extensive, nonstationary environments. In practical applications, ICS installations consist of multiple carts, each equipped with several bearings that operate continuously over extended periods and along diverse trajectories. Under these conditions, selecting an appropriate time–frequency window through visual inspection, heuristic rules, or trial-and-error methods is impractical and non-reproducible. Although a single fixed window may ultimately be employed for analysis, its value must be determined automatically and objectively from the data to ensure consistency across various operating regimes. This necessity drives the development of principled, data-driven window selection criteria that maintain the simplicity of fixed-window representations while eliminating the need for manual tuning and addressing scalability limitations.

A significant limitation inherent in the current Gaussianization and density regularization methodologies is rooted in their operational objectives rather than their statistical validity. Gaussianization techniques predominantly aim to transform data towards an ideal global distribution, typically characterized by a zero-mean, unit-variance Gaussian, through processes such as marginal normalization, probabilistic modeling, or invertible transformations. Although these methods are effective in mitigating skewness and heavy tails, they generally maintain inter-class overlap within embedded spaces. In contrast, density-homogenization strategies focus on equalizing the sample density but fail to explicitly address the locally induced multimodality, often resulting in entangled class manifolds. Consequently, density imbalance and overlap persist in low-dimensional embeddings, thereby constraining the efficacy of clustering and anomaly detection in one class. This gap underscores the necessity of a preprocessing strategy that directly engages with embedding geometry and adapts to locally induced statistical modes, rather than imposing global distributional conformity.

To address these limitations, we propose a three-stage framework that integrates Theil-index-based spectrogram analysis, feature compression, and distribution reshaping for robust bearing-fault diagnosis in independent cart systems (ICS). The first stage introduces Theil-Aware Intelligent Windowing (TAIW), a Theil-index-driven strategy that automatically selects optimal window sizes to enhance the spectrogram representation of nonstationary vibration signals. In the second stage, spectrogram computation and feature compression were performed to obtain compact and informative representation. The third stage applies Distribution Reshaping via Bimodal Transformation (DRBT). This novel method converts arbitrary or irregular feature distributions into compact representations that reduce class overlap and are better suited for clustering and unsupervised analysis. The integrated pipeline provides adaptive time–frequency analysis and geometry-friendly feature shaping, enabling reliable diagnosis under the complex dynamics of ICS.

Key contributions

The principal contributions of this study are as follows:

- A lightweight, fully unsupervised diagnostic framework is formulated for Independent Cart Systems (ICS), obviating the necessity for extensive labeled datasets and diminishing computational overhead in comparison to deep learning methodologies.
- *Theil-Aware Intelligent Windowing (TAIW)* is introduced as an innovative Theil-index-driven approach for automatic window-size selection in spectrogram computation, enabling enhanced extraction of fault signatures from nonstationary vibration signals without manual parameter tuning.
- *Distribution Reshaping via Bimodal Transformation (DRBT)* is developed as a novel feature-transformation technique that converts irregular, overlapping, or multimodal distributions into compact, Gaussian-like clusters, thereby mitigating false alarms and misclassifications in unsupervised learning.
- An integrated three-stage pipeline is established, combining automatic spectrogram window selection and distribution reshaping to produce representations that are both spectrally informative and cluster-compatible, thereby enhancing robustness under variable-speed and transient operating conditions.
- The proposed framework is rigorously evaluated on three open-access bearing datasets, including MOIRA–UNIMORE, CWRU, and Politecnico di Torino, demonstrating reliable fault-detection performance under the complex translational–rotational dynamics characteristic of industrial transport platforms.

The remainder of this paper is organized as follows. Section 2 details the proposed framework (TAIW for window selection, t-SNE, UMAP embedding, and DRBT). Section 3 introduces the public datasets used to validate the proposed framework. Section 4 presents and discusses the experimental results. Section 5 concludes the paper and outlines the directions for future work.

2. Methodology

This section outlines the proposed vibration-based fault diagnosis framework for independent cart systems (ICS). The workflow consists of three stages (Fig. 1): (i) Theil-Aware Intelligent Windowing (TAIW) selects an analysis window and overlap that optimally preserves the salient time–frequency structure; (ii) a spectrogram is computed using the selected window, and frame-level features are embedded into two dimensions via t-distributed stochastic neighbor embedding (t-SNE) [36,37] or uniform manifold approximation and projection (UMAP) [39,40]; and (iii) Distribution Reshaping via Bimodal Transformation (DRBT) enhances cluster compactness and separation to facilitate downstream detection and classification (see Algorithm 1 for the end-to-end implementation).

2.1. Implementation summary

Stage I — TAIW. Given $(\mathcal{W}, \mathcal{O})$, the algorithm iterates over candidate pairs (w, ov), computes a spectrogram for each, evaluates the *Theil-based concentration scores* along time and frequency, $Q_t(w)$ and $Q_f(w)$, combines them as $Q_{tf}(w) = Q_t(w)\,Q_f(w)$, and selects

$$(w^\star, \mathrm{ov}^\star) \in \arg\max_{(w,\mathrm{ov})\in\mathcal{W}\times\mathcal{O}} Q_{tf}(w).$$

Stage II — Spectrogram and feature compression. The spectrogram is recomputed using $(w^\star, \mathrm{ov}^\star)$, and each frame is treated as a sample in the feature space. A 2-D t-SNE or UMAP embedding was then computed, representing the latent manifold of spectrotemporal patterns in a compact form that is suitable for density analysis.

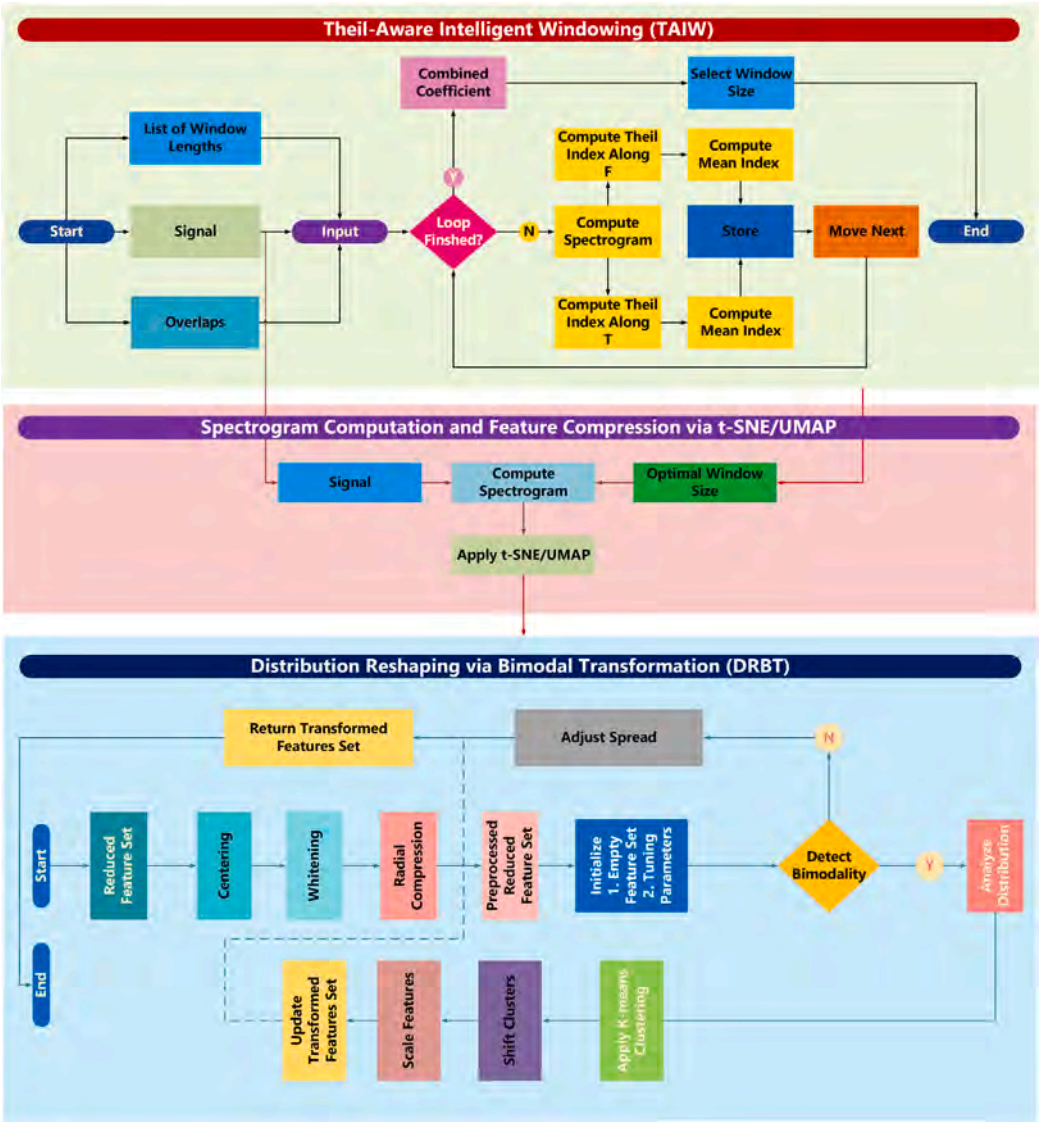

Fig. 1. Proposed three-stage diagnostic framework. Stage 1: Theil-Aware Intelligent Windowing (TAIW) for automatic window size selection. Stage 2: Spectrogram computation and feature compression. Stage 3: Distribution Reshaping via Bimodal Transformation (DRBT) to obtain compact, cluster-friendly features for reliable fault diagnosis in ICS.

Stage III — DRBT. A concise, deterministic sequence of operations, including centering/scaling, whitening, radial compaction, and adaptive cluster shift along a bimodal axis, was applied to tighten the clusters, homogenize feature densities, and enlarge inter-cluster margins. The entire transformation is non-iterative, data-driven, and computationally efficient.

2.2. Notation and conventions

Table 1 enumerates symbols and terms utilized throughout, without introducing equations. Units are provided where applicable.

2.3. Theil-Aware Intelligent Windowing (TAIW)

Selecting an appropriate time–frequency resolution is essential for analyzing non-stationary vibration signals. The proposed Theil-Aware Intelligent Windowing (TAIW) method autonomously determines an optimal window length w by maximizing an energy-concentration criterion based on the Theil index. This information-theoretic measure, which is traditionally used to quantify inequality, was employed to assess the degree of energy localization in the spectrogram. Higher Theil values indicate a stronger temporal and spectral concentration of fault-related events than lower values.

Algorithm 1 End-to-End Fault Detection Pipeline (TAIW → Embedding → DRBT)

Require: Vibration signal $x[n]$, sampling rate f_s, candidate sets $\mathcal{W}, \mathcal{O}$, N_{FFT}, embedding method $\Phi \in \{\text{t-SNE}, \text{UMAP}\}$, DRBT parameters Θ_{DRBT}, detector D (e.g., OCSVM/IF), train/test split rule

Ensure: Fault decision (or anomaly score) per sample/window and evaluation metrics

Stage I (TAIW): Compute $(w^\star, \text{ov}^\star) \leftarrow \arg\max_{(w,\text{ov})} Q_{tf}(w, \text{ov})$ using Algorithm 2

Stage II (TF features): Compute STFT with $(w^\star, \text{ov}^\star)$ and obtain power spectrogram $P_{w^\star}$

Form frame-level feature matrix $Z \in \mathbb{R}^{N \times d}$ from $P_{w^\star}$ (each time frame is one sample)

Compute 2-D embedding $X \leftarrow \Phi(Z) \in \mathbb{R}^{N \times 2}$

Stage III (DRBT): Compute reshaped embedding $X_{\text{transformed}} \leftarrow \text{DRBT}(X; \Theta_{\text{DRBT}})$ using Algorithm 3

Split $X_{\text{transformed}}$ into training/testing according to the experimental protocol

Train detector D using training data (healthy-only for one-class settings)

Compute anomaly scores / predicted labels on test data using D

Output decisions and compute evaluation metrics (Accuracy/Precision/Recall/F1, etc.)

Table 1
Notation used in Section 2.3 (TAIW) and Section 2.5 (DRBT). Dimensions are shown in brackets.

Symbol	Type/Dimensions	Meaning
$x[n]$	real vector $[L]$	Vibration signal (one channel), length L samples
f_s	scalar [Hz]	Sampling rate
$\mathcal{M} \subset \mathbb{Z}$, $m_{\min}, m_{\max}$	index set/integers	Exponent indices for candidate windows ($w = 2^m$)
$\mathcal{W}$, w	Set/integer [samples]	Candidate window set; window length $w \in \mathcal{W}$
α	scalar $(0, 1)$	Overlap fraction for STFT (per window)
r	integer [samples]	Sample overlap, $r = \lfloor \alpha w \rfloor$
N_{FFT}	integer	FFT size used in STFT
$S_w[k, n]$	complex matrix $[K \times N_t]$	STFT for window w (freq bins K, time frames N_t)
$P_w[k, n]$	nonnegative matrix $[K \times N_t]$	One-sided power spectrogram, $P_w = \lvert S_w \rvert^2 + \varepsilon$
ε	small ≥ 0	Numerical clamp for logs/zeros in P_w
k, n	indices	Frequency/time-frame indices ($k = 0 \dots K-1$, $n = 0 \dots N_t - 1$)
$\mathcal{T}(\cdot)$	scalar functional	Theil index on a nonnegative vector
$T_w^{(t)}(n)$	scalar	Theil index on time slice $P_w[:, n]$
$T_w^{(f)}(k)$	scalar	Theil index on frequency track $P_w[k, :]$
$Q_t(w)$, $Q_f(w)$	scalars	Aggregated Theil scores over time/frequency
$Q_{tf}(w)$	scalar	Joint score for selection, $Q_{tf} = Q_t \cdot Q_f$
$w^\star$	integer [samples]	Selected window, $w^\star \in \arg\max_{w \in \mathcal{W}} Q_{tf}(w)$
X	real matrix $[N \times 2]$	2-D t-SNE embeddings (rows = frames), input to DRBT
μ	real vector $[2]$	Mean of rows of X
X_c	real matrix $[N \times 2]$	Centered data, $X - \mathbf{1}\mu^\top$
Σ	real matrix $[2 \times 2]$	Covariance of X_c
V, d	$[2 \times 2]$, $[2]$	Eigenvectors/eigenvalues of Σ (clamped by ε)
W	real matrix $[2 \times 2]$	Whitening transform, $W = V\,\text{diag}(d^{-1/2})$
X_w	real matrix $[N \times 2]$	Whitened data, $X_c W$
r_i, s_i	scalars	Radius $\lVert x_i^{(w)} \rVert_2$ and radial scale $s_i = \tanh(r_i)/(r_i + \varepsilon)$
X_{compact}	real matrix $[N \times 2]$	After radial compaction
$\hat{p}(x)$, h, $K(\cdot)$	density/scalars	KDE, bandwidth, and Gaussian kernel in modality check
$p_{\min}$	scalar	Minimum peak prominence for bimodality detection
$c_{\text{low}}, c_{\text{high}}$	scalars	Centers from 1-D k-means (k=2) on a feature
Δ, σ	scalars	Inter-center separation and feature std. for reshaping
γ	scalar	Max shift fraction (controls cluster shift)
α_{shift}	scalar	Actual shift factor in DRBT (notation to avoid clash with overlap α)
β, δ, ϵ, η	scalars	Contraction controls (bimodal case)
$\beta_{\text{nonbimodal}}$	scalar	Contraction for unimodal case
$X_{\text{transformed}}$	real matrix $[N \times 2]$	Final DRBT output used for detection/classification

Industrial motivation. In industrial environments, condition monitoring systems continuously collect data on vibrations, currents, and system variables from different machines. Although each machine is observed for only a brief period each day, the cumulative effect of monitoring numerous machines repeatedly daily and weekly over several years results in substantial data. Each dataset typically comprises multiple channels corresponding to different sensors, rendering the manual selection of appropriate spectrogram parameters for each machine and channel impractical. Consequently, computing thousands of spectrograms across machines, time intervals, and measurement channels necessitates an automated mechanism capable of adaptively determining suitable time, frequency, and joint time–frequency resolutions. The TAIW algorithm directly addresses this scalability challenge by providing a fully automatic data-driven criterion for selecting optimal window parameters, thereby facilitating the efficient large-scale monitoring of industrial fleet.

2.3.1. Definition

For a nonnegative vector $y \in \mathbb{R}^M_{\geq 0}$ with a mean value denoted as $\bar{y} = \frac{1}{M}\sum_{i=1}^{M} y_i$, the (mean-normalized) Theil index is defined by the equation

$$\mathcal{T}(y) = \frac{1}{M}\sum_{i=1}^{M} \frac{y_i}{\bar{y}} \log\Big(\frac{y_i}{\bar{y}}\Big), \tag{1}$$

where the convention $0 \log 0 = 0$ is applied, which is numerically clamped by ϵ. By applying (1) to the columns (time slices) and rows (frequency tracks) of the spectrogram $P_w[k,n] = |S_w[k,n]|^2 + \varepsilon$, we obtain

$$T_w^{(t)}(n) = \mathcal{T}\big(P_w[:,n]\big), \qquad n = 0,\dots,N_t - 1, \tag{2}$$

$$T_w^{(f)}(k) = \mathcal{T}\big(P_w[k,:]\big), \qquad k = 0,\dots,K-1. \tag{3}$$

The aggregate per-window scores are calculated as

$$Q_t(w) = \frac{1}{N_t}\sum_{n=0}^{N_t-1} T_w^{(t)}(n), \qquad Q_f(w) = \frac{1}{K}\sum_{k=0}^{K-1} T_w^{(f)}(k), \tag{4}$$

and the joint criterion for window selection is defined by

$$Q_{tf}(w) = Q_t(w)\,Q_f(w), \qquad w^\star \in \arg\max_{w\in\mathcal{W}} Q_{tf}(w). \tag{5}$$

This selection criterion favors window lengths that yield both compact time-domain impulses and spectrally coherent components, which are characteristic of fault-induced vibrations.

Algorithm 2 TAIW: Theil-Aware Intelligent Windowing

Require: $x[n]$, f_s, $\mathcal{M}$, N_{FFT}, α
Ensure: Best window $w^\star$ and curves $Q_t(w)$, $Q_f(w)$, $Q_{tf}(w)$
1: **for** each $m \in \mathcal{M}$ **do**
2: $w \leftarrow 2^m$; $r \leftarrow \lfloor \alpha w \rfloor$
3: Compute STFT $\rightarrow S_w$
4: Power spectrogram $P_w \leftarrow |S_w|^2 + \varepsilon$
5: Column Theil indices (time slices) $\rightarrow T_w^{(t)}(n) = \mathcal{T}(P_w[:,n])$
6: Row Theil indices (frequency tracks) $\rightarrow T_w^{(f)}(k) = \mathcal{T}(P_w[k,:])$
7: Aggregates $\rightarrow Q_t(w) = \frac{1}{N_t}\sum_n T_w^{(t)}(n)$, $Q_f(w) = \frac{1}{K}\sum_k T_w^{(f)}(k)$
8: Joint score $Q_{tf}(w) = Q_t(w)\,Q_f(w)$
9: Select best window: $w^\star \leftarrow \arg\max_w Q_{tf}(w)$

2.3.2. Interpretation and implementation notes

Physical rationale. The function $Q_t(w)$ quantifies the distribution of spectral energy across frequencies within each temporal frame, whereas $Q_f(w)$ assesses its persistence over time for each frequency bin. Consequently, maximizing their product, $Q_{tf}(w)$, facilitates a time–frequency representation in which transient impulses, such as those generated by inner-race bearing faults, are sharply localized while maintaining temporal coherence along the characteristic fault frequencies of the bearings. This approach ensures that the selected window effectively captures the impulsive fault energy without spreading it across adjacent frames.

Mathematical remark. Given that $\mathcal{T}(y) = \ln M - H(p)$, where $p_i = y_i / \sum_j y_j$, the criterion in (5) can be understood as the minimization of the joint Shannon entropy of time–frequency energy shares, thereby maximizing the concentration. From an information-theoretic perspective, the maximization of $Q_t Q_f$ indirectly enhances the joint Rényi entropy contrast between uniform and concentrated energy distributions, thereby ensuring balanced compactness along both axes.

Energy definition. The Theil index is applicable to non-negative values; therefore, the energy spectrogram $P_w = |S_w|^2$ was used. A slight numerical offset ε is introduced to prevent singularities in $\log(\cdot)$ for bins that approach zero.

Practical settings. For a real-valued vibration signal $x[n]$ sampled at a frequency of f_s Hz, the TAIW method determines a spectrogram window length from a discrete set of candidates, denoted as

$$\mathcal{W} = \{\, w_m = 2^m \mid m \in \mathcal{M} \subset \mathbb{Z},\ m_{\min} \leq m \leq m_{\max} \,\},$$

where $\mathcal{M}$ represents the set of permissible integer exponents. The parameter $m_{\min}$ (e.g., 6 for $w_{\min} = 64$) is selected to avoid excessively short windows, whereas $m_{\max}$ is chosen such that $w_{m_{\max}} \leq \lfloor L/4 \rfloor$, where L denotes the signal length. For each candidate $w \in \mathcal{W}$, the Short-Time Fourier Transform (STFT) $S_w[k,n]$ is computed using a Hamming window of length w and an overlap $\alpha \in (0,1)$. The candidate window set $\mathcal{W} = \{2^m \mid m_{\min} \leq m \leq m_{\max}\}$ should encompass physically relevant time scales, ranging from a few cycles of bearing rotation to the anticipated duration of transient impulses. The FFT size N_{FFT} is typically fixed to a power of two, ensuring an adequate frequency resolution without excessive memory usage.

