UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dottorato di Ricerca in Ingegneria dell'Innovazione Industriale

Ciclo XXXVII

Hybrid Modelling Techniques for Condition Monitoring of Motion Control Applications

Candidato: Abdul Jabbar

Relatore (Tutor): Prof. Marco Cocconcelli

Coordinatore del Corso di Dottorato: Prof. Franco Zambonelli

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Hybrid Modelling Techniques for Condition Monitoring of Motion Control Applications

---

by

Abdul Jabbar

Submitted to the *Dipartimento di Scienze e Metodi dell'Ingegneria*, *Università degli Studi di Modena e Reggio Emilia* in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Industrial Innovation Engineering

December 2025

With faith, discipline and selfless devotion to duty,
there is nothing worthwhile that you cannot achieve
— M. A. Jinnah

To my parents, whose perseverance, sacrifices, and unwavering belief have laid the foundation for all my achievements;

To my wife, whose patience, strength, and steadfast support have guided me through every challenge;

and to my children, *Jibreel* and *Mikaeel*, whose love has been a profound source of joy and inspiration throughout this journey.

Acknowledgements

I wish to express my profound gratitude to **Prof. Marco Cocconcelli**, my supervisor, for his unwavering guidance, encouragement, and insightful discussions throughout this research. His expertise and vision have been instrumental in my technical and personal development. I am deeply appreciative of the independence he afforded me in my work, which allowed me to grow as an autonomous researcher and for the considerate and respectful manner in which he consistently supported me. His mentorship extended well beyond academic supervision, offering thoughtful advice and fostering an environment in which I always felt comfortable, understood, and never under unnecessary pressure. His kindness, patience, and holistic support have played crucial roles in shaping both my scientific outlook and confidence as a researcher. I am equally indebted to **Dr. Gianluca D'Elia** for his steadfast support, constructive feedback, and scientific mentorship during the course of this work.

I extend my sincere gratitude to **Prof. Riccardo Rubini** and **Prof. Matteo Strozzi** for their invaluable suggestions and for fostering an inspiring and highly supportive research environment. I am equally appreciative of my colleagues, **Cosimo Fonte** and **Pasquale Grosso**, whose friendliness, openness, and unwavering willingness to assist contributed to a uniquely welcoming atmosphere in the laboratory. Their collaborative spirit, approachability, and numerous engaging discussions significantly enriched my research journey and made daily work genuinely enjoyable. It is rare to encounter such a warm and collegial environment, and I feel privileged to have been part of it.

I am grateful to **Tetra Pak Italia** for providing the experimental test rig and facilitating its installation at the DISMI Laboratory at UNIMORE. I am particularly thankful to **Davide Borghi**, **Luca Capelli**, and **Jacopo Cavalaglio Camargo Molano** for their technical insights, practical support, and close collaboration throughout the experimental phase.

Special thanks are due to **Mr. Giovanni Paladini** at Beckhoff Italy for their invaluable assistance in resolving the challenges associated with the experimental setup. I am also deeply appreciative of **Prof. Leonardo Orazi** for granting access to his laboratory facilities and equipment, which were essential for conducting the laser-guided fault injection experiments, and **Manuel Mazzonetto** for his collaboration and support during the fault injection campaign.

Finally, I gratefully acknowledge the **European Commission** for supporting this research through the *Marie Skłodowska-Curie Innovative Training Network (ITN) MOIRA Project (Grant Agreement No. 955681)*, which made this work possible.

Reggio Emilia, December 2025 A. J.

Abstract

Unlike conventional rotary systems, where bearings remain fixed around a single axis and experience only rotational motion, bearings in independent cart systems exhibit coupled translational–rotational motion and highly non-stationary dynamics. This simultaneous translational-rotational motion of the bearings along the guide rail results in time-varying vibration signatures, complicating fault diagnosis using conventional methods. Furthermore, each cart in an independent cart system is supported by multiple rolling bearings, and industrial machines involve hundreds of carts operating simultaneously, resulting in a large number of interacting bearings, the condition monitoring of which is both challenging and highly desirable. This thesis addresses these challenges through a series of methodological and experimental contributions. First, a comprehensive and publicly accessible vibration dataset was developed from an extensive experimental campaign on a linear-motor-driven Independent Cart System. The dataset includes multiple fault types and severities, single- and multi-cart configurations, diverse speed profiles, and synchronized system variables, such as position, velocity, motor current, and following error. This is the first dataset specifically designed for translational–rotational bearing dynamics and establishes a foundation for benchmarking modern diagnostic algorithms.