Computational complexity. Let $W = |\mathcal{W}|$. Each candidate requires a single STFT evaluation with a cost of $\mathcal{O}(N \log N_{FFT})$ for a signal of length N, followed by two linear passes for the Theil evaluation. The total complexity is therefore

$$\mathcal{O}\big(W\, N \log N_{FFT}\big),$$

which are dominated by the FFT operations. In practice, W is small, and memory access costs dominate the runtime for long-signal processing.

2.4. Spectrogram compression via t-SNE/UMAP

Upon determining $(w^\star, \alpha^\star)$, the spectrogram $P_{w^\star}$ is recalculated, with each time frame considered a feature vector that characterizes a local vibration pattern. To achieve dimensionality reduction and reveal latent structures, a nonlinear embedding technique, such as t-SNE or UMAP, was employed.

$$X = \Phi(P_{w^\star}) \in \mathbb{R}^{N\times 2}, \tag{6}$$

where $\Phi(\cdot)$ denotes the embedding operator that preserves local topology.

Rationale. Linear methods, such as PCA, are inadequate for capturing the nonlinear manifolds generated by time–frequency variability. In contrast, t-SNE and UMAP maintain the proximity of adjacent frames while unfolding the curved trajectories resulting from varying rotational speeds and transient excitations. Because these embeddings preserve the local geometry but not the absolute scale, the resulting two-dimensional manifold often exhibits a non-uniform density or elongated clusters. This observation necessitates subsequent Distribution Reshaping via Bimodal Transformation (DRBT).

2.5. Distribution Reshaping via Bimodal Transformation (DRBT)

Distribution Reshaping via Bimodal Transformation (DRBT) constitutes the final stage of the proposed pipeline. It operates directly on the low-dimensional embeddings, typically the 2-D t-SNE or UMAP outputs, to homogenize feature densities, suppress heavy tails, and enhance cluster compactness without distorting manifold topology. Essentially, DRBT regularizes the distribution geometry of the embedded data, ensuring that classes or conditions, such as healthy and faulty states, occupy compact, Gaussian-like regions rather than irregular or elongated clusters.

Motivation. The low-dimensional manifolds derived from t-SNE or UMAP frequently exhibit uneven or multimodal densities owing to the local distance-preserving nature of the embedding, which is not maintained globally. These irregularities impede unsupervised separation, resulting in uneven cluster shapes and overlapping boundaries. From a diagnostic perspective, this results in ambiguous fault regions and an unreliable clustering. The DRBT addresses this limitation by implementing an adaptive data-driven reshaping procedure that balances global compactness and local separability.

Industrial relevance. In large-scale industrial monitoring systems, dimensionality-reduced data from multiple sensors and machines often exhibit significant variability owing to differences in load, speed, and environmental conditions. Without density homogenization, these embeddings produce highly irregular feature spaces, complicating automatic classification or anomaly detection. DRBT offers a lightweight deterministic correction that restores statistical regularity across datasets, enabling reliable unsupervised detection even when data are collected under diverse operating regimes.

Principle and overview. The transformation process was performed in two stages. The initial stage, termed *compactification*, involves centering and whitening the data, followed by nonlinear radial compaction, which mitigates heavy tails and reshapes arbitrary distributions into weakly bimodal forms. The subsequent stage, referred to as *modality-aware reshaping*, identifies bimodality along each feature axis and adaptively shifts and contracts the data to merge secondary modes and standardize the variance. Formally, given $X \in \mathbb{R}^{N\times 2}$, the overall mapping is represented as

$$X \;\longmapsto\; X_{\text{compact}} \;\longmapsto\; X_{\text{transformed}},$$

where X_{compact} denotes the whitened and radially compressed representation, and $X_{\text{transformed}}$ signifies the bimodality-adjusted output.

2.5.1. Step 1: Compactification

Centering. Each feature vector $x_i \in \mathbb{R}^2$ is initially centered as follows:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad X_c = X - \mathbf{1}\mu^\top. \tag{7}$$

Whitening. The covariance of the centered data is calculated using the equation

$$\Sigma = \frac{1}{N-1} X_c^\top X_c, \tag{8}$$

and is subsequently decomposed as

$$\Sigma = V \,\mathrm{diag}(d)\, V^\top, \tag{9}$$

where $d = (d_1, d_2)$ represent the eigenvalues. To ensure numerical stability, the eigenvalues are adjusted as follows:

$$d_k \leftarrow \max(d_k, \varepsilon),$$

and the whitening transformation is defined by

$$W = V \,\mathrm{diag}(d^{-1/2}), \qquad X_w = X_c W, \tag{10}$$

ensuring that $\mathrm{Cov}(X_w) \approx I$.

Radial compaction. To mitigate the influence of heavy-tailed samples, each whitened vector $x_i^{(w)}$ undergoes radial scaling as follows:

$$r_i = \|x_i^{(w)}\|_2, \tag{11}$$

$$s_i = \frac{\tanh(r_i)}{r_i + \varepsilon}, \tag{12}$$

$$x_i^{(\mathrm{comp})} = s_i\, x_i^{(w)}. \tag{13}$$

Consequently, the compactified dataset is represented as

$$X_{\mathrm{compact}} = [x_1^{(\mathrm{comp})}; x_2^{(\mathrm{comp})}; \ldots; x_N^{(\mathrm{comp})}]. \tag{14}$$

Given that $\tanh(r) \le 1$, substantial outliers are smoothly contracted towards the unit circle, whereas points proximate to the origin remain largely unaffected. This nonlinearity effectively attenuates high-energy deviations and stabilizes the density tails.

2.5.2. Step 2: Modality-aware reshaping

Each feature dimension of X_{compact} was independently analyzed to detect bimodality or excessive variance in the data.

Bimodality detection. For each feature $f \in \mathbb{R}^N$, a Gaussian kernel density estimate is calculated as

$$\hat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - f_i}{h}\right), \tag{15}$$

where

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2},$$

and $h > 0$ represents the bandwidth. Local maxima x^* are identified by

$$\frac{d}{dx}\hat{p}(x^*) = 0, \qquad \frac{d^2}{dx^2}\hat{p}(x^*) < 0,$$

and are retained only if their prominence satisfies

$$\mathrm{prominence}\left(\hat{p}, x^*\right) > p_{\min}.$$

A feature is classified as bimodal when more than one such local maximum is present.

Bimodal case (Reshape-worthy). Upon the detection of bimodality, the function f is divided using one-dimensional k-means clustering with $k = 2$, resulting in the identification of centers c_{low} and c_{high}, with a separation defined as $\Delta = c_{\mathrm{high}} - c_{\mathrm{low}} > 0$.

Cluster shift The higher cluster is adjusted towards the lower cluster as follows:

$$\alpha = \min\left(\gamma, \gamma\frac{\Delta}{\sigma}\right), \qquad f_i \leftarrow f_i - \alpha\Delta, \quad \forall i \in C_{\mathrm{high}}, \tag{16}$$

where $\sigma = \mathrm{std}(f)$ and $\gamma > 0$ serves as a constraint on the maximum shift fraction.

Contraction Subsequently, all samples are contracted around the mean $m = \mathrm{mean}(f)$:

$$\beta = \max\left(\delta, \epsilon - \eta\frac{\Delta}{\sigma}\right), \qquad f \leftarrow m + \beta(f - m). \tag{17}$$

In this context, $\delta > 0$ ensures a minimum level of contraction, $\epsilon > 0$ establishes the base contraction, and $\eta > 0$ modulates the contraction's reduction as the inter-cluster separation Δ/σ increases. This methodology effectively reduces the intermodal distance while preserving the stability of the overall variance.

Non-bimodal case (Mild scatter adjustment). In instances where strong bimodality is not observed and the feature variance surpasses a specified threshold, a uniform contraction is implemented as follows:

$$f \leftarrow m + \beta_{\text{nonbimodal}} \frac{(f - m)}{\sigma}. \tag{18}$$

The parameter $\beta_{\text{nonbimodal}} > 0$ regulates the extent of scatter reduction while ensuring the preservation of unimodality. Specifically, smaller values result in tighter compaction, whereas larger values maintain the variance.

Final output. Upon the reshaping of both features, the transformed data are reconstructed as

$$X_{\text{transformed}} = [f^{(1)}_{\text{reshaped}}, f^{(2)}_{\text{reshaped}}]. \tag{19}$$

Theoretical note. The reshaping operation ensures bounded variance, as expressed by the equation:

$$\text{Var}(f_{\text{reshaped}}) \leq \beta^2 \, \text{Var}(f), \tag{20}$$

which guarantees that compactness increases monotonically while maintaining the integrity of the topology. Minor perturbations in γ result in linearly proportional changes in variance, thereby affirming numerical stability in the presence of parameter uncertainty.

Computational aspects. All operations are algebraic and non-iterative, resulting in a computational complexity of $\mathcal{O}(N)$ per feature. In practical applications, the DRBT achieves convergence in a single iteration because the coefficients (α, β) are analytically derived from local density estimates rather than through iterative optimization.

Algorithm 3 DRBT: Distribution Reshaping and Bimodality Transformation

Require: Dataset $X \in \mathbb{R}^{N\times 2}$, ($\gamma$, δ, ϵ, η (w.r.t. Δ/σ), $p_{\min}$, σ_{th}, $\beta_{\text{nonbimodal}}$
Ensure:
- Transformed dataset $X_{\text{transformed}}$ with Gaussianized distribution and controlled variance

Centering: $\mu \leftarrow \frac{1}{N}\sum_i x_i, \quad X_c \leftarrow X - \mu$
Whitening: $(V, d) \leftarrow \texttt{eig}(\texttt{cov}(X_c)), \quad X_w \leftarrow X_c V \, \text{diag}(d^{-1/2})$
for each sample i **do**
  $r_i \leftarrow \|X_w(i,:)\|_2$
  $X_{\text{compact}}(i,:) \leftarrow \tanh(r_i)/(r_i + \varepsilon) \cdot X_w(i,:)$
for each feature f in X_{compact} **do**
  Estimate $\hat{p}(x)$ via Gaussian KDE
  Detect peaks x^*: $\hat{p}'(x^*) = 0$, $\hat{p}''(x^*) < 0$, $\text{prom}(\hat{p}, x^*) > p_{\min}$
  if #{peaks} > 1 **then**
    Partition $f \rightarrow C_{\text{low}}, C_{\text{high}}$ (k-means, $k = 2$)
    $\Delta \leftarrow c_{\text{high}} - c_{\text{low}}$
    $\alpha \leftarrow \min(\gamma, \gamma\Delta/\sigma)$
    Shift: $f_i \leftarrow f_i - \alpha\Delta, \quad \forall i \in C_{\text{high}}$
    $\beta \leftarrow \max(\delta, \epsilon - \eta\Delta/\sigma)$
    Contract: $f \leftarrow m + \beta(f - m), \quad m = \text{mean}(f)$
  else
    $f \leftarrow m + \beta_{\text{nonbimodal}}(f - m)/\sigma$
$X_{\text{transformed}} \leftarrow [f^{(1)}_{\text{reshaped}}, f^{(2)}_{\text{reshaped}}]$

2.6. Theoretical properties and stability

The proposed DRBT mapping constitutes a deterministic single-pass transformation that exclusively employs algebraic and monotonic operations. Although it does not incorporate iterative optimization, its theoretical properties can be rigorously delineated in terms of compactness and numerical stability. These attributes ensure that the transformation yields stable and bounded feature reshaping, making it suitable for unsupervised diagnostic application.

Bounded shift fraction (γ). The cluster-shift operation $f_i \leftarrow f_i - \alpha\Delta$ with

$$\alpha = \gamma \cdot \min(1, \Delta/\sigma), \qquad 0 < \alpha \leq \gamma,$$

is affine within each cluster and ensures that the displacement magnitude does not exceed $\gamma\Delta$. This mechanism prevents the overshoot for widely separated modes ($\Delta \gg \sigma$). A larger γ facilitates mode merging but poses a risk of overcompression, whereas smaller values maintain separation.

Table 2
Role of DRBT hyperparameters and typical practical ranges.

Parameter	Role	Typical Range/Effect
ϵ	Radial near-origin damping	$[0.01, 0.95]$; larger ↑ smooths small r
γ	Max shift fraction	$[0.05, 0.95]$; larger ↑ merges modes faster
δ	Min post-shift contraction	$[0.2, 0.7]$; prevents collapse ($\beta \geq \delta$)
ϵ	Base contraction cap	$(\delta, 1]$; larger ↑ preserves variance
η	Separation-aware softening	$[0.1, \epsilon - \delta]$; clipped by Δ/σ
$p_{\min}$	KDE prominence gate	$[0.01, 0.2]$ of max density; robustness vs. sensitivity
σ_{th}	Variance gate	$[0.1, 0.6]\times$ median σ per feature
$\beta_{\text{nonbimodal}}$	Unimodal scatter control	$(0, 1]$; smaller ↓ scatter more

Contraction gate (δ, ϵ, η). The subsequent contraction step,

$$\beta = \max\left(\delta,\ \epsilon - \eta\frac{\Delta}{\sigma}\right), \qquad f \leftarrow m + \beta(f - m),$$

is piecewise-affine with a Lipschitz constant β. Selecting $0 < \delta < \epsilon \leq 1$ ensures non-expansiveness ($\beta \leq 1$) and establishes a hard lower bound $\beta \geq \delta$ to prevent collapse. The antagonistic term $\eta(\Delta/\sigma)$ mitigates contraction when the two modes are distant, thereby avoiding excessive compaction of well-separated clusters. A safe condition ensuring $\beta \in [\delta, \epsilon]$ for all Δ/σ is

$$\eta \leq \epsilon - \delta.$$

KDE prominence and variance gates. The kernel-density prominence threshold $p_{\min}$ and the variance threshold σ_{th} function as stability gates, confining bimodality detection to peaks that are sufficiently strong and well-separated. This method reduces spurious branching caused by noise and weak oscillations, thus improving robustness at low signal-to-noise ratios (SNR).

Non-bimodal contraction. For unimodal features with $\sigma > \sigma_{\text{th}}$, the transformation

$$f \leftarrow m + \beta_{\text{nonbimodal}}\frac{(f - m)}{\sigma}$$

implements a linear contraction characterized by a Lipschitz constant $\beta_{\text{nonbimodal}} \in (0, 1]$. Smaller values of this constant enforce a stronger reduction in the scatter, whereas larger values preserve a greater degree of variance.

Invariances. The processes of centering and whitening confer translation invariance and scale normalization, while the radial transformation Φ is rotation-equivariant. Feature-wise reshaping was conducted on a whitened basis aligned with the principal axes. Consequently, the DRBT is insensitive to isotropic rescaling and exhibits moderate robustness to rotations.

Contraction and stability guarantees. The radial map $\Phi(x) = s(r)x$, where $s(r) = \tanh(r)/(r+\varepsilon)$, satisfies the condition $\|\Phi(x)\|_2 = h(\|x\|_2)$ with $0 \leq h(r) < r$ for $r > 0$ and $h(0) = 0$, thereby ensuring strict contractiveness in norm ($L_{\Phi} < 1$). Each feature-wise map is piecewise-affine with a Lipschitz constant $\max\{\beta, \beta_{\text{nonbimodal}}\} \leq 1$, which guarantees global non-expansiveness:

$$\|\mathcal{T}(X) - \mathcal{T}(Y)\|_F \leq L\,\|X - Y\|_F, \qquad L \leq \max\{L_{\Phi}, \beta, \beta_{\text{nonbimodal}}\} \leq 1.$$

Consequently, DRBT is numerically stable and robust to minor perturbations.

Functional convergence. Although the DRBT is designed as a single-pass mapping, repeated application $\mathcal{T}^r(X)$ results in a norm-decreasing sequence. According to the Banach fixed-point theorem, if at least one operation is strictly contractive ($\varepsilon > 0$, $\beta < 1$, or $\beta_{\text{nonbimodal}} < 1$), the mapping converges to a unique fixed point with a geometric rate $\leq L$. This establishes functional (non-iterative) convergence.

Overall guarantee. Given bounded X, thresholds $(p_{\min}, \sigma_{\text{th}})$, and parameters satisfying $0 < \delta < \epsilon \leq 1$, $\gamma > 0$, clipped η, and $\varepsilon > 0$, the DRBT mapping $\mathcal{T}$ is continuous, non-expansive, and globally bounded. Consequently, it (i) resists numerical blow-up, (ii) prevents cluster collapse, and (iii) ensures deterministic and stable feature reshaping under all operating conditions.

2.7. Hyperparameter selection rationale

The DRBT transformation introduces a constrained set of hyperparameters $\{\gamma, \delta, \epsilon, \eta, p_{\min}, \sigma_{\text{th}}, \beta_{\text{nonbimodal}}\}$, each governing a specific and bounded geometric operation within the algorithm. Unlike optimization-based approaches, DRBT does not minimize an explicit objective function; instead, its hyperparameters are selected according to *structural constraints imposed by the update rules themselves*, rather than empirical tuning towards a precise optimum.

Bimodality detection is exclusively determined by the minimum peak prominence, denoted as $p_{\min}$, which is applied to the kernel density estimate (KDE) of each whitened feature dimension. Given that the KDE peak magnitudes are interpreted in a relative, scale-consistent manner within each distribution, $p_{\min}$ is appropriately specified as a fractional threshold rather than an absolute value. Empirical observations indicate that values below $p_{\min} \approx 0.01$ result in spurious peak detection due to sampling noise, whereas values exceeding $p_{\min} \approx 0.2$ can suppress genuine secondary modes. Consequently, a moderate value within this range is selected to identify only structurally significant bimodality while maintaining robustness against noise.

The variance threshold σ_{th} does not contribute to bimodality detection; rather, it functions as a *unimodal scatter gate*. In the absence of bimodality detection, σ_{th} assesses whether a feature demonstrates sufficient dispersion to justify a controlled variance adjustment. This distinction ensures that bimodal and unimodal conditions are managed by separate, non-overlapping update rules, thereby preventing unstable mode separation when the marginal variance is low. The specified range of σ_{th} (Table 2) ensures numerical stability while maintaining meaningful scatter corrections.

The inter-mode shift parameter γ governs the maximum fraction of inter-cluster separation that can be translated during bimodal correction. Notably, the effective shift is normalized by the ratio Δ/σ and explicitly constrained by γ, ensuring both scale invariance and bounded translation. Consequently, γ is not a free scaling parameter: small values result in inadequate mode alignment, whereas values approaching unity lead to nearly complete merging of the detected modes. Therefore, a representative value in the upper-mid range was selected to achieve a meaningful correction without excessive collapses. Adaptive contraction is regulated by the triplet (ϵ, η, δ). Here, ϵ defines the nominal contraction cap, η modulates the contraction strength as a function of bimodal separation, and δ imposes a strict lower bound to prevent numerical collapse. The update rule $s = \max\left(\delta,\ \epsilon - \eta \frac{\Delta}{\sigma}\right)$ implies that δ must remain strictly positive, while ϵ must exceed δ to allow adaptive behavior. Selecting δ too small results in overcontraction, whereas substantial values suppress the corrective action of the DRBT. The adopted values lie safely within the feasible region defined by these constraints, thereby ensuring controlled compaction without loss of structural information. Finally, $\beta_{\text{nonbimodal}}$ regulates the unimodal high-scatter regime, where updates driven by bimodality are inapplicable to data. This parameter implements variance-normalized scaling around the feature mean, maintaining the overall distributional shape while preventing uncontrolled dispersion.

3. Experimental datasets and validation protocol

The proposed framework was validated using three publicly accessible datasets, each representing a diverse range of operating conditions and mechanical configurations. (i) the MOIRA–UNIMORE bearing dataset for independent cart systems (ICS) [33,76,77], (ii) the Politecnico di Torino bearing dataset [78], and (iii) Case Western Reserve University (CWRU) dataset [79]. Collectively, these datasets provide complementary validation domains encompassing translational–rotational linear drive systems, laboratory test benches, and conventional motor–bearing assemblies.

Data diversity. Although the MOIRA–UNIMORE dataset includes additional system variables, such as cart currents and position feedback [80,81], only the vibration channels were employed in this study to ensure methodological consistency across all datasets. Each dataset contained multiple fault severities and sensor configurations, enabling the cross-domain assessment of the proposed pipeline under varying loads, speeds, and fault conditions. A fundamental difference among these datasets lies in motion: in the Politecnico di Torino and CWRU test benches, bearings operate under purely rotational dynamics around a fixed shaft, whereas in the independent cart system, the bearings experience coupled translational and rotational motion as the cart moves along the rail [34,35,82–85]. The hybrid kinematic behavior of the ICS platform introduces time-varying excitation patterns, making it particularly suitable for evaluating the robustness of the proposed method under non-stationary operating conditions.

Validation protocol. For each dataset, the vibration signals were processed through the TAIW–DRBT pipeline without parameter tuning or dataset-specific adjustments. Identical hyperparameters were used throughout the evaluation of the generalization capabilities of the proposed framework. All analyses were conducted in MATLAB R2024b on a 12th Gen Intel® Core™ i9–12900HK (2.50 GHz) workstation with 32 GB RAM. The performance evaluation was based on cluster compactness, separation, and anomaly detection accuracy using one-class SVM and Isolation Forest detectors.

4. Results and discussion

This section describes the experimental validation of the proposed framework. The evaluation was conducted in three progressive stages: (i) a controlled analysis utilizing synthetic signals to verify the interpretability and accuracy of the TAIW mechanism; (ii) extensive validation across three publicly available vibration datasets to assess the generalization capability of the proposed pipeline; and (iii) a robustness assessment through sensitivity analysis, additive noise perturbation, and comparison with state-of-the-art methods. The results collectively demonstrate the proposed framework's ability to adaptively extract time–frequency representations, reshape latent feature distributions, and enhance class separability under stationary and non-stationary operating conditions.

4.1. Evaluation of TAIW on synthetic signals

Four synthetic waveforms ($f_s = 5\,\text{kHz}$, $T = 6\,\text{s}$) were developed to demonstrate the adaptive selection of spectrogram windows by the TAIW, which optimizes the joint time–frequency concentration (Figs. 2 and 3). Each signal exhibited a unique time–frequency morphology, facilitating evaluation under diverse stationarity and modulation conditions:

- **Signal 1 (FM tone + noise):** a slowly frequency-modulated sinusoid centered at $\sim 300\,\text{Hz}$ with additive Gaussian noise ($\sigma = 0.1$).
- **Signal 2 (bursty** $700\,\text{Hz}$ **+ noise):** six $\sim 80\,\text{ms}$ packets of a $700\,\text{Hz}$ tone embedded in low-level noise.
- **Signal 3 (step frequency + noise):** a tone transitioning from $450\,\text{Hz}$ to $800\,\text{Hz}$ at $T/2$ with mild noise ($\sigma = 0.04$).
- **Signal 4 (crossing chirps + noise):** two linear chirps, $350\,\text{Hz} \rightarrow 900\,\text{Hz}$ and $900\,\text{Hz} \rightarrow 350\,\text{Hz}$, intersecting mid-record.

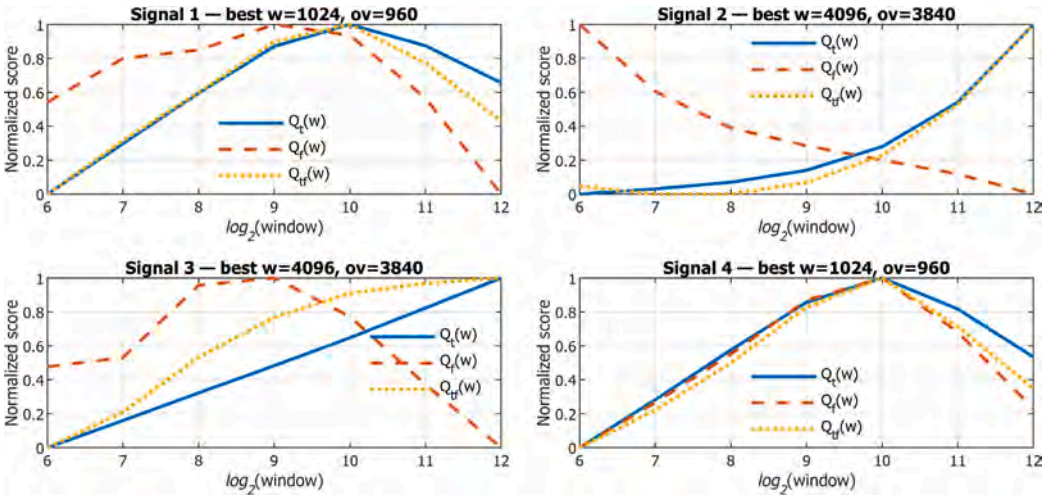

Fig. 2. TAIW coefficient curves for the four synthetic signals. Normalized $Q_t(w)$ (solid), $Q_f(w)$ (dashed), and $Q_{tf}(w)$ (dotted) versus $\log_2$ window length. Titles indicate the selected $w^\star$ and overlap $\mathrm{ov}^\star$.

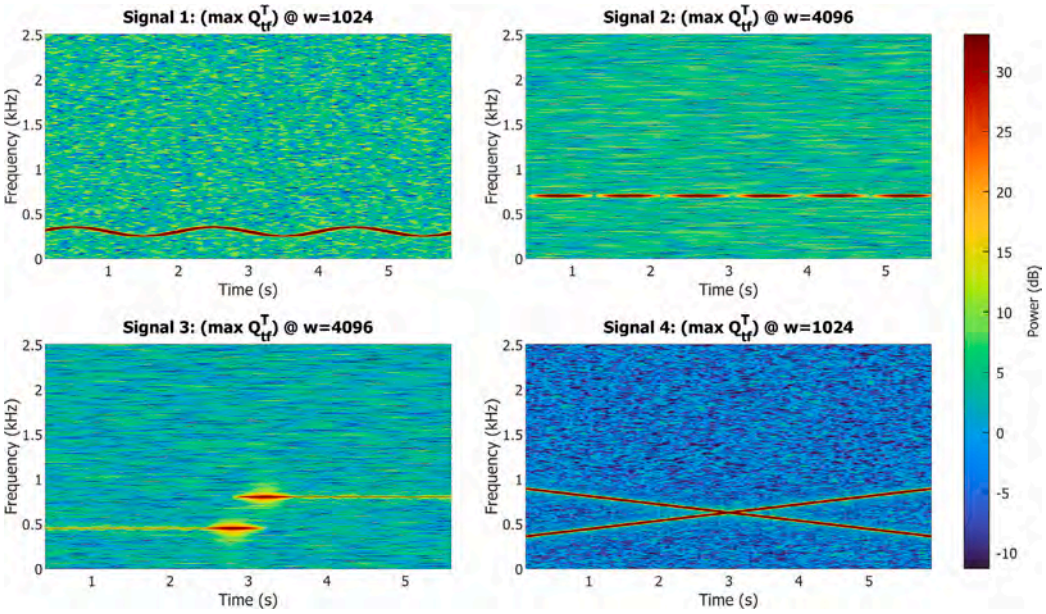

Fig. 3. Spectrograms computed using the TAIW-selected window and overlap for each synthetic signal.

Fig. 2 illustrates the normalized Theil-based scores $Q_t(w)$, $Q_f(w)$, and their product $Q_{tf}(w)$ as a function of $\log_2$(window). For *Signals 1* and *4*, the criterion reaches its maximum at $w = 1024$ samples (≈ 0.205 s) with $\mathrm{ov} = 960$ (15/16 of w), maintaining temporal detail in the slowly evolving ridge and chirp crossing area. For *Signals 2* and *3*, Q_{tf} prefers a longer window $w = 4096$ (≈ 0.819 s, $\mathrm{ov} = 3840$), enhancing the spectral sharpness within the burst train and the quasi-stationary segments preceding and following the frequency step.

The spectrograms shown in Fig. 3 corroborates these observations: shorter windows (1, 4) improve time localization, whereas longer windows (2, 3) enhance the narrowband structure. Thus, the Theil-aware criterion Q_{tf} aligns with expert visual intuition, selecting windows that reveal meaningful time–frequency organization without manual adjustment.

4.2. Application of the proposed pipeline on public datasets

The proposed TAIW–DRBT pipeline was applied to the three datasets introduced in Section 3.

Table 3
Class labels and fault specifications of the MOIRA–UNIMORE bearing dataset. 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 follow F{IR/OR}{width}; depth is not reported due to deep-groove geometry, and relative severity comparisons use width only.

4.2.1. MOIRA–UNIMORE dataset (Translational–rotational system)

System and motion dynamics. The testbed employed is an *ICS* based on the Beckhoff Extended Transport System (XTS) [30–32], a modular linear-motor platform where magnetized movers are propelled along a stator track via a contactless electromagnetic force (Fig. 4). Unlike belt- or chain-driven conveyors, ICS eliminates the need for intermediate mechanics, such as pulleys or gearboxes, thereby reducing frictional losses and maintenance requirements while enhancing motion precision and controllability [30–32,77, 86].

Architecture and mechanics. The system comprises a closed-loop 3000 mm track, which includes eight straight motor modules (250 mm each) and two 180° clothoid curves (500 mm each). A steel guide rail running parallel to the motor modules serves as a mechanical interface for the cart. Each cart utilizes three V-groove ball bearings — two identical upper bearings and a larger central lower bearing — where the outer races roll along the rail, whereas the inner races remain stationary. The mover's permanent magnets interact with the time-varying magnetic fields of the stator array, facilitating *contactless electromagnetic propulsion*. Nonetheless, mechanical contact is maintained at the bearing–rail interface, and a preload mechanism equalizes the contact forces and constrains lateral play.

Sensing and data acquisition. Five accelerometers were mounted on the rail: two triaxial sensors near the curved modules and three mono-axial sensors on the upper and lower straight sections of the rail. The signals were acquired using a Beckhoff ELM3602-0002 IEPE module (24-bit, up to 50 kHz). System variables such as the cart position, velocity, following error, and drive current were also recorded but excluded from the present analysis, which focused only on the vibration data. A detailed description of the setup, fault injection procedure, and dataset organization is provided in the MOIRA–UNIMORE Bearing Dataset publication [33].

Experimental conditions. The data analyzed correspond to *Experiment Type 2* of the MOIRA–UNIMORE dataset [33], wherein a single cart executed a back-and-forth motion over a 500 mm segment between the absolute positions of 250 mm and 750 mm along the curved module situated on the right side of the 3 m closed-loop track (Fig. 5). This section of the rail, highlighted in red, induces coupled translational–rotational motion as the cart navigates the curve, resulting in pronounced non-stationarity in both speed and load distribution across the three bearings. During each cycle, the cart accelerates, maintains a nominal velocity, and decelerates before reversing its direction, as depicted in the position–speed profile in Fig. 6. The cart speed varied between 0 mm/s and approximately 1000 mm/s. Signals were recorded under no fault, inner-race fault, and outer race fault conditions, with fault severities as described in Table 3. Each experiment was repeated for multiple cycles and realizations to ensure statistical consistency.

Vibration channels. The vibration channels of the MOIRA–UNIMORE bearing dataset correspond to the accelerometer placements shown in Fig. 5. Tri-axial accelerometers were installed on both sides of the rail near the curved modules, providing signals `X_Guide_Right`, `Y_Guide_Right`, `Z_Guide_Right`, and their symmetric counterparts `X_Guide_Left`, `Y_Guide_Left`, `Z_Guide_Left`. In addition, three mono-axial accelerometers were positioned on the upper and lower straight sections and on the top aluminum rod, resulting in channels `mono_PCB_Top`, `mono_PCB_Bottom`, and `mono_ifm_Top`. Here, "mono" signifies mono-axial accelerometers, while "PCB" and "ifm" denote the sensor manufacturers; "Top" and "Bottom" indicate their respective mounting positions along the track. Collectively, these nine vibration channels capture both local and global vibration transmission paths across the rail, offering complementary sensitivities to the translational and rotational motions of the cart. For further details on the data acquisition configuration, the reader is referred to [33].

Embedding analysis and quantitative evaluation. The t-SNE and UMAP embeddings were calculated for all monoaxial vibration channels and for the triaxial sensor on the right side of the rail, both before and after the DRBT transformation, to assess the cluster compactness and separability achieved by the proposed TAIW–DRBT pipeline. To ensure reproducibility, all t-SNE embeddings were generated using fixed, random seeds. Figs. 7–9 depict the comparative low-dimensional representations across channels.

Across all channels, the pre-transformation embeddings (upper panels in each figure) showed a significant overlap between healthy and faulty conditions. Following the application of the DRBT (lower panels), the data points were organized into distinct, well-defined clusters corresponding to fault types (`FIR` and `FOR`) and notch widths ranging from 0.25 mm to 1.5 mm for the inner race and 0.25, 0.5, and 2.0 mm for the outer race. This transformation reduces the intra-class dispersion and enhances the inter-class boundaries, thereby improving the interpretability and discriminative structure of the embedded feature space.

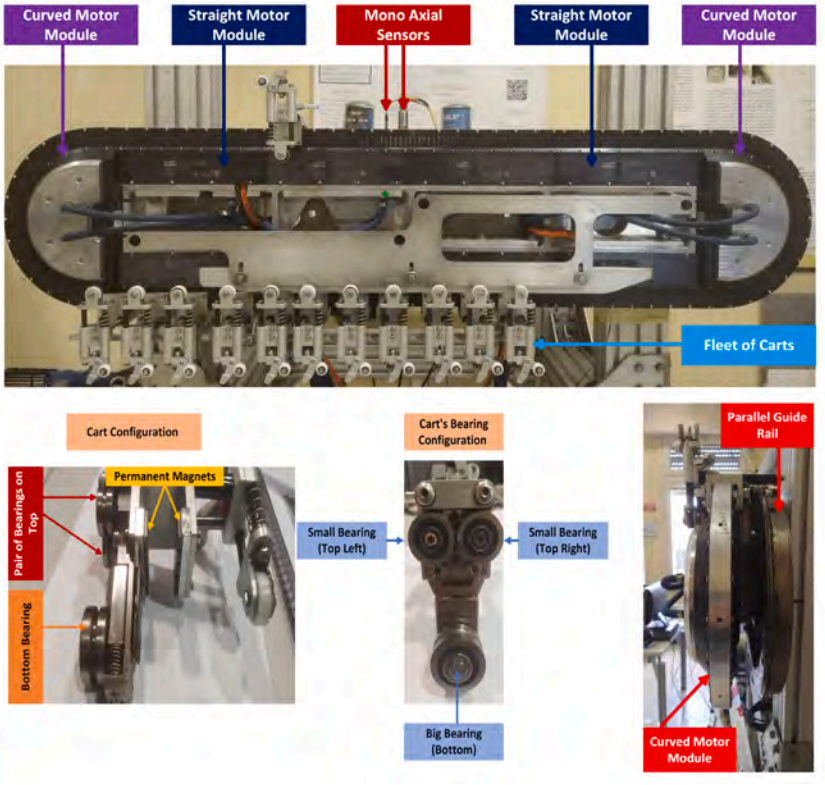

Fig. 4. An overview of the experimental setup utilizing the Beckhoff XTS platform is presented. **Top:** A comprehensive view of the closed-loop track is depicted, illustrating the modular configuration and the multiple carts affixed along the guide rail. **Bottom left:** A detailed side view of an individual 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 offers mechanical guidance.

Fig. 5. 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 comprised straight (C1–C4, A1–A4) and curved (D1–D2, B1–B2) modules, facilitating 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 situated on the curved section. The dimension labels specify the distances between the key module boundaries (in mm). The red arc on the right delineates the 500 mm back-and-forth reversal zone employed during testing to induce bidirectional motion.

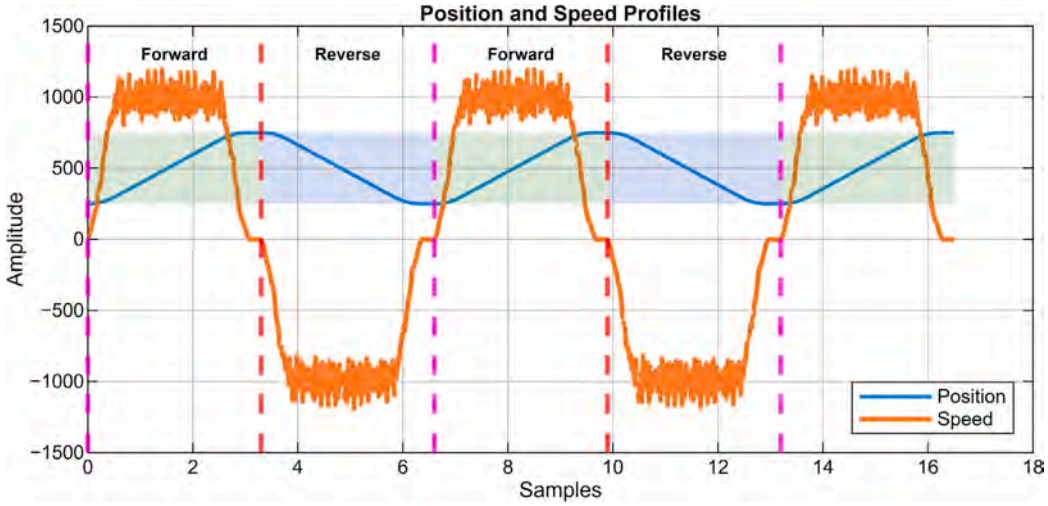

Fig. 6. The position and velocity profile of the cart during Experiment Type 2 is illustrated, with the forward and reverse phases indicated by shaded regions and dashed lines.

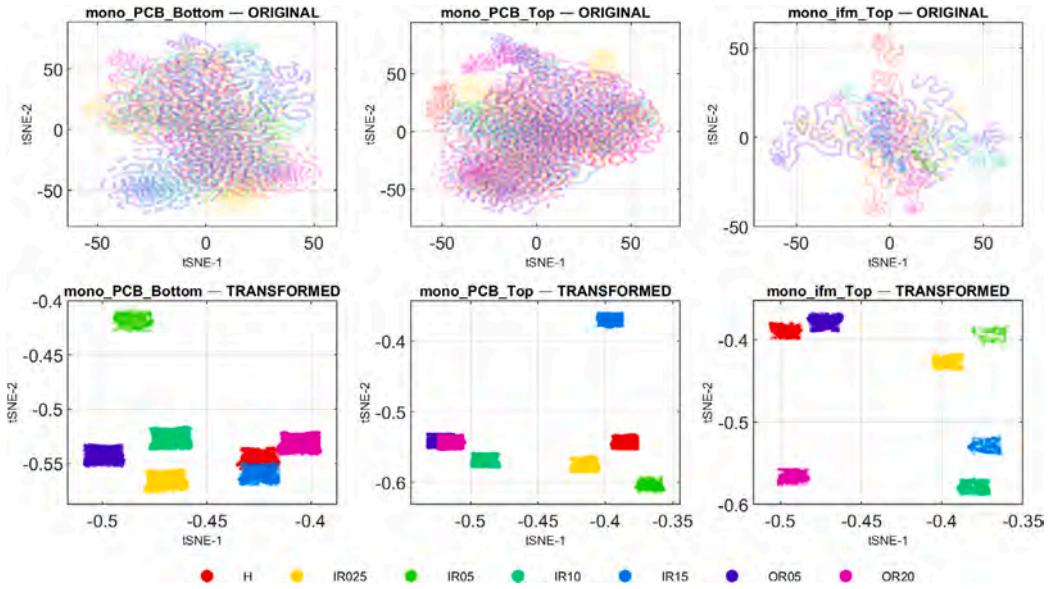

Fig. 7. The t-SNE embeddings, both prior to and following the DRBT transformation, for mono-axial accelerometer channels (`mono_PCB_Bottom`, `mono_PCB_Top`, and `mono_ifm_Top`) demonstrate significant improvements. Post-transformation, the clusters exhibit enhanced compactness and improved fault separability across all bearing conditions.

Quantitative validation using one-class SVM (OCSVM) and Isolation Forest (IF) is presented in Table 4 for t-SNE-reduced features. The post-transformation F1-scores exceeded 0.98 across nearly all channels, indicating a strong correlation between feature compactness and anomaly detection accuracy. The most significant performance improvements were observed in `Z_Guide_Right` and `mono_ifm_Top`, where the DRBT mapping mitigated the nonlinear distortions associated with speed variation.

To further evaluate the consistency of the DRBT transformation across various embedding techniques, Table 5 presents the results for UMAP-reduced features. Consistent with the t-SNE findings, all channels demonstrated significant performance enhancements post-transformation, with F1-scores nearing unity. This observation confirms that DRBT improves class compactness and separability in both stochastic (t-SNE) and manifold-based (UMAP) embeddings, highlighting its generalizability across nonlinear projection domains.

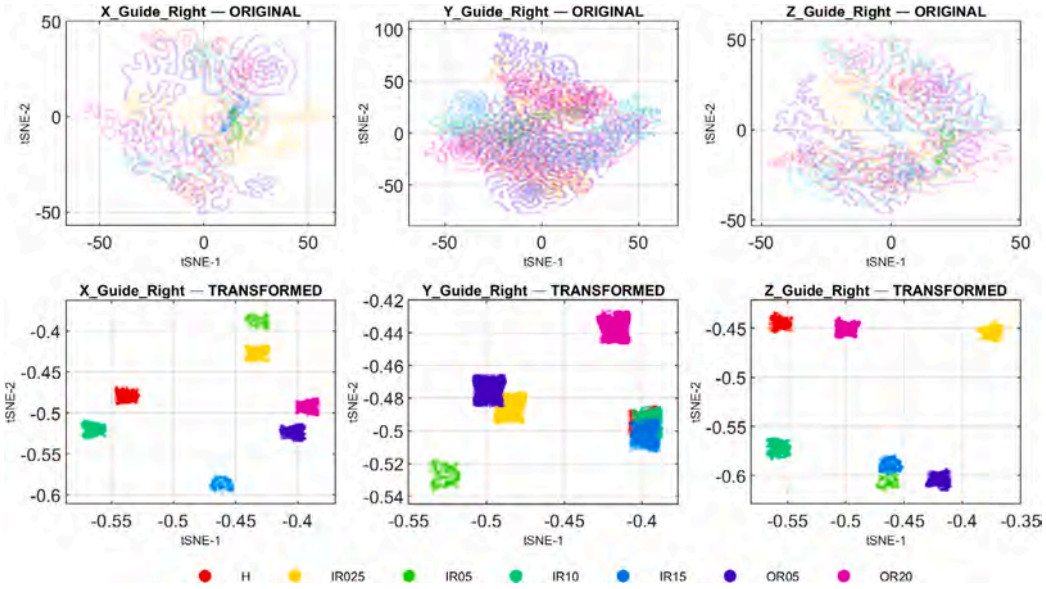

Fig. 8. The t-SNE embeddings for the right-side tri-axial accelerometer channels (`X_Guide_Right`, `Y_Guide_Right`, and `Z_Guide_Right`) were analyzed both prior to and following the DRBT transformation. The embeddings post-transformation exhibit compact, distinctly separable clusters, effectively replacing the overlapping distributions evident in the original feature space.