Based on this dataset, this study initially introduces two distribution-reshaping transformations. The first method, *Bimodal Distribution Detection and Transformation*, presents a training-free approach that mitigates bimodality, reduces class overlap, and compacts feature distributions derived from multichannel vibration signals. When combined with statistical feature ranking and dimensionality reduction, it significantly enhances the separability between healthy and faulty conditions and improves the performance of anomaly detection methods, such as support vector-based detectors and isolation-based models. The second method, the *Adaptive Distribution Transformation*, extends this framework by addressing skewed, multimodal, and high-variance feature distributions that emerge under complex operating conditions. This transformation consistently enhances cluster compactness and separability in principal component and nonlinear low-dimensional embeddings, surpassing established transformations. It demonstrated strong robustness to broadband sensor noise, narrowband interference, and structurally shaped distortions, and generalized effectively across both single- and multi-cart experiments. The thesis further develops the *Iterative Weighted Compact Clustering Transformation*, a lightweight and fully unsupervised method that restructures dispersed or overlapping feature distributions into compact and linearly separable forms. It employs a non-expansive and convergent update rule with a linear computational cost and achieves superior separability without requiring labels or model training.

Additionally, this study introduces *Theil-Aware Intelligent Windowing*, a framework that autonomously selects the optimal spectrogram window size using a Theil-index formulation,

enabling consistent and high-resolution representations across variable speeds and transient regimes. When combined with the newly proposed *Distribution Reshaping via Bimodal Transformation*, the pipeline converts irregular, curved, and density-imbalanced manifolds into cluster-friendly geometries, significantly improving unsupervised separability. Validation across multiple datasets, including the developed Independent Cart System dataset, the Case Western Reserve University dataset, and the Politecnico di Torino dataset, demonstrated substantial gains in cluster compactness, separability, and fault detection accuracy. Finally, this study introduces entropy-based exhaustive and accelerated non-exhaustive search strategies for automatic spectrogram window selection. These methods enable the reliable identification of optimal window sizes while significantly reducing the computational burden of time–frequency analysis. The proposed strategies are particularly valuable for large-scale industrial deployments involving multi-machine systems and long-term multichannel data archives, where thousands of channels and years' worth of historical data must be processed consistently and efficiently. By preserving diagnostic fidelity while improving scalability, this approach makes adaptive window selection feasible for both real-time monitoring and embedded implementations. Overall, this thesis provides the first unified and unsupervised diagnostic framework tailored for translational–rotational industrial transport systems. By integrating a high-quality open-access dataset, adaptive time–frequency analysis, and a suite of innovative distribution-reshaping transformations, this study advances machine condition monitoring under nonstationary operating conditions and establishes a scalable, interpretable, and computationally efficient foundation for next-generation industrial diagnostics.

Keywords: Asset Management, Ball Bearings, Clustering Algorithms, Condition Monitoring, Fault Detection, Fault Diagnosis, Independent cart systems, Non-stationary, Predictive Maintenance, Rotating Machines, Statistical Distributions, Unsupervised Learning

Astratto

A differenza dei sistemi rotativi convenzionali, in cui i cuscinetti rimangono fissi attorno a un unico asse e subiscono solo un movimento rotatorio, i cuscinetti nei sistemi a carrelli indipendenti presentano un movimento traslatorio-rotatorio accoppiato e dinamiche altamente non stazionarie. Questo movimento traslatorio-rotatorio simultaneo dei cuscinetti lungo la propria guida provoca vibrazioni variabili nel tempo, complicando la diagnosi dei guasti con i metodi convenzionali. Inoltre, ogni carrello in un sistema indipendente è supportato da più cuscinetti volventi e le applicazioni industriali d questi sistemi coinvolgono centinaia di carrelli che operano simultaneamente, con un conseguente elevato numero di cuscinetti, il cui monitoraggio è sia altamente auspicabile ma impegnativo. Questa tesi affronta tali sfide attraverso una serie di contributi metodologici e sperimentali. In primo luogo, è stato sviluppato un set di dati completo e accessibile al pubblico, frutto di una vasta campagna sperimentale su un sistema a carrelli indipendenti azionato da motori lineari. Il set di dati include diversi tipi e livelli di gravità di guasti, configurazioni a carrello singolo e multiplo, diversi profili di velocità e variabili di sistema sincronizzate, quali posizione, velocità, corrente del motore ed errore di inseguimento. Si tratta del primo set di dati specificamente progettato per la dinamica dei cuscinetti traslazionali-rotazionali e costituisce una base di riferimento per il benchmarking dei moderni algoritmi diagnostici.