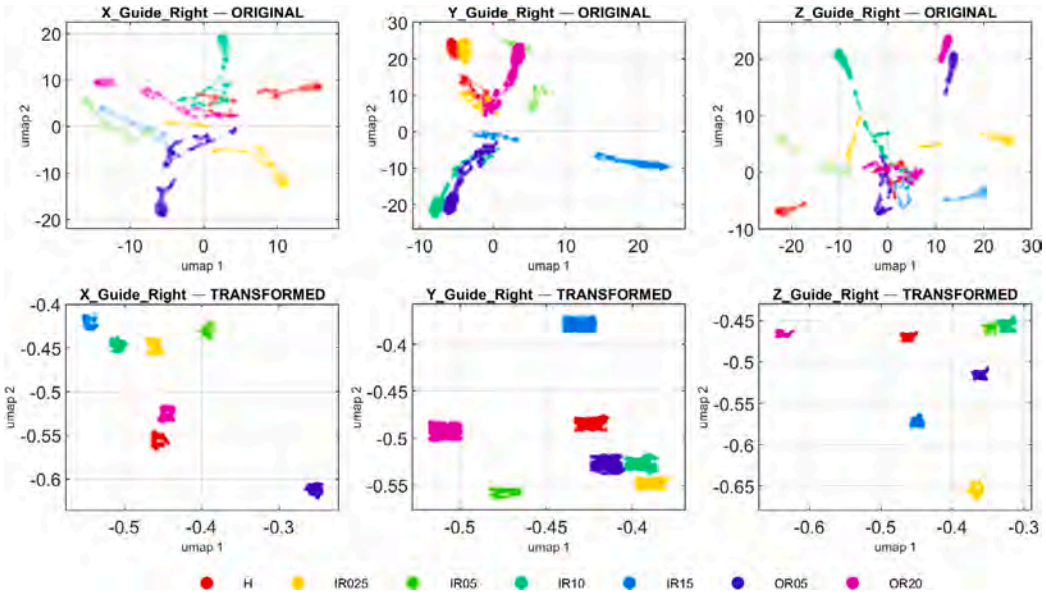

Fig. 9. The UMAP embeddings for the right-side tri-axial accelerometer channels, both prior to and following the DRBT transformation, demonstrate that the transformation consistently improves inter-class separation and intra-class compactness. This finding confirms the robustness of the transformation across nonlinear embedding techniques.

4.2.2. Politecnico di Torino dataset (Purely rotational system)

Experimental context. The proposed framework was further validated using the *Politecnico di Torino* bearing dataset [78], a publicly available benchmark for evaluating fault diagnosis algorithms under purely rotational conditions. Data acquisition was conducted using a dedicated test bench equipped with two triaxial IEPE accelerometers (totaling six channels) mounted on bearings B1 and B3, as depicted in the figure of text rif in [78]. Signals were synchronously captured using an OROS OR38 data acquisition system with 24-bit delta–sigma converters at a sampling frequency of 51.2 kHz. Each recording spanned 10 s, equating to 512,000 samples per channel, with no multiplexing applied to maintain inter-channel coherence.

Table 4
The F1-scores for each channel, both prior to and following the application of DRBT, were calculated using one-class SVM (OCSVM) and Isolation Forest (IF) on features reduced via **t-SNE**.

Channel	OCSVM		IF	
	Original	Transformed	Original	Transformed
X_Guide_Right	0.5039	0.9840	0.7080	0.9808
Y_Guide_Right	0.6669	0.9065	0.7224	0.9086
Z_Guide_Right	0.6602	0.9753	0.7768	0.9726
Mono_PCB_Bottom	0.5723	0.9881	0.6301	0.9858
Mono_PCB_Top	0.5223	0.9889	0.5641	0.9885
Mono_ifm_Top	0.7033	0.9892	0.7418	0.9822

Table 5
The F1-scores for each channel, both prior to and following the application of DRBT, were calculated using one-class SVM (OCSVM) and Isolation Forest (IF) on features reduced via **UMAP**.

Channel	OCSVM		IF	
	Original	Transformed	Original	Transformed
X_Guide_Right	0.5962	0.9814	0.7103	0.9782
Y_Guide_Right	0.6288	0.9135	0.7226	0.9051
Z_Guide_Right	0.6815	0.9732	0.7695	0.9697
Mono_PCB_Bottom	0.6031	0.9850	0.6524	0.9821
Mono_PCB_Top	0.5534	0.9858	0.5876	0.9862
Mono_ifm_Top	0.7260	0.9889	0.7539	0.9825

Table 6
Mapping of the original *Politecnico di Torino* dataset labels to the nomenclature employed in this study.

Original label	Current label	Defect description	Diameter (μm)
0A	H	No defect (Healthy)	–
1A	FIR450	Inner race indentation	450
2A	FIR250	Inner race indentation	250
3A	FIR150	Inner race indentation	150
4A	FR450	Roller indentation	450
5A	FR250	Roller indentation	250
6A	FR150	Roller indentation	150

The dataset comprises one healthy bearing and six faulted bearings, each exhibiting varying fault severities and locations. These were initially labeled 0A–6A, with defects induced by a controlled conical indentation using a Rockwell indenter applied on either the inner race or a single roller. The defect diameters were approximately 150, 250, and 450 μm, as listed in Table 6. To conform to the MOIRA–UNIMORE naming convention, the faults were relabeled as H, FIR450, FIR250, FIR150, FR450, FR250, and FR150, where "F" denotes a fault, "IR" signifies an inner-race defect, and "R" indicates a roller defect. All tests were conducted at five nominal shaft speeds (100, 200, 300, 400, and 500 Hz) and four static load levels (0, 1000, 1400, and 1800 N). Each bearing underwent an identical procedure, beginning at 0 N and 100 Hz, with incremental increases in load and speed until steady-state conditions were achieved. A total of 119 data files were collected (17 per class across all load–speed combinations). Each file contained six vibration channels corresponding to the two triaxial sensors. For the present analysis, the data were grouped by load level, yielding seven independent data subsets that were processed using the same TAIW–DRBT pipeline.

Embedding visualizations. The t-SNE embeddings for the *Politecnico di Torino* dataset were examined across various load levels to assess the impact of the DRBT transformation on feature distributions under different operating conditions. Figs. 10–13 present representative results for both fault families—inner-race (FIR) and roller (FR) defects, at load levels of 1000 N and 1400 N. In the pre-transformation embeddings (upper panels of each figure), the class boundaries are indistinct, with a significant overlap between fault severities. Following the application of the DRBT transformation (lower panels), the data points form compact, well-separated clusters for each defect class, indicating a marked enhancement in the intra-class cohesion and inter-class separability. This visual compactness was consistently observed across all shaft speeds (100–400 Hz), affirming the robustness of the transformation against variations in the rotational speed. Notably, the improvement was more pronounced for the roller fault family, in which the pre-transformation manifolds exhibited substantial nonlinear distortions owing to contact dynamics.

To further assess the consistency of the DRBT transformation across various nonlinear embedding techniques, UMAP projections were generated for the same load and speed configurations as those used in the t-SNE analysis. Figs. 14–17 illustrate the UMAP embeddings for the inner-race (FIR) and roller (FR) fault families at load levels of 1000 N and 1400 N, respectively. Consistent with the t-SNE results, the pre-transformation distributions (upper panels) exhibited broad entangled manifolds with significant overlap among fault classes, particularly at lower speeds (100–200 Hz). Following the DRBT transformation (lower panels), the data become notably more compact and distinct for both fault types and severities.

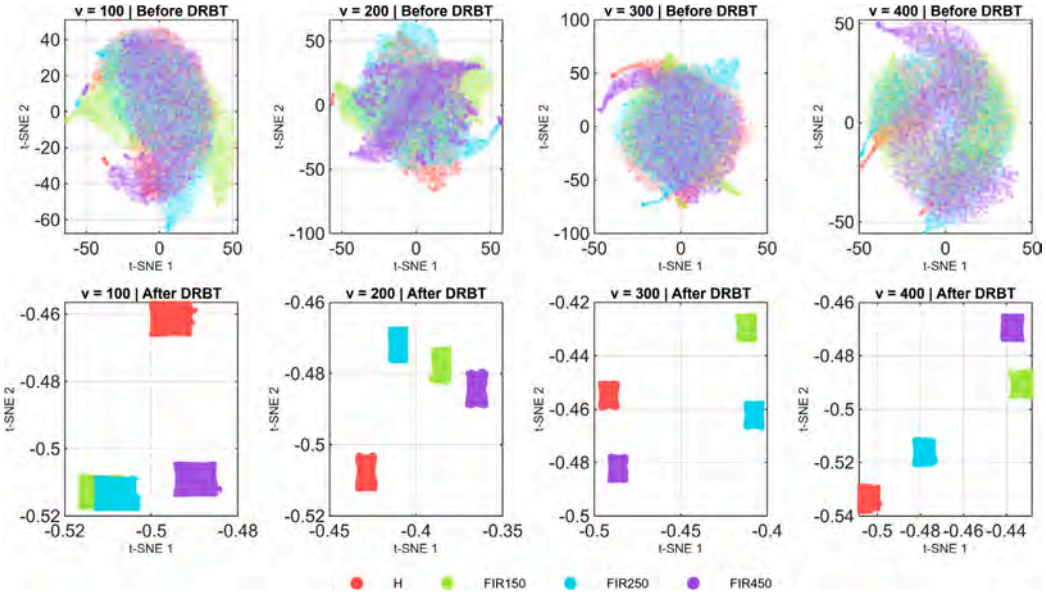

Fig. 10. The t-SNE embeddings are depicted before (top row) and after (bottom row) the DRBT transformation for the inner-race fault family (FIR) at a load level of **1000 N**. The transformation results in compact, linearly separable clusters across increasing shaft speeds (100–400 Hz).

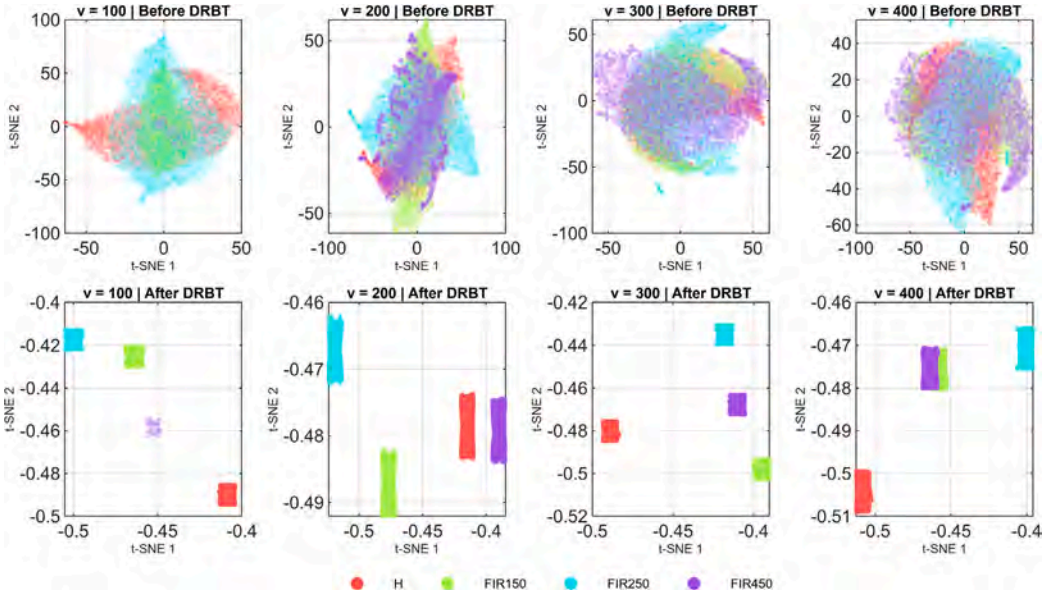

Fig. 11. The t-SNE embeddings, both prior to and following the DRBT transformation, for the inner-race fault family (FIR) at a load level of **1400 N**, demonstrate that the DRBT effectively reduces intra-class spread and enhances the separation among fault severities (FIR150--FIR450).

However, some residual overlap persists under specific operating conditions. For the FIR family, the FIR150 class partially overlaps with the healthy cluster at 300 Hz and 1400 N (Fig. 15), indicating that the low damage energy of this minor defect produces vibration signatures similar to the baseline noise at high rotational speeds. Similarly, in the FR family, limited overlap between the FR150 and healthy class is observed at 100 Hz for 1000 N and at 400 Hz for 1400 N (Figs. 16–17). Despite these minor ambiguities, the overall class topology remained consistent, with distinct clustering of higher-severity faults (FR450, FIR450) and improved alignment of intra-class densities. These findings confirm that the DRBT transformation generalizes effectively across different embedding manifolds, reducing the intra-class variance while preserving the intrinsic relationships between healthy and early stage fault conditions.

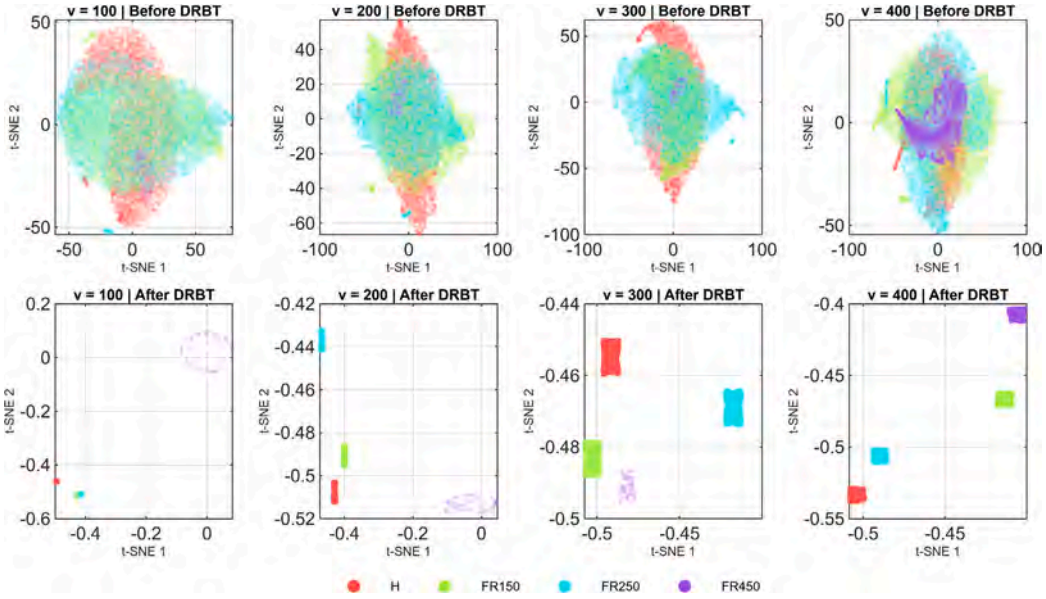

Fig. 12. The t-SNE embeddings are depicted before (top) and after (bottom) the DRBT transformation for the roller fault family (FR) at a load level of **1000 N**. This transformation effectively compacts the fault clusters and rectifies local manifold distortions observed in the pre-transformation space.

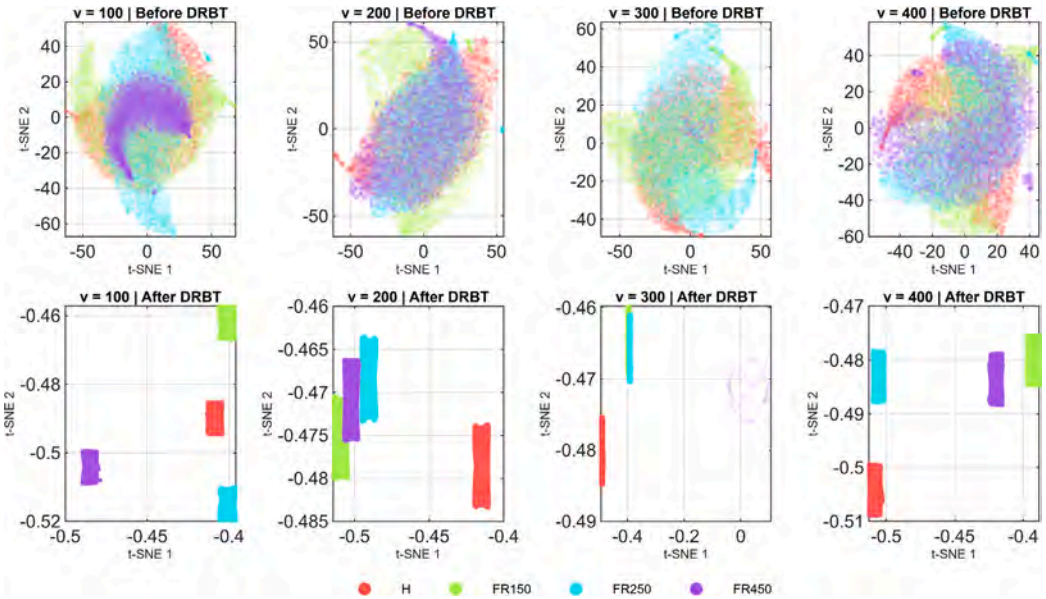

Fig. 13. The t-SNE embeddings, both prior to and following the DRBT transformation, for the roller fault family (FR) at a load level of **1400 N**, demonstrate that the DRBT consistently achieves compactness and distinct boundary formation across various shaft speeds. This transformation notably enhances the separability of smaller roller defects.

Quantitative evaluation. The results presented in Tables 7 and 8 indicate a consistent and significant enhancement in the anomaly detection performance after applying the proposed DRBT transformation. Across all speed and load combinations, both detectors, OCSVM and IF, exhibited notably higher F1-scores in the "After" columns compared to the "Before" columns, thereby confirming that the transformation improved class separability and reduced nonlinear overlap in the reduced feature spaces. For the t-SNE reduced features (Table 7), the baseline F1-scores before transformation exhibited considerable variability (approximately 0.12–0.67 for OCSVM and 0.24–0.72 for IF), reflecting their sensitivity to the operating conditions. Upon applying the DRBT, the F1-scores increase sharply to above 0.95 in nearly all instances, approaching unity at several frequencies (e.g., 300 Hz at 0 N). Thus, the

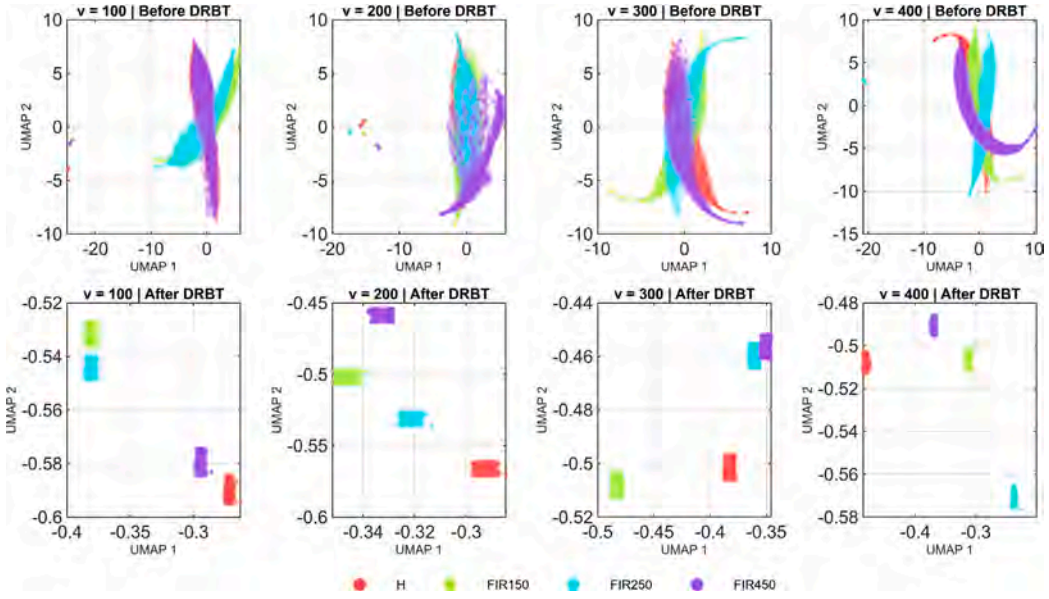

Fig. 14. UMAP embeddings are depicted before (top) and after (bottom) the DRBT transformation for the inner-race fault family (FIR) at a load level of 1000 N. The DRBT transformation markedly improves fault separability across all shaft speeds (100–400 Hz).

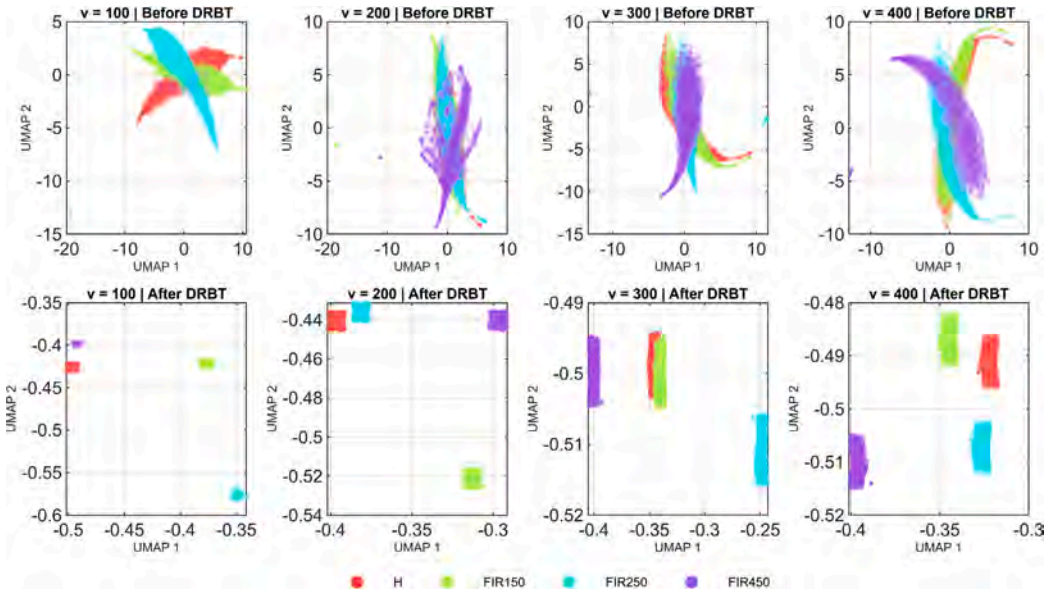

Fig. 15. UMAP embeddings, both prior to and following the DRBT transformation, are presented for the inner-race fault family (FIR) at a load level of 1400 N. A minor overlap between FIR150 and healthy data is discernible at 300 Hz; however, the overall cluster separability remains substantial.

proposed method effectively mitigates the distortions introduced by varying speeds and loads, resulting in nearly ideal one-class discrimination.

A similar pattern was observed for the UMAP features (Table 8). Before DRBT, there was moderate to high variability among speeds and loads, with OCSVM values ranging from 0.30–0.83 and IF values ranging from 0.40–0.83. Following the transformation, nearly all F1-scores exceeded 0.97, reaffirming that the reshaped manifolds consistently preserved the class-specific structure across the operating regimes of the pump.

Overall, these quantitative results confirm that the DRBT significantly enhances the uniformity and robustness of both detectors under non-stationary conditions, which is consistent with the qualitative improvements observed in the manifold visualizations.

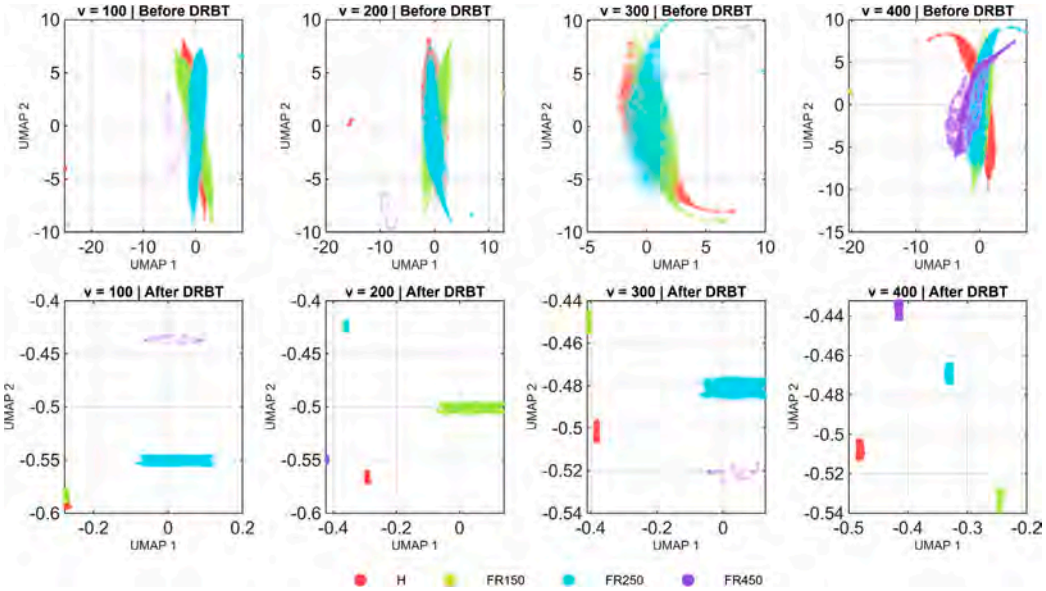

Fig. 16. UMAP embeddings are depicted prior to (top) and following (bottom) the DRBT transformation for the roller fault family (FR) at a load level of 1000 N. The FR150 and healthy clusters demonstrate minor overlap at 100 Hz, whereas other faults remain distinctly separable post-transformation.

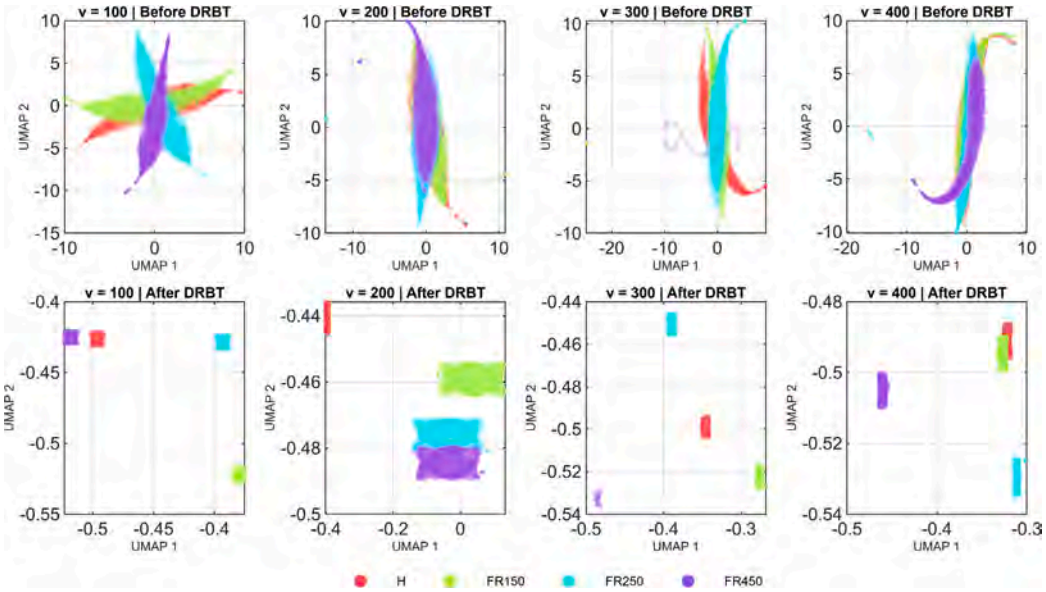

Fig. 17. UMAP embeddings, both prior to and following the DRBT transformation, are presented for the roller fault family (FR) at a load level of 1400 N. The FR150 class exhibits partial overlap with the healthy cluster at 400 Hz, whereas faults of greater severity (FR250, FR450) manifest as compact, isolated regions.

4.2.3. CWRU dataset (Purely rotational system)

Experimental context. To assess the generalization capability of the proposed framework, additional experiments were conducted using the *CWRU bearing dataset*, which represents a stationary rotational system with localized defects. To ensure methodological consistency and comparability with the existing literature, this study adheres to the diagnostic categorization proposed by [87], who systematically evaluated the entire CWRU dataset and classified each case based on its classification. In this benchmark, the dataset was divided into six categories — Y1, Y2, P1, P2, N1, and N2 — based on the clarity of fault manifestation in the time and frequency domains. Specifically, Y1 and Y2 correspond to fully diagnosable data exhibiting clear or moderately distorted fault signatures, whereas P1 and P2 denote cases with partial diagnosability. Conversely, N1 and N2 include data that are either not

Table 7
The F1 scores related to speed and load for OCSVM and IF on **t-SNE**-reduced features are presented both prior to and following the application of the proposed DRBT transformation. These results pertain to the Healthy class and the Inner Race Fault cases `FIR150`, `FIR250`, and `FIR450`.

Speed (Hz)	Load (N)	OCSVM		IF	
		Before	After	Before	After
100	0	0.5185	0.9853	0.5746	0.9849
	1000	0.3530	0.9764	0.4410	0.9757
	1400	0.5302	0.9553	0.5525	0.9546
200	0	0.6709	0.9705	0.6826	0.9859
	1000	0.4525	0.9723	0.4661	0.9722
	1400	0.4620	0.9710	0.4654	0.9706
300	0	0.4644	0.9920	0.4831	0.9925
	1000	0.2168	0.9745	0.2381	0.9749
	1400	0.4744	0.9769	0.5584	0.9767
400	0	0.1198	0.9767	0.2993	0.9759
	1000	0.4071	0.9760	0.5078	0.9754
	1400	0.5314	0.9751	0.7259	0.9746

Table 8
The F1 scores related to speed and load for OCSVM and IF on **UMAP**-reduced features are presented both prior to and following the application of the proposed DRBT transformation. These results pertain to the Healthy class and the Inner Race Fault cases `FIR150`, `FIR250`, and `FIR450`.

Speed (Hz)	Load (N)	OCSVM		IF	
		Before	After	Before	After
100	0	0.6918	0.9866	0.6742	0.9857
	1000	0.6241	0.9745	0.7528	0.9745
	1400	0.8230	0.9494	0.8130	0.9506
200	0	0.6389	0.9846	0.6483	0.9846
	1000	0.5205	0.9733	0.5359	0.9730
	1400	0.4059	0.9704	0.4073	0.9705
300	0	0.5766	0.9927	0.6508	0.9922
	1000	0.3610	0.9750	0.6588	0.9752
	1400	0.2980	0.9272	0.6629	0.9316
400	0	0.5763	0.9754	0.8331	0.9756
	1000	0.4296	0.9752	0.7863	0.9759
	1400	0.5831	0.9760	0.6802	0.9760

diagnosable owing to confounding mechanical effects or are indistinguishable from noise, containing only weak harmonic traces in spectral envelopes. Notably, the authors of the tutorial [87] recommended that future diagnostic algorithms be validated primarily on datasets belonging to the *P* and *N* categories, as they present the most challenging and realistic fault conditions for assessing the robustness of the diagnosis.

Following the recommendations of [87], this study focuses exclusively on the N1 and N2 subsets, which correspond to the most noise-dominated and diagnostically challenging cases in the CWRU dataset. Among these, the focus was restricted to the 48 kHz drive-end bearing fault recordings, which included ball, inner, and outer-race (centered) defects explicitly marked as N1 or N2 in the benchmark study. Accordingly, the dataset was relabeled such that all healthy signals were assigned the label `H`, inner race faults were grouped under `IR`, ball faults under `Ball`, and the centered outer race fault was grouped under `ORC`. This relabeling ensured a consistent labeling convention with the other datasets analyzed in this study and provided a rigorous test scenario that focused exclusively on the least diagnosable CWRU conditions.

Embedding visualizations. Fig. 18 illustrates the t-SNE embeddings at various rotational speeds (1730–1797rpm) before and after the application of the DRBT. Initially, the class manifolds exhibited significant nonlinear intermixing and spiral-shaped distortions, particularly between the *Healthy* and *Ball Fault* samples. This behavior is characteristic of vibration-induced amplitude modulation and stochastic excitation in rotating systems, resulting in complex, curved manifolds with nonlinear embeddings. After transformation, the manifolds undergo notable contraction and density equalization, forming compact and nearly orthogonal clusters with distinct boundaries. This transformation effectively reduces the curvature and local density bias, allowing the underlying fault structure to become more distinguishable. Consequently, all four classes became linearly separable, with significantly reduced intra-class spreads and balanced intercluster distances.

A similar progression was observed in the UMAP-reduced feature space (Fig. 19). Prior to DRBT, the UMAP projections displayed curvilinear trajectories and partial overlaps, particularly among the *Healthy*, *Ball*, and *Outer Race* states. These overlaps result from UMAP's continuous mapping nature of UMAP, which preserves the neighborhood topology but not necessarily the inter-manifold

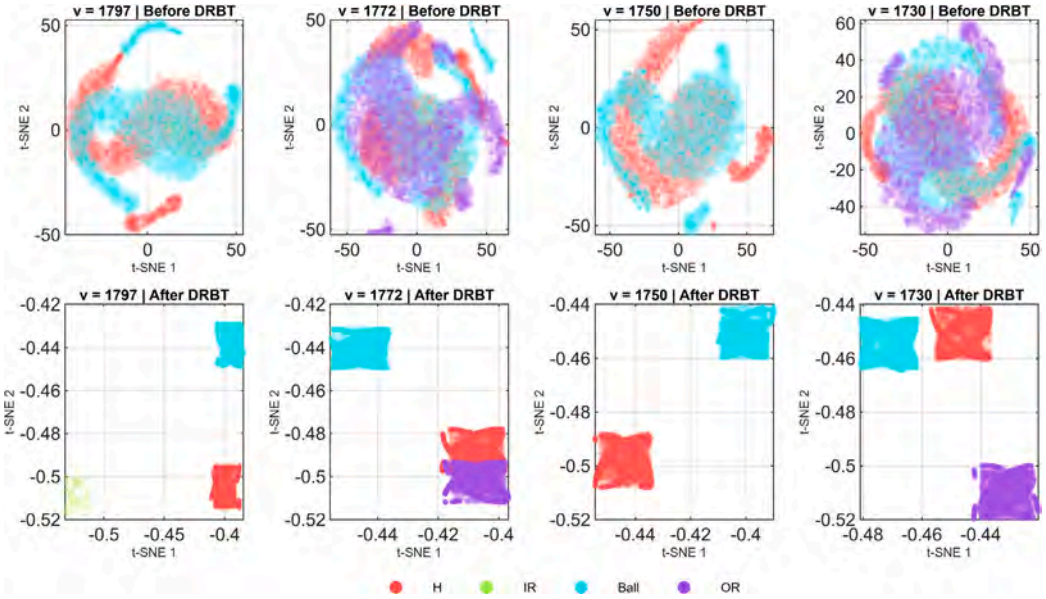

Fig. 18. The t-SNE embeddings of the *CWRU* dataset at various rotational speeds (v, expressed in rpm), both before and after the application of DRBT, illustrate that the proposed transformation effectively mitigates manifold curvature, equalizes density, and enhances the separability among the four health conditions.

margins. Following DRBT, the feature distributions become highly compact and evenly spaced, with reduced within-class variance and enhanced boundary distinctness. Notably, the *Inner Race* cluster, which previously exhibited partial continuity with the *Healthy* class, became fully separated, confirming the robustness of this change.

The consistent improvement across both t-SNE and UMAP projections demonstrates that the DRBT functions as a *post-manifold distribution equalizer*, effectively reshaping feature densities, regardless of the embedding technique. By flattening local density gradients and regularizing inter-cluster spacing, DRBT mitigates the geometric distortion introduced by nonlinear reductions and enhances class discriminability. These results confirm the suitability of the proposed method as a universal preprocessing stage for unsupervised fault diagnosis, thereby facilitating improved clustering and anomaly detection across variable fan operating speeds of the fan.

Quantitative evaluation. The results presented in Tables 9 and 10 indicate a consistent and significant enhancement in the anomaly detection performance following the application of the proposed DRBT transformation on t-SNE and UMAP embeddings for all examined rotational speeds in the CWRU dataset. Prior to transformation, the F1-scores of both the OCSVM and IF models displayed moderate variability and reduced separability between the healthy and faulty states, particularly at mid-range speeds (1750–1772 rpm), where the spectral signatures of different fault locations partially overlapped. Post-DRBT, the F1-scores exceeded 0.93 in all instances, with most configurations achieving or surpassing 0.96 for both t-SNE and UMAP embeddings.

These findings corroborate the visual trends observed in Figs. 18 and 19, where the DRBT transformation effectively enhanced the cluster compactness and separation across the four bearing conditions (healthy, inner race, ball fault, and outer race fault). The nonlinear mapping minimized the density distortions inherent to t-SNE and UMAP, ensuring that the manifolds corresponding to each fault class remained locally coherent and globally distinct. Furthermore, the stability of the F1-scores across varying speeds underscores the robustness of the proposed pipeline against non-stationary conditions and speed-dependent modulation, which is a critical challenge in real-world rotating machinery diagnostics.

In summary, the convergence of qualitative and quantitative evidence across three distinct datasets — MOIRA–UNIMORE, Politecnico di Torino, and CWRU — validated the proposed framework as a generalizable, data-driven diagnostic methodology that enhances the separability of bearing fault classes in nonlinear feature spaces without requiring dataset-specific tuning.

4.3. Robustness to additive noise

In real-world scenarios, vibration measurements are inevitably affected by environmental noise and the instruments used for measurement. To evaluate robustness, we introduced three distinct types of noise solely into the vibration channels, leaving the drive/command signals unchanged, and then repeated the entire process (transform → reduction → clustering) following the standard procedures in condition monitoring (see, e.g., [88]).

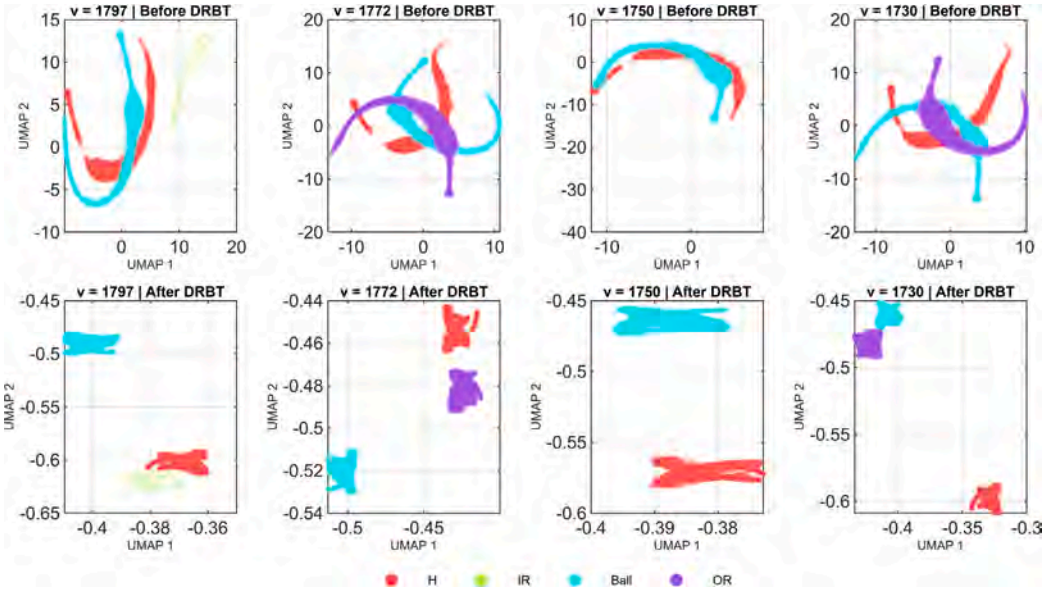

Fig. 19. UMAP embeddings of the *CWRU* dataset at various rotational speeds (v, expressed in rpm) are depicted both before and after the implementation of DRBT. The DRBT technique successfully converts the initially curved and overlapping manifolds into compact, well-separated clusters with uniform density.

Table 9
F1-scores corresponding to different rotational speeds (v, rpm) obtained using OCSVM and IF on **t-SNE**-reduced features, both before and after application of the proposed DRBT transformation on the CWRU dataset.

Speed (rpm)	OCSVM		IF	
	Before	After	Before	After
1730	0.5072	0.9593	0.5853	0.9604
1750	0.3601	0.9323	0.4441	0.9323
1772	0.4067	0.9017	0.5975	0.9102
1792	0.2957	0.9292	0.4807	0.9322

Note: All entries represent the F1-score for the specified speed and model configuration.

Table 10
F1-scores corresponding to different rotational speeds (v, rpm) obtained using OCSVM and IF on **UMAP**-reduced features, both before and after application of the proposed DRBT transformation on the CWRU dataset.

Speed (rpm)	OCSVM		IF	
	Before	After	Before	After
1730	0.6412	0.9641	0.9113	0.9640
1750	0.3543	0.9292	0.8145	0.9274
1772	0.6321	0.9647	0.8923	0.9646
1792	0.6269	0.9284	0.8766	0.9267

Note: All entries represent the F1-score for the specified speed and model configuration.

4.3.1. Noise models and injection protocol

(1) Broadband sensor/quantization noise (AWGN + uniform). To assess robustness against additive interference, zero-mean Additive White Gaussian Noise (AWGN) was injected into each vibration channel at three signal-to-noise ratio (SNR) levels, namely 10 dB, 20 dB, and 30 dB, representing severe, moderate, and mild noise conditions, respectively. In addition, a small zero-mean uniform disturbance was included to emulate the quantization effects and the low-amplitude bias jitter. The corrupted signal is expressed as

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{n}_g + \mathbf{u}_t, \quad \mathbf{n}_g \sim \mathcal{N}(0, \sigma_n^2), \quad u_t \sim \mathcal{U}(-a, a),$$

with $a = \rho\,\mathrm{RMS}(\mathbf{x}), \quad \rho = 0.02$. Here, σ_n^2 denotes the Gaussian noise variance, which is adjusted to achieve the target SNR. The SNR in decibels is defined as $\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}(P_{\mathrm{sig}}/P_{\mathrm{noise}})$, applied independently to each analysis window and channel.

(2) Narrowband, amplitude-modulated interference. To replicate sporadic tonal disturbances, such as structural modes and rotational tones, we developed a narrowband noise signal (Fig. 20).

$$n_{\mathrm{NB}}(t) = A_c\, e^{\delta t}\left(w(t) * h_{[f_1,f_2]}(t)\right),$$

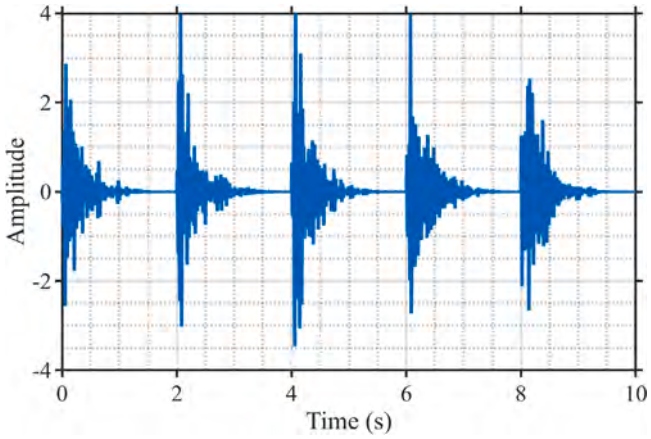

Fig. 20. Example of the amplitude-modulated narrowband interference used in Category 2 (f_s=50 kHz, band-pass [20, 40] Hz, pulse repetition f_p=0.5 Hz, exponential decay δ). The decaying bursts mimic intermittent tonal contamination aligned with the mechanical cycle.

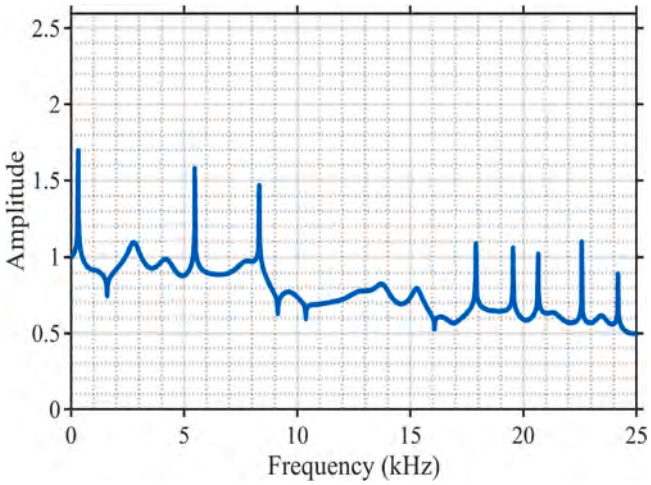

Fig. 21. Pseudo-FRF profile (Category 3) generated with N_p=20, peak_{RES}=2, $\text{peak}_{\text{ARES}}$=0.75, and ADR=0.5. This profile sets the target spectral envelope for the structural noise synthesized via IFFT.

where $w(t)$ represents zero-mean white noise, $h_{[f_1,f_2]}(t)$ is the impulse response of a 4th-order Butterworth band-pass filter with passbands $[f_1, f_2]$, $*$ signifies convolution, A_c is the carrier amplitude scale, and $\delta<0$ indicates the exponential decay of amplitude. The pulses were repeated over a period T_p, where $T_p=1/f_p$. In our experiments, we employed a sampling rate of f_s=50 kHz, with band edges at f_1=20 Hz and f_2=40 Hz, and a pulse rate of f_p=0.5 Hz (hence, T_p=2 s).

(3) Structural noise via Pseudo-FRF shaping. Structural noise was crafted using a spectrum that aligned with a pseudo-frequency response function (FRF) profile (Fig. 21) PFRF[r], where r denotes positive frequency bins, and the profile length is $M = \lfloor N/2 \rfloor$ for a time series containing N samples. This profile was designed as a smooth aggregation of sinusoidal terms with random phases to mimic broad resonant structures. It was then refined and tapered to emulate the resonances and high frequency roll-offs. The profile was subsequently adjusted to fit within the amplitude range $[\text{peak}_{\text{ARES}}, \text{peak}_{\text{RES}}]$ and multiplied by a linear decay factor ADR $\in (0, 1]$. A complex spectrum $H[r]$ was created with random phases and a magnitude shaped by the profile

$$|H[r]| \propto \text{PFRF}[r] \cdot \left(r_1[r]\, r_2[r] + c\right), \qquad \angle H[r] \sim \mathcal{U}[0, 2\pi),$$

where $r_1[r]$ and $r_2[r]$ are independent and identically distributed $\mathcal{U}(0, 1)$ variables that introduce minor random amplitude variations, and $c > 0$ is a bias to prevent zero bins. Hermitian symmetry was applied to both positive and negative frequencies to ensure that the inverse discrete Fourier transform (IDFT) produced real signal outputs. The time series was then derived as $\text{SN}(t) = \Re\{\text{IFFT}(H)\}$ followed by peak normalization. Unless stated otherwise, we used N_p=20 sinusoidal components to build the profile with peak_{RES}=2, $\text{peak}_{\text{ARES}}$=0.75, and ADR=0.5.

The impact of noise on the original low-dimensional embeddings is clearly illustrated in Figs. 22–24. Under significant interference, UMAP projections of the original features showed considerable expansion. Class manifolds become elongated, and

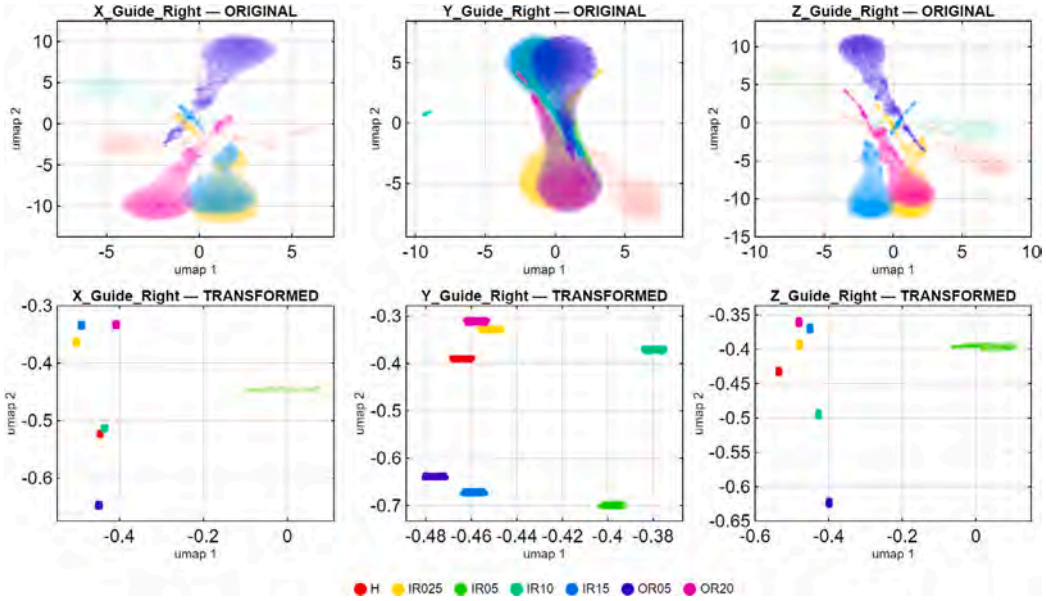

Fig. 22. UMAP embeddings of the MOIRA vibration channels (*X_Guide_Right*, *Y_Guide_Right*, *Z_Guide_Right*) were analyzed both prior to and following the application of DRBT under conditions of broadband AWGN combined with uniform noise **(10 dB SNR with 2% uniform jitter)**.

the inter-class overlap increases, particularly among intermediate fault severities (IR05–IR15). In addition, the geometric structure varied distinctly across the X, Y, and Z vibration channels. These distortions are most pronounced at lower signal-to-noise ratios (SNR) and gradually diminish as the noise level decreases. However, even at higher SNRs, residual density imbalances and class interpenetration remain evident in the original embedding. In contrast, after applying the DRBT technique, all the fault categories consistently condensed into compact, high-density clusters across different noise conditions. The transformed embeddings displayed minimal overlap, stable cluster centroids, and negligible sensitivity to noise intensity and channel direction. Importantly, there was no qualitative degradation in the class compactness or separability when transitioning from high- to low-SNR conditions. This indicates that the proposed transformation effectively mitigates broadband interference at the feature-manifold level.

A similar pattern emerged under narrowband amplitude-modulated interference (Fig. 25), which introduced decaying tonal bursts that interacted with different phases of cart motion. The original UMAP embeddings displayed long curved trajectories that reflected these transient resonances, with the Z-axis exhibiting the most pronounced arc-shaped distortions. Following DRBT, these nonlinear curves vanished, and each fault class occupied a well-separated and channel-consistent region in the t-SNE plot. The transformed space no longer reflects instantaneous noise energy or pulse timing; instead, it is dominated by condition severity, demonstrating DRBT's capability of the DRBT to eliminate temporally structured nonstationary noise.

Structural noise shaped by a pseudo-FRF envelope presents the most challenging scenario (Fig. 26). Prior to DRBT, the t-SNE embeddings collapsed into dense, overlapping clouds influenced almost entirely by the imposed spectral envelope, rendering categories such as OR05/OR20 and IR10/IR15 nearly indistinguishable. The geometry is dictated by a resonance-like noise profile rather than the underlying class differences. After DRBT, this effect is fully suppressed: clusters re-emerge with clear separation, uniform density, and no visible imprint of FRF-shaped spectral contamination. This underscores that the DRBT successfully removes structured resonant distortions and restores the class-centric geometry, even under severe spectral interference.

These observations confirm that the proposed framework effectively maintains robust, discriminative feature representations despite varying noise intensities. This highlights its suitability for vibration-based fault diagnosis in realistic industrial environments that are subject to stochastic disturbances.

4.4. Comparison with baseline and state-of-the-art transformation methods

To assess the efficacy of the proposed DRBT in comparison to widely utilized distribution-reshaping and density-homogenizing techniques within the reduced space, Fig. 27 offers a qualitative analysis based on UMAP embeddings following the application of each transformation directly to the two-dimensional UMAP coordinates. This configuration ensures an equitable and direct comparison of the capacity of each method to regularize a low-dimensional manifold, independent of any variations originating from high-dimensional preprocessing.

The baseline methods encompass classical power transforms (Box–Cox, Yeo–Johnson), the hyperbolic power transformation (HPT), rotation-based iterative Gaussianization (RBIG), probabilistic integral transforms (PIT-GMM, PIT-KDE), and the CDF-TS density homogenizing procedure. For reference, the unaltered UMAP embedding is depicted in the top-left panel.

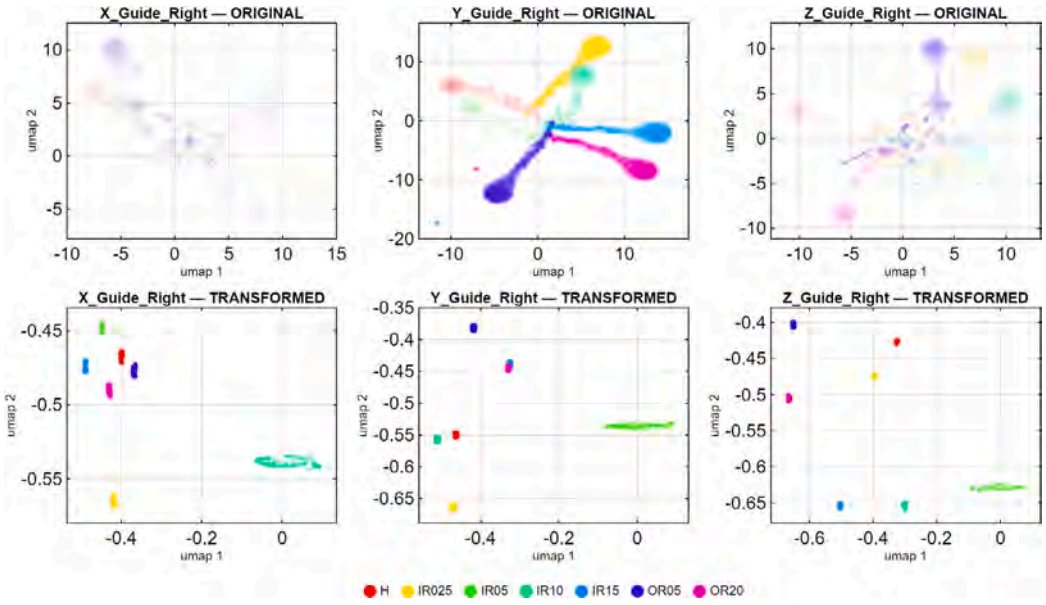

Fig. 23. UMAP embeddings of the MOIRA vibration channels (*X_Guide_Right*, *Y_Guide_Right*, *Z_Guide_Right*) were analyzed both prior to and following the application of DRBT under conditions of broadband AWGN combined with uniform noise **(20 dB SNR with 2% uniform jitter)**.

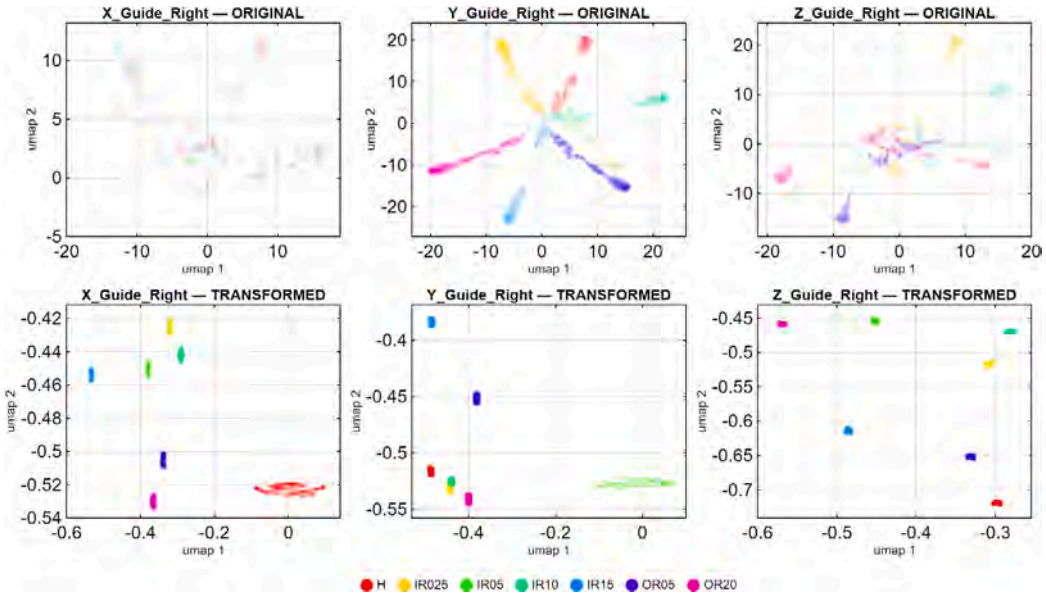

Fig. 24. UMAP embeddings of the MOIRA vibration channels (*X_Guide_Right*, *Y_Guide_Right*, *Z_Guide_Right*) were analyzed both prior to and following the application of DRBT under conditions of broadband AWGN combined with uniform noise **(30 dB SNR with 2% uniform jitter)**.

Across all baseline transformations, the reshaped UMAP embeddings continued to display non-uniform densities, elongated trajectory-like manifolds, and significant inter-class overlap. Box–Cox and Yeo–Johnson offer only marginal improvements over the raw UMAP and largely maintain its original curvature — most notably in the intertwined IR05–IR10–IR15 region. HPT results in excessive stretching along the first UMAP axis, effectively collapsing the geometry and diminishing class separability. While RBIG is effective for Gaussianization in higher dimensions, it yields a nearly homogeneous, structure-poor cloud when applied to 2-D UMAP data owing to iterative whitening and sample size sensitivity. PIT-GMM and PIT-KDE partially homogenize point densities but fail to eliminate nonlinear curvature and retain intertwined structures shaped by multimodal local neighborhoods. CDF-TS similarly flattens the densities but preserves the oscillatory distortions inherited from the original UMAP layout.

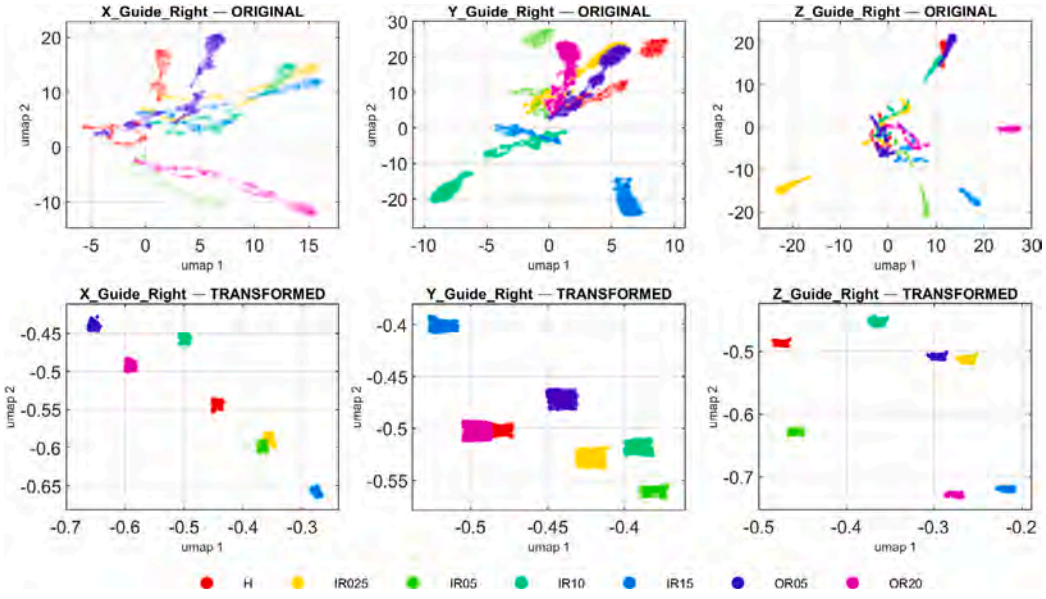

Fig. 25. UMAP embeddings, both prior to and following the application of DRBT, were examined under conditions of narrowband, amplitude-modulated interference (band-pass 20–40 Hz with exponentially decaying pulses). The initial embeddings exhibited curved and extended structures influenced by tonal bursts and non-stationary motion dynamics. The application of the DRBT effectively eliminated these geometric artifacts, resulting in the recovery of compact and channel-consistent fault clusters.

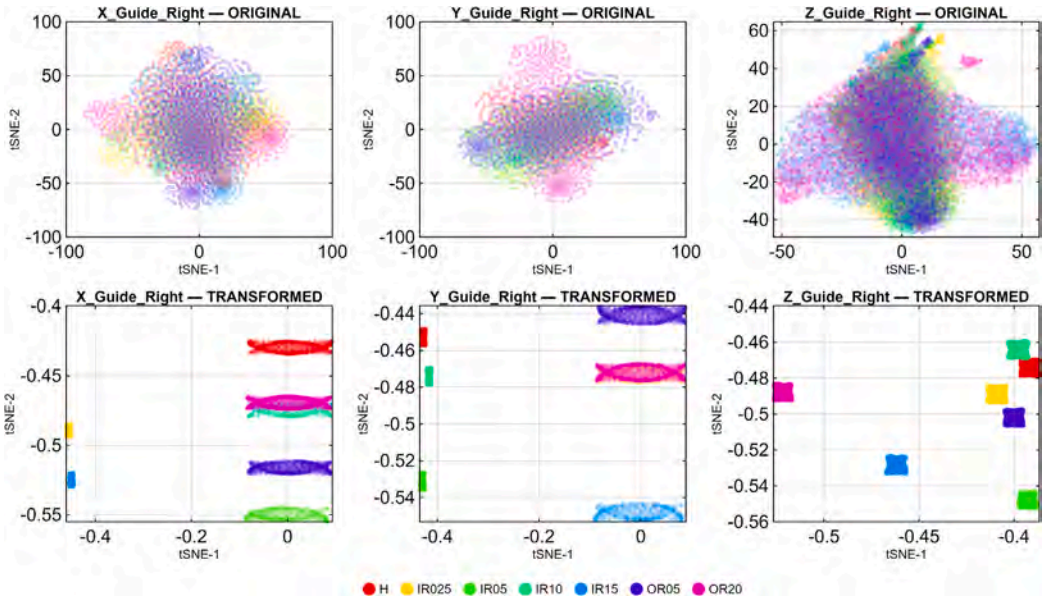

Fig. 26. The t-SNE embeddings, both prior to and after the application of DRBT under conditions of structural noise induced by a pseudo-FRF spectral envelope, exhibited distinct characteristics. Initially, the embeddings demonstrated significant class intermixing and resonance-driven distortions. In contrast, the application of DRBT resulted in class clusters that were uniformly dense and well-separated, indicating a high degree of robustness against structured spectral contamination.

In contrast, the DRBT generated highly compact, uniformly dense, and distinctly separated clusters for all fault categories and severities. In contrast to baseline methods, the DRBT eliminates the nonlinear curvature, equalizes the cluster density, and restores a class-centric, channel-consistent low-dimensional geometry. Notably, this is achieved without reintroducing the effects of operating conditions (e.g., speed trajectories or load-related drifts), which persist in other transformations. This facilitates accurate discrimination even between subtle conditions, such as IR05 vs. IR10 and OR05 vs. OR20. Overall, this comparative evaluation demonstrates that the DRBT is not merely a variance-stabilizing or Gaussianizing operator but a principled reshaping mechanism

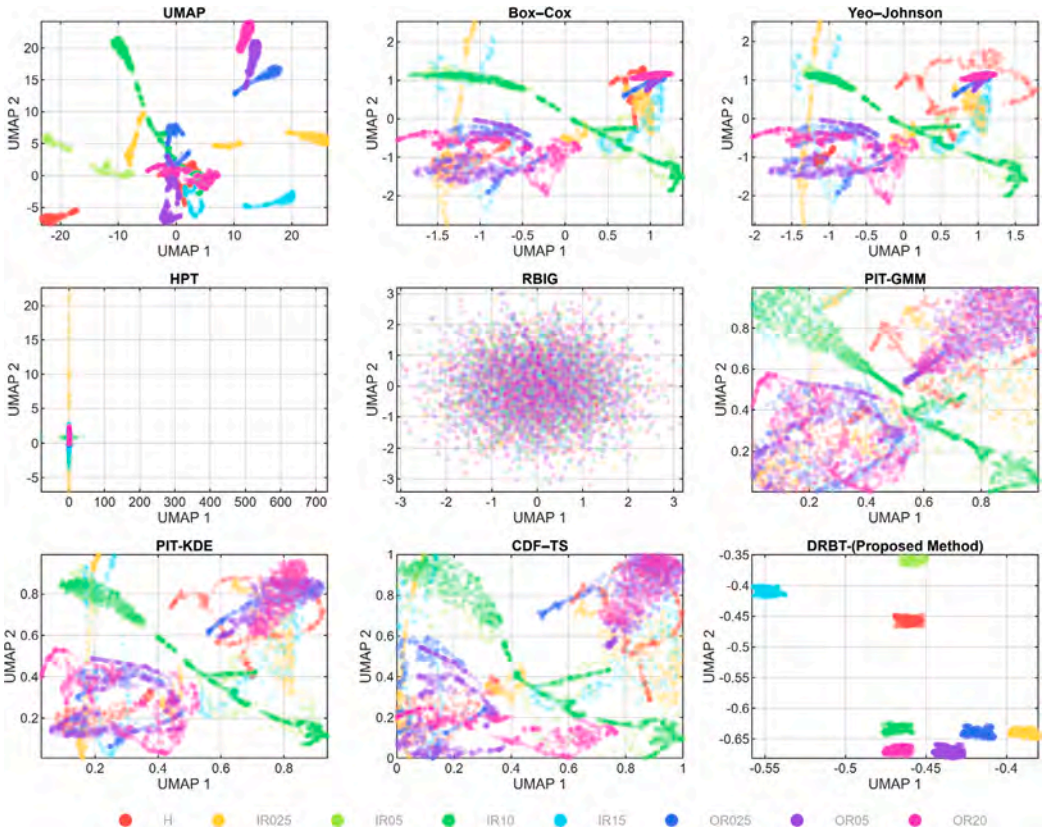

Fig. 27. The comparative assessment of distribution-reshaping techniques applied directly to two-dimensional UMAP embeddings reveals several key findings. The unprocessed UMAP projection showed elongated trajectories and significant interclass overlaps. Traditional power transformations, such as Box–Cox and Yeo–Johnson, effectively reduce marginal skewness but fail to address the nonlinear curvature inherent in the original manifold. The hyperbolic power transformation (HPT) results in excessively stretched embeddings with a collapsed class structure. In contrast, RBIG produces an almost homogeneous point cloud owing to iterative whitening in a low-dimensional space. Probability integral transform methods, including PIT-GMM and PIT-KDE, partially homogenize the density while preserving intertwined geometric patterns, and CDF-TS flattens densities while retaining oscillatory distortions from the original UMAP layout. In contrast, the proposed DRBT technique generates compact, uniformly dense, and well-separated clusters across all fault categories, demonstrating superior geometry reshaping and enhanced class discriminability.

specifically suited for condition-monitoring manifolds under nonstationary operations. Its ability to regularize already reduced embeddings distinguishes it from existing methods and results in low-dimensional representations that are reliably separable and informative.

4.5. Comparison with end-to-end deep anomaly detection methods

To complement the comparative analysis and address the relevance of recent state-of-the-art industrial methods, an additional evaluation was conducted using representative *end-to-end deep anomaly detection frameworks*. Vibration signals acquired from the same accelerometer channel of the MOIRA–UNIMORE bearing dataset were transformed into time–frequency representations using spectrogram images, which served as inputs to the deep models. Specifically, three widely adopted deep anomaly detection approaches were considered: EfficientAD [89], PaDiM [90], and Fully Convolutional Data Description (FCDD) [91]. All models were trained exclusively on healthy data and evaluated under identical experimental conditions to ensure a fair and consistent comparison. In parallel, an Isolation Forest classifier was applied to features obtained through the proposed transformation pipeline, namely *TAIW → t-SNE → DRBT*, enabling a direct comparison between fully end-to-end deep learning methods and a lightweight, statistically driven feature transformation framework.

Fig. 28 presents the row-normalized confusion matrices obtained for all four approaches. EfficientAD and PaDiM exhibited limited anomaly recall on this vibration-based industrial dataset despite achieving high normal-state recognition accuracy. This behavior highlights the difficulty in transferring generic vision-oriented deep anomaly detection models to non-image vibration signals without a domain-specific adaptation. FCDD achieves a more balanced detection performance owing to its fully convolutional one-class formulation, which is tailored to describe anomalies. Notably, the proposed *TAIW → t-SNE → DRBT* pipeline combined

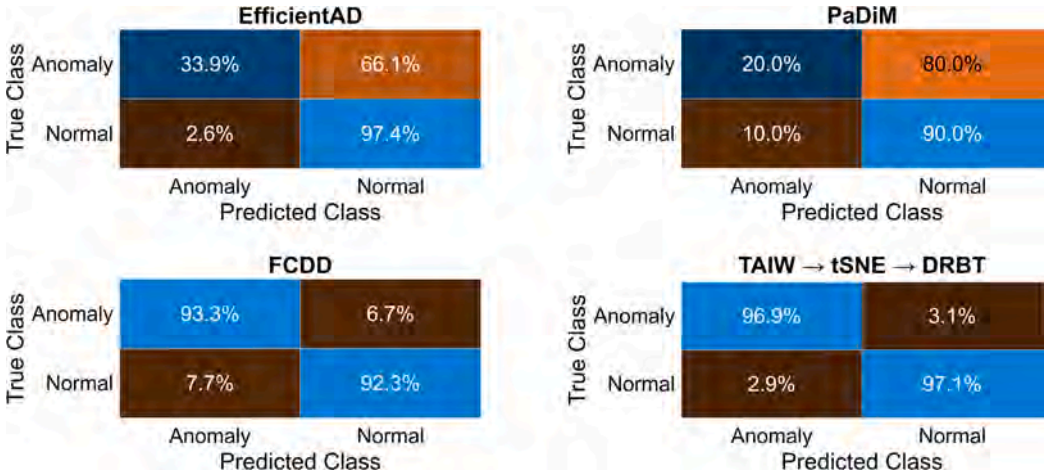

Fig. 28. Row-normalized confusion matrices comparing anomaly detection performance on the MOIRA–UNIMORE bearing dataset using spectrogram representations derived from a single vibration channel. From top left to bottom right: EfficientAD, PaDiM, Fully Convolutional Data Description (FCDD), and the proposed TAIW → t-SNE → DRBT pipeline followed by Isolation Forest. The comparison highlights the relative strengths and limitations of end-to-end deep anomaly detection methods versus the proposed statistically driven transformation-based approach in an industrial vibration monitoring context.

with the Isolation Forest demonstrates the most consistent and symmetric classification behavior, achieving high anomaly detection accuracy while maintaining a low false-alarm rate. These results indicate that although deep end-to-end anomaly detection methods provide strong baselines, the proposed statistically grounded transformation framework offers superior robustness, interpretability, and suitability for nonstationary vibration-based bearing-fault diagnosis in industrial environments.

4.6. Sensitivity analysis of DRBT hyperparameters under noisy conditions

To further validate the robustness of the proposed DRBT transformation and address concerns regarding the rigor of hyperparameter selection, a one-parameter-at-a-time (OPAT) sensitivity analysis was conducted under realistic noise conditions. This analysis was performed using the MOIRA dataset augmented with broadband measurement data. These disturbances combined zero-mean Additive White Gaussian Noise (AWGN) at an SNR of 10 dB with a small uniform perturbation to simulate quantization and bias jitter, as detailed in Section 4.3.1. Each DRBT hyperparameter was independently varied over the range $\{0.01, 0.95\}$, whereas all other parameters were maintained at their nominal values, as listed in Table 2. Following the application of the DRBT, anomaly detection was performed using a one-class SVM (OCSVM), and the performance was quantified using the F1-score. The resulting sensitivity curves are summarized in Fig. 29.

The minimum peak prominence parameter, denoted as $p_{\min}$, demonstrates a distinct threshold-dependent behavior. When $p_{\min}$ was set below approximately 0.2, consistently high F1-scores were achieved. In contrast, increasing the parameter beyond this threshold resulted in a significant decrease in performance. This observation corroborates the notion that excessively stringent prominence thresholds can suppress the structurally significant secondary modes in the presence of background noise. In contrast, moderate values effectively differentiate genuine bimodality from noise-induced artifacts. The inter-mode shift parameter γ exhibits a shallow U-shaped sensitivity profile. Insufficient values lead to inadequate alignment of the separated modes, whereas values approaching unity result in near-complete merging, potentially obscuring class separability. Notably, the degradation outside the optimal region is gradual, indicating a tolerance for moderate deviations and confirming that γ serves as a bounded control rather than a finely tuned gain.

The contraction lower bound δ exhibits a monotonic decline in performance as its value increases. This phenomenon aligns with its theoretical function as a safeguard against collapse: huge values impede meaningful contraction, thereby effectively turning off the corrective mechanism of the DRBT, whereas small positive values maintain stability while permitting adaptive compaction. In contrast, the nominal contraction cap ϵ, separation-aware modulation parameter η, variance gate σ_{th}, and unimodal fallback factor $\beta_{\text{nonbimodal}}$ exhibited largely flat sensitivity curves across the examined range. This invariance suggests that once the fundamental stability constraints are met, the performance of the DRBT is largely unaffected by the specific selection of these parameters. Such behavior is advantageous in practice, as it demonstrates that the transformation does not require precise hyperparameter tuning to achieve robust anomaly detection.

The OPAT sensitivity analysis conducted under the conditions of combined AWGN and uniform noise demonstrates that DRBT functions within extensive and well-defined stability regions. It was observed that only a limited subset of parameters, particularly $p_{\min}$ and δ, necessitated conservative selection to prevent degenerate behavior. In contrast, the remaining parameters exhibited significant robustness. These results substantiate the hyperparameter rationale outlined in Section 2.7 and affirm that the reported values represent typical operating points rather than fragile optima.

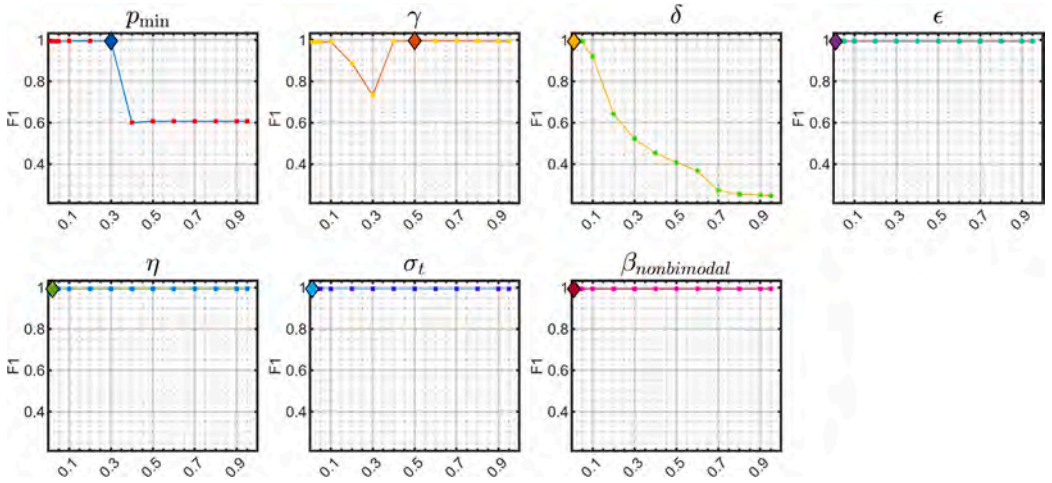

Fig. 29. One-parameter-at-a-time (OPAT) sensitivity analysis was performed on the DRBT hyperparameters under conditions of noise. The F_1 score of the OCSVM detector is depicted as a function of each parameter, with all other parameters held constant at their nominal values (Table 2). Sensitivity was evaluated using the MOIRA dataset, which was corrupted by additive white Gaussian noise (SNR = 10 dB) and supplemented with a minor uniform perturbation to simulate quantization and bias jitter. The solid diamond markers indicate the maximum F1 score for a specific parameter, whereas the solid squares denote the operating points used in the experimental evaluation. Parameters related to bimodality detection and contraction bounds (e.g., $p_{\min}$ and δ) exhibited the expected performance degradation when they were outside their permissible ranges. In contrast, scale-normalized parameters (γ, ϵ, η, σ_t, and $\beta_{\text{nonbimodal}}$) demonstrated considerable stability, thereby confirming the robustness of the proposed transformation.

5. Conclusion

This study presents a lightweight and fully unsupervised diagnostic framework that integrates Theil-index-driven time–frequency window selection (TAIW) with modality-aware distribution reshaping transformation (DRBT) for the detection of bearing faults under non-stationary conditions. TAIW facilitates automatic window selection, thereby eliminating the need for manual spectrogram tuning, whereas DRBT restructures density-imbalanced embeddings into compact and well-separated clusters without requiring dataset-specific optimization.

Experimental validation using the MOIRA–UNIMORE Independent Cart System dataset demonstrated that the proposed pipeline consistently transforms weakly separable embeddings into highly discriminative feature spaces, resulting in substantial improvements in the one-class detection performance. Cross-dataset evaluation on the Politecnico di Torino and CWRU benchmarks further confirmed the general applicability of the framework across fundamentally different bearing kinematics and operating regimes.

Robustness analyses under conditions of additive white Gaussian noise, uniform perturbations, narrowband interference, and structured disturbances indicated stable performance across realistic industrial conditions. Sensitivity analysis revealed wide stability regions, with only limited parameter constraints necessary to avoid spurious bimodality detection or excessive contractions.

In summary, the integrated TAIW–tSNE/UMAP embedding–DRBT pipeline offers a practical and versatile unsupervised solution for condition monitoring of non-stationary industrial systems. Future research should focus on real-time deployment, multichannel fusion, and integration with emerging self-supervised paradigms.

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 as potential competing interests: Abdul Jabbar reports financial support was provided by 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.

Acknowledgments

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

Data availability

The data used in this study is publicly available and published in a separate article.

References

[1] S.M. Tayyab, S. Chatterton, P. Pennacchi, Intelligent defect diagnosis of rolling element bearings under variable operating conditions using convolutional neural network and order maps, Sensors 22 (5) (2022) 2026, http://dx.doi.org/10.3390/s22052026.
[2] M.H. Farhat, X. Chiementin, F. Chaari, F. Bolaers, M. Haddar, Order-based identification of bearing defects under variable speed condition, Appl. Sci. 11 (9) (2021) http://dx.doi.org/10.3390/app11093962, URL https://www.mdpi.com/2076-3417/11/9/3962.
[3] Z. Liu, D. Peng, M.J. Zuo, J. Xia, Y. Qin, Improved Hilbert–Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings, ISA Trans. 125 (2022) 426–444, http://dx.doi.org/10.1016/j.isatra.2021.07.011, URL https://www.sciencedirect.com/science/article/pii/S0019057821003785.
[4] T. Wu, J. Chen, C. Wang, Characterization of gear faults in variable rotating speed using Hilbert-Huang transform and instantaneous dimensionless frequency normalization, Mech. Syst. Signal Process. 30 (2012) 103–122, http://dx.doi.org/10.1016/j.ymssp.2012.01.022, URL https://www.sciencedirect.com/science/article/pii/S0888327012000234.
[5] Y. Jiang, J. Zhou, X. Wu, T. Liu, X. Liu, Vision-based bearing fault diagnosis under non-stationary conditions using optimized short-time concentrated transform method, Reliab. Eng. Syst. Saf. 262 (2025) 111183, http://dx.doi.org/10.1016/j.ress.2025.111183, URL https://www.sciencedirect.com/science/article/pii/S0951832025003849.
[6] B. Peng, Y. Bi, B. Xue, M. Zhang, S. Wan, A survey on fault diagnosis of rolling bearings, Algorithms 15 (10) (2022) http://dx.doi.org/10.3390/a15100347, URL https://www.mdpi.com/1999-4893/15/10/347.
[7] S.R. Saufi, Z.A.B. Ahmad, M.S. Leong, M.H. Lim, Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review, IEEE Access 7 (2019) 122644–122662, http://dx.doi.org/10.1109/ACCESS.2019.2938227.
[8] P. Luo, Z. Liu, Unsupervised bearing fault diagnosis using masked self-supervised learning and swin transformer, Machines 13 (9) (2025) http://dx.doi.org/10.3390/machines13090792, URL https://www.mdpi.com/2075-1702/13/9/792.
[9] Z. Shu, D. Peng, H. Wang, C. Mao, Z. Yu, Time-frequency perception guided multi-level contrastive learning for rotating machinery fault diagnosis, Expert Syst. Appl. 293 (2025) 128664, http://dx.doi.org/10.1016/j.eswa.2025.128664.
[10] Q. Zhou, W. Ma, Y. Zhang, J. Guo, Bearing fault diagnosis for variable working conditions via lightweight transformer and homogeneous generalized contrastive learning with inter-class repulsive discriminant, Eng. Appl. Artif. Intell. 139 (2025) 109548, http://dx.doi.org/10.1016/j.engappai.2024.109548.
[11] K. Alqunun, M.B. Bechiri, M. Naoui, A. Khechekhouche, I. Marouani, T. Guesmi, B.M. Alshammari, A. AlGhadhban, A. Allal, An efficient bearing fault detection strategy based on a hybrid machine learning technique, Sci. Rep. 15 (1) (2025) 18739, http://dx.doi.org/10.1038/s41598-025-02439-4.
[12] M.R. Islam, Y.-H. Kim, J.-Y. Kim, J.-M. Kim, Detecting and learning unknown fault states by automatically finding the optimal number of clusters for online bearing fault diagnosis, Appl. Sci. 9 (11) (2019) http://dx.doi.org/10.3390/app9112326, URL https://www.mdpi.com/2076-3417/9/11/2326.
[13] L.C. Brito, G.A. Susto, J.N. Brito, M.A.V. Duarte, Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction, Informatics 8 (4) (2021) http://dx.doi.org/10.3390/informatics8040085, URL https://www.mdpi.com/2227-9709/8/4/85.
[14] D.E. Spina, L.F. de O. Campos, W.F. de Arruda, A. Melo, M.F. de S. Alves, G.L. Rabello, T.K. Anzai, J.C. Pinto, Comparison of autoencoder architectures for fault detection in industrial processes, Digit. Chem. Eng. 12 (2024) 100162, http://dx.doi.org/10.1016/j.dche.2024.100162, URL https://www.sciencedirect.com/science/article/pii/S2772508124000243.
[15] O. Das, D. Bagci Das, D. Birant, Machine learning for fault analysis in rotating machinery: A comprehensive review, Heliyon 9 (6) (2023) e17584, http://dx.doi.org/10.1016/j.heliyon.2023.e17584.
[16] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, X. Chen, Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study, IEEE Trans. Instrum. Meas. 70 (2021) 3525828, http://dx.doi.org/10.1109/TIM.2021.3116309.
[17] Q. Xiao, M. Yang, J. Yan, et al., Feature decoupling integrated domain generalization network for bearing fault diagnosis under unknown operating conditions, Sci. Rep. 14 (2024) 30848, http://dx.doi.org/10.1038/s41598-024-81489-6.
[18] I. Kim, S. Wook Kim, J. Kim, H. Huh, I. Jeong, T. Choi, J. Kim, S. Lee, Single domain generalizable and physically interpretable bearing fault diagnosis for unseen working conditions, Expert Syst. Appl. 241 (2024) 122455, http://dx.doi.org/10.1016/j.eswa.2023.122455.
[19] Y. Chen, J. Shi, C. Shen, H. Yang, Z. Hua, W. Huang, Z. Zhu, Time-frequency aware feature disentanglement learning for intelligent bearing fault diagnosis under variable speed conditions, Expert Syst. Appl. 303 (2026) 130664, http://dx.doi.org/10.1016/j.eswa.2025.130664.
[20] X. Yan, D. She, Y. Xu, Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions, Expert Syst. Appl. 216 (2023) 119479, http://dx.doi.org/10.1016/j.eswa.2022.119479.
[21] T. Li, Z. Peng, H. Xu, Q. He, Parameterized domain mapping for order tracking of rotating machinery, IEEE Trans. Ind. Electron. 70 (7) (2023) 7406–7416, http://dx.doi.org/10.1109/TIE.2022.3201311.
[22] M.A. Sa'd, T. Jalonen, S. Kiranyaz, M. Gabbouj, Quadratic time-frequency analysis of vibration signals for diagnosing bearing faults, 2024, arXiv arXiv:2401.01172.
[23] Y. Li, H. Xia, 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, pp. 4925–4930.
[24] L. Zhang, Y. Lv, W. Huang, C. Yi, Bearing fault diagnosis under various operation conditions using synchrosqueezing transform and improved two-dimensional convolutional neural network, Meas. Sci. Technol. 33 (8) (2022) 085002, http://dx.doi.org/10.1088/1361-6501/ac69b1.
[25] D. Liu, L. Cui, W. Cheng, Flexible generalized demodulation for intelligent bearing fault diagnosis under nonstationary conditions, IEEE Trans. Ind. Inform. 19 (3) (2023) 2717–2728, http://dx.doi.org/10.1109/TII.2022.3192597.
[26] L. Zeng, J. Jian, X. Chang, S. Wang, A meta-learning method for few-shot bearing fault diagnosis under variable working conditions, Meas. Sci. Technol. 35 (5) (2024) 056205, http://dx.doi.org/10.1088/1361-6501/ad28e7.
[27] H. Fan, Z. Ren, X. Cao, X. Zhang, J. Huang, A GTI&ada-act LMCNN method for intelligent fault diagnosis of motor rotor-bearing unit under variable conditions, IEEE Trans. Instrum. Meas. 73 (2024) 3508314, http://dx.doi.org/10.1109/TIM.2024.3351265.
[28] D. Peng, Z. Liu, H. Wang, Y. Qin, L. Jia, A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains, IEEE Access 7 (2019) 10278–10293, http://dx.doi.org/10.1109/ACCESS.2018.2888842.

[29] S.M.W.u.H. Naqvi, A. Arif, A. Khan, F. Bangash, G.J. Sirewal, B. Huang, AI-driven resilient fault diagnosis of bearings in rotating machinery, Sensors 25 (22) (2025) http://dx.doi.org/10.3390/s25227092.
[30] Hepco Motion, GFX – guidance system for beckhoff XTS linear transport system, 2025, https://www.hepcomotion.com/product/driven-track-systems/gfx-guidance-system-for-beckhoff-xts-linear-transport-system, (Accessed: 2025-08-21).
[31] Beckhoff Automation, Extended transport systems (XTS), 2025, https://www.beckhoff.com/en-en/products/motion/xts-linear-product-transport/?msclkid=fd69f89dec35190c45d89f054681eef2, (Accessed: 2025-08-21).
[32] Rockwell Automation, Independent cart technology, 2025, https://www.rockwellautomation.com/en-us/products/hardware/independent-cart-technology.html, (Accessed: 2025-08-21).
[33] A. Jabbar, M. Cocconcelli, G. D'Elia, D. Borghi, L. Capelli, J. Cavalaglio Camargo Molano, M. Strozzi, R. Rubini, MOIRA-UNIMORE bearing data set for independent cart systems, Appl. Sci. 15 (7) (2025) http://dx.doi.org/10.3390/app15073691, URL https://www.mdpi.com/2076-3417/15/7/3691.
[34] J. Cavalaglio Camargo Molano, L. Capelli, R. Rubini, D. Borghi, M. Cocconcelli, A bearing fault model for Independent Cart Conveyor System and its validation, Appl. Acoust. 159 (2020) 107069, http://dx.doi.org/10.1016/j.apacoust.2019.107069.
[35] M. Cocconcelli, J. Cavalaglio Camargo Molano, R. Rubini, L. Capelli, D. Borghi, Bearing fault model for an independent cart conveyor, in: A. Fernandez Del Rincon, F. Viadero Rueda, F. Chaari, R. Zimroz, M. Haddar (Eds.), Advances in Condition Monitoring of Machinery in Non-Stationary Operations, Springer International Publishing, Cham, 2019, pp. 211–220.
[36] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[37] L. van der Maaten, Barnes-hut-SNE, 2013, arXiv:1301.3342.
[38] L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, 2018, arXiv preprint arXiv:1802.03426, URL https://arxiv.org/abs/1802.03426.
[39] L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, 2020, arXiv:1802.03426, URL https://arxiv.org/abs/1802.03426.
[40] J. Healy, L. McInnes, Uniform manifold approximation and projection, Nat. Rev. Methods Prim. 4 (2024) 82, http://dx.doi.org/10.1038/s43586-024-00363-x.
[41] Y. Meng, J. Zhou, F. Lei, D. Li, R. Liu, A novel class of non-Gaussian system performance assessment and controller parameter tuning methods, ISA Trans. 154 (2024) 199–212, http://dx.doi.org/10.1016/j.isatra.2024.08.031.
[42] V. Laparra, G. Camps-Valls, J. Malo, Iterative Gaussianization: From ICA to random rotations, IEEE Trans. Neural Netw. 22 (4) (2011) 537–549, http://dx.doi.org/10.1109/TNN.2011.2114351.
[43] J.E. Johnson, V. Laparra, M. Piles, G. Camps-Valls, Gaussianizing the Earth: Multidimensional information measures for earth data analysis, IEEE Geosci. Remote. Sens. Mag. 9 (4) (2021) 191–208, http://dx.doi.org/10.1109/MGRS.2021.3066260.
[44] V. Laparra, A. Hepburn, J.E. Johnson, J. Malo, Orthonormal convolutions for the rotation based iterative Gaussianization, in: 2022 IEEE International Conference on Image Processing, ICIP, 2022, pp. 4018–4022, http://dx.doi.org/10.1109/ICIP46576.2022.9897849.
[45] K. Yu, Z. Wu, J. Sun, Y. Zhang, Y. Xu, Z. Wei, S. Zheng, Accelerating hyperspectral anomaly detection with enhanced multivariate Gaussianization based on FPGA, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–12, http://dx.doi.org/10.1109/TGRS.2024.3476152.
[46] J.A. Padrón-Hidalgo, V. Laparra, G. Camps-Valls, Unsupervised anomaly and change detection with multivariate Gaussianization, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–10, http://dx.doi.org/10.1109/TGRS.2021.3116186.
[47] G.E.P. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2) (1964) 211–243, http://dx.doi.org/10.1111/j.2517-6161.1964.tb00553.x.
[48] I.-K. Yeo, R.A. Johnson, A new family of power transformations to improve normality or symmetry, Biometrika 87 (4) (2000) 954–959, http://dx.doi.org/10.1093/biomet/87.4.954.
[49] P. Wang, Z. Long, Z. Lv, Z. Wang, Fault detection for non-Gaussian processes using multiple canonical correlation analysis models and box-cox transformation, IEEE Access 7 (2019) 68707–68717, http://dx.doi.org/10.1109/ACCESS.2019.2914960.
[50] J. Tang, Y. You, Y. Zhao, C. Guo, Z. Li, B. Yang, Fault diagnosis of HVAC system sensors: A method based on box-cox transformation and multi-model fusion, Energy Rep. 13 (2025) 3489–3503, http://dx.doi.org/10.1016/j.egyr.2025.03.012.
[51] M. Yusoff, Y. Mahmud, P.A.R. Azmi, M.T.M. Sallehud-din, The improvement of SMOTE-ENN-XGBoost through Yeo Johnson strategy on Dissolved Gas Analysis dataset, Energy Rep. 13 (2025) 6281–6290, http://dx.doi.org/10.1016/j.egyr.2025.05.013.
[52] T. Chen, L. Guo, H. Gao, D. Wang, T. Feng, Y. Yu, Investigations on improved Box-Cox sparsity measures for machine condition monitoring, ISA Trans. 157 (2025) 466–480, http://dx.doi.org/10.1016/j.isatra.2024.12.010.
[53] B. Chen, W. Zhang, J. Xi Gu, D. Song, Y. Cheng, Z. Zhou, F. Gu, A.D. Ball, Product envelope spectrum optimization-gram: An enhanced envelope analysis for rolling bearing fault diagnosis, Mech. Syst. Signal Process. 193 (2023) 110270, http://dx.doi.org/10.1016/j.ymssp.2023.110270.
[54] C. López, D. Wang, Á. Naranjo, K.J. Moore, Box-cox-sparse-measures-based blind filtering: Understanding the difference between the maximum kurtosis deconvolution and the minimum entropy deconvolution, Mech. Syst. Signal Process. 165 (2022) 108376, http://dx.doi.org/10.1016/j.ymssp.2021.108376, URL https://www.sciencedirect.com/science/article/pii/S0888327021007305.
[55] D. Peng, W. Teng, C. Gao, B. Tong, Y. Liu, Cyclic band Box-Cox sparse measures based blind filtering and its application to bearing fault diagnosis, Measurement 218 (2023) 113054, http://dx.doi.org/10.1016/j.measurement.2023.113054, URL https://www.sciencedirect.com/science/article/pii/S0263224123006188.
[56] J. Li, Y. Wang, J. Xiang, Z. Chen, A simulation guided BCSM blind deconvolution for fault diagnosis in wind turbine bearing, Meas. Sci. Technol. 36 (3) (2025) 036143, http://dx.doi.org/10.1088/1361-6501/adb6c6.
[57] Z. Zhao, Y.-G. Zhao, P.-P. Li, Efficient approach for dynamic reliability analysis based on uniform design method and Box-Cox transformation, Mech. Syst. Signal Process. 172 (2022) 108967, http://dx.doi.org/10.1016/j.ymssp.2022.108967, URL https://www.sciencedirect.com/science/article/pii/S0888327022001479.
[58] M.-C. Chen, C.-C. Hsu, B. Malhotra, M.K. Tiwari, An efficient ICA-DW-SVDD fault detection and diagnosis method for non-Gaussian processes, Int. J. Prod. Res. 54 (17) (2016) 5208–5218, http://dx.doi.org/10.1080/00207543.2016.1161250.
[59] A. Giantomassi, F. Ferracuti, S. Iarlori, G. Ippoliti, S. Longhi, Electric motor fault detection and diagnosis by kernel density estimation and Kullback–Leibler divergence based on stator current measurements, IEEE Trans. Ind. Electron. 62 (3) (2015) 1770–1780, http://dx.doi.org/10.1109/TIE.2014.2370936.
[60] C.I. Lang, F.-K. Sun, B. Lawler, J. Dillon, A.A. Dujaili, J. Ruth, P. Cardillo, P. Alfred, A. Bowers, A. Mckiernan, D.S. Boning, One class process anomaly detection using kernel density estimation methods, IEEE Trans. Semicond. Manuf. 35 (3) (2022) 457–469, http://dx.doi.org/10.1109/TSM.2022.3181468.
[61] Q. Qian, H. Pu, T. Tu, Y. Qin, Variance discrepancy representation: A vibration characteristic-guided distribution alignment metric for fault transfer diagnosis, Mech. Syst. Signal Process. 217 (2024) 111544, http://dx.doi.org/10.1016/j.ymssp.2024.111544.
[62] J. Cai, Y. Xiao, Time-frequency analysis method of bearing fault diagnosis based on the generalized S transformation, J. Vibroeng. 19 (6) (2017) 4221–4230, http://dx.doi.org/10.21595/jve.2017.18244.
[63] B. Sun, H. Li, C. Wang, Z. Ma, X. Guan, Optimized Weights Time-Frequency Analysis: A novel method for fault diagnosis in rotating machinery under time-varying speeds, Mech. Syst. Signal Process. 192 (2025) 112345, http://dx.doi.org/10.1016/j.ymssp.2025.112345.
[64] Q. He, Time–frequency manifold for nonlinear feature extraction in machinery fault diagnosis, Mech. Syst. Signal Process. 35 (1) (2013) 200–218, http://dx.doi.org/10.1016/j.ymssp.2012.08.018, URL https://www.sciencedirect.com/science/article/pii/S088832701200324X.

[65] X. Ding, Q. Li, L. Lin, Q. He, Y. Shao, Fast time-frequency manifold learning and its reconstruction for transient feature extraction in rotating machinery fault diagnosis, Measurement 141 (2019) 380–395, http://dx.doi.org/10.1016/j.measurement.2019.04.030, URL https://www.sciencedirect.com/science/article/pii/S0263224119303355.
[66] X. Ding, Q. He, Y. Shao, W. Huang, Transient feature extraction based on time–frequency manifold image synthesis for machinery fault diagnosis, IEEE Trans. Instrum. Meas. 68 (11) (2019) 4242–4252, http://dx.doi.org/10.1109/TIM.2018.2890316.
[67] L. Ciabattoni, F. Ferracuti, A. Freddi, A. Monteriù, Statistical spectral analysis for fault diagnosis of rotating machines, IEEE Trans. Ind. Electron. 65 (5) (2018) 4301–4310, http://dx.doi.org/10.1109/TIE.2017.2762623.
[68] A.C. Jahagirdar, K.K. Gupta, Cumulative distribution sharpness profiling based bearing fault diagnosis framework under variable speed conditions, IEEE Sensors J. 21 (13) (2021) 15124–15132, http://dx.doi.org/10.1109/JSEN.2021.3072368.
[69] G. Fu, Q. Wei, Y. Yang, Bearing fault diagnosis with parallel CNN and LSTM, Math. Biosci. Eng. 21 (2) (2024) 2385–2406, http://dx.doi.org/10.3934/mbe.2024105, URL https://www.aimspress.com/article/doi/10.3934/mbe.2024105.
[70] Z. Chen, H. Zhang, Z. Tao, Y. Wang, Intelligent fault diagnosis of rolling bearing based on time–frequency processing and DLKA-YOLO, Frankl. Open 10 (2025) 100228, http://dx.doi.org/10.1016/j.fraope.2025.100228, URL https://www.sciencedirect.com/science/article/pii/S2773186325000180.
[71] M.F. Siddique, F. Saleem, M. Umar, C.H. Kim, J.-M. Kim, A hybrid deep learning approach for bearing fault diagnosis using continuous wavelet transform and attention-enhanced spatiotemporal feature extraction, Sensors 25 (9) (2025) http://dx.doi.org/10.3390/s25092712, URL https://www.mdpi.com/1424-8220/25/9/2712.
[72] J. Chen, C. Lin, B. Yao, L. Yang, H. Ge, Intelligent fault diagnosis of rolling bearings with low-quality data: A feature significance and diversity learning method, Reliab. Eng. Syst. Saf. 237 (2023) 109343, http://dx.doi.org/10.1016/j.ress.2023.109343, URL https://www.sciencedirect.com/science/article/pii/S0951832023002570.
[73] L. Ye, X. Ma, C. Wen, Rotating machinery fault diagnosis method by combining time-frequency domain features and CNN knowledge transfer, Sensors 21 (24) (2021) http://dx.doi.org/10.3390/s21248168, URL https://www.mdpi.com/1424-8220/21/24/8168.
[74] G.-B. Jang, S.-B. Cho, Feature space transformation for fault diagnosis of rotating machinery under different working conditions, Sensors 21 (4) (2021) http://dx.doi.org/10.3390/s21041417, URL https://www.mdpi.com/1424-8220/21/4/1417.
[75] D.K.B. Kulevome, M. Qiu, F. Cao, E. Opoku-Mensah, Evaluation of time-frequency representations for deep learning-based rotating machinery fault diagnosis, Int. J. Eng. Technol. Innov. (2025) URL https://api.semanticscholar.org/CorpusID:280826018.
[76] A. Jabbar, M. Mazzonetto, L. Orazi, M. Cocconcelli, Ultrafast laser damaging of ball bearings for the condition monitoring of a fleet of linear motors, in: PHM Society European Conference, Vol. 8, 2024, p. 10, http://dx.doi.org/10.36001/phme.2024.v8i1.4136.
[77] A. Jabbar, C. Fonte, G. D'Elia, M. Cocconcelli, 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, ISMA, Leuven, Belgium, 2024, pp. 1698–1711.
[78] A.P. Daga, A. Fasana, S. Marchesiello, L. Garibaldi, The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data, Mech. Syst. Signal Process. 120 (2019) 252–273, http://dx.doi.org/10.1016/j.ymssp.2018.10.010.
[79] CWRU, Case western reserve university (CWRU) bearing data center, 2005, https://engineering.case.edu/bearingdatacenter, (Accessed on 20 January 2025).
[80] A. Jabbar, G. D'Elia, M. Cocconcelli, Distribution reshaping transformation for bearing fault diagnosis in independent cart systems, IEEE Access 13 (2025) 200403–200430, http://dx.doi.org/10.1109/ACCESS.2025.3636190.
[81] A. Jabbar, M. Cocconcelli, G. D'Elia, Bimodal distribution detection and transformation for Gaussian merging for bearing fault classification of independent cart systems under nonstationary conditions, Measurement 262 (2026) 119900, http://dx.doi.org/10.1016/j.measurement.2025.119900, URL https://www.sciencedirect.com/science/article/pii/S0263224125032592.
[82] A. Jabbar, G. D'Elia, M. Cocconcelli, Experimental setup for non-stationary condition monitoring of independent cart systems, in: U. Kumar, R. Karim, D. Galar, R. Kour (Eds.), International Congress and Workshop on Industrial AI and EMaintenance, Springer Nature Switzerland, 2024, pp. 517–530, http://dx.doi.org/10.1007/978-3-031-39619-9-38.
[83] A. Jabbar, M. Cocconcelli, G. d'Elia, M. Strozzi, R. Rubini, Results on experimental data analysis of independent cart systems in non-stationary conditions, in: Surveillance, Vibrations, Shock and Noise, Toulouse, France, 2023, URL https://hal.science/hal-04165905, hal-04165905.
[84] J.C.C. Molano, S. Rossi, M. Cocconcelli, R. Rubini, Dynamic model of an independent carts system, in: G. Boschetti, A. Gasparetto (Eds.), Advances in Italian Mechanism Science, Springer International Publishing, Cham, 2017, pp. 379–387.
[85] J.C.C. Molano, L. Scurria, C. Source, M. Cocconcelli, T. Tamarozzi, Virtual training of machine learning algorithm using a multibody model for bearing diagnostics on independent cart system, in: Proceedings of the 2020 International Conference on Noise and Vibration Engineering (ISMA 2020) and 2020 International Conference on Uncertainty in Structural Dynamics (USD 2020), KU Leuven, Leuven, Belgium, 2020, pp. 2013–2024, Paper presented at ISMA 2020 and USD 2020.
[86] J. Cavalaglio Camargo Molano, L. Capelli, R. Rubini, D. Borghi, M. Cocconcelli, A bearing fault model for Independent Cart Conveyor System and its validation, Appl. Acoust. 159 (2020) 107069, http://dx.doi.org/10.1016/j.apacoust.2019.107069, URL https://www.sciencedirect.com/science/article/pii/S0003682X18310806.
[87] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, Mech. Syst. Signal Process. 64–65 (2015) 100–131, http://dx.doi.org/10.1016/j.ymssp.2015.04.021, URL https://www.sciencedirect.com/science/article/pii/S0888327015002034.
[88] A. Jablonski, Condition Monitoring Algorithms in MATLAB®, Springer, Cham, 2021, http://dx.doi.org/10.1007/978-3-030-62749-2.
[89] K. Batzner, L. Heckler, R. König, EfficientAD: Accurate visual anomaly detection at millisecond-level latencies, in: 2024 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2024, pp. 127–137, http://dx.doi.org/10.1109/WACV57701.2024.00020.
[90] T. Defard, A. Setkov, A. Loesch, R. Audigier, PaDiM: A patch distribution modeling framework for anomaly detection and localization, in: A. Del Bimbo, R. Cucchiara, S. Sclaroff, G.M. Farinella, T. Mei, M. Bertini, H.J. Escalante, R. Vezzani (Eds.), Pattern Recognition. ICPR International Workshops and Challenges, Springer International Publishing, Cham, 2021, pp. 475–489.
[91] P. Liznerski, L. Ruff, R.A. Vandermeulen, B.J. Franks, M. Kloft, K.-R. Muller, Explainable deep one-class classification, 2020, arXiv arXiv:2007.01760.