Sulla base di questo set di dati, lo studio introduce inizialmente due trasformazioni di distribuzione. Il primo metodo, Bimodal Distribution Detection and Transformation, presenta un approccio senza addestramento che mitiga la bimodalità, riduce la sovrapposizione delle classi e compatta le distribuzioni delle caratteristiche derivate dai segnali di vibrazione multicanale. Se combinato con la classificazione statistica delle caratteristiche e la riduzione della dimensionalità, migliora significativamente la separabilità tra condizioni sane e difettose e migliora le prestazioni dei metodi di rilevamento delle anomalie, come i Support Vector-based Detectors e gli isolation-based models. Il secondo metodo, la Trasformazione Adattiva della Distribuzione, estende questa idea affrontando le distribuzioni di caratteristiche asimmetriche, multimodali e ad alta varianza che emergono in condizioni operative complesse. Questa trasformazione migliora costantemente la compattezza e la separabilità dei cluster nelle componenti principali e negli incorporamenti non lineari a bassa dimensionalità, superando le trasformazioni consolidate. Ha dimostrato una forte robustezza nei confronti del rumore dei sensori a banda larga, delle interferenze a banda stretta e delle distorsioni strutturali, e si è generalizzata efficacemente sia negli esperimenti con un solo carrello che in quelli con più carrelli. La tesi sviluppa ulteriormente la trasformazione iterativa ponderata del clustering compatto, un metodo leggero e completamente non supervisionato che ristruttura le distribuzioni di caratteristiche disperse o sovrapposte in forme compatte e linearmente separabili. Impiega una regola di aggiornamento non espansiva

e convergente con un costo computazionale lineare, raggiungendo una separabilità superiore senza richiedere etichette o addestramento del modello.

Inoltre, questo studio introduce il Theil-Aware Intelligent Windowing, un framework che seleziona autonomamente la dimensione ottimale della finestra dello spettrogramma utilizzando una formulazione dell'indice di Theil, consentendo rappresentazioni coerenti e ad alta risoluzione su velocità variabili e regimi transitori. Se combinata con la nuova proposta di rimodellamento della distribuzione tramite trasformazione bimodale, la pipeline converte manifold irregolari, curvi e con densità sbilanciata in geometrie adatte ai cluster, migliorando significativamente la separabilità non supervisionata. La convalida su più set di dati, tra cui il set di dati Independent Cart System sviluppato in Unimore, il set di dati della Case Western Reserve University e il set di dati del Politecnico di Torino, ha dimostrato miglioramenti sostanziali in termini di compattezza dei cluster, separabilità e accuratezza del rilevamento dei guasti. Infine, questo studio introduce strategie di ricerca esaustiva basate sull'entropia e strategie di ricerca non esaustiva accelerata per la selezione automatica delle finestre dello spettrogramma. Questi metodi consentono l'identificazione affidabile delle dimensioni ottimali delle finestre, riducendo significativamente il carico computazionale dell'analisi tempo-frequenza. Le strategie proposte sono particolarmente preziose per le implementazioni industriali su larga scala che coinvolgono sistemi multi-macchina e archivi di dati multicanale a lungo termine, dove migliaia di canali e anni di dati storici devono essere elaborati in modo coerente ed efficiente. Preservando la fedeltà diagnostica e migliorando la scalabilità, questo approccio rende la selezione adattiva delle finestre fattibile sia per il monitoraggio in tempo reale che per le implementazioni integrate. Nel complesso, questa tesi fornisce il primo quadro diagnostico unificato e non supervisionato specifico per i sistemi di trasporto industriale traslazionale-rotazionale. Integrando un set di dati open access di alta qualità, un'analisi adattiva tempo-frequenza e una serie di trasformazioni innovative di rimodellamento della distribuzione, questo studio fa progredire il monitoraggio delle macchine in condizioni operative non stazionarie e stabilisce una base scalabile, interpretabile ed efficiente dal punto di vista computazionale per la diagnostica industriale di prossima generazione.

Parole chiave: Gestione delle risorse, cuscinetti a sfera, algoritmi di clustering, monitoraggio delle condizioni, rilevamento dei guasti, diagnosi dei guasti, sistemi a carrelli indipendenti, non stazionario, manutenzione predittiva, macchine rotanti, distribuzioni statistiche, apprendimento non supervisionato

Contents

List of Figures

List of Tables

List of Acronyms

CBM	Condition-Based Monitoring
CM	Condition Monitoring
CWRU	Case Western Reserve University
IMS	Intelligent Maintenance Systems
XJTU-SY	Xi'an Jiaotong University, and Sumyoung Technology
SEU	Southeast University
PHM	Prognostics and Health Management
ICS	Independent Cart System
MOIRA	Monitoring Large-scale Complex Systems
UNIMORE	University of Modena and Reggio Emilia
PSD	Power Spectral Density
IR	Inner Race
OR	Outer Race
PCA	Principal Component Analysis
t-SNE	t-Distributed Stochastic Neighbor Embedding
KDE	Kernel Density Estimate
RBIG	Rotation-Based Iterative Gaussianization
STFT	Short Time Fourier Transform
WT	Wavelet Transform
BDD-T	Bimodal Distribution Detection and Transformation
ADT	Adaptive Distribution Transformation
IWCCT	Iterative Weighted Compact Clustering Transformation
DRBT	Distribution Reshaping via Bimodal Transformation
TAIW	Theil-Aware Intelligent Windowing
OC-SVM	One-Class Support Vector Machines
IF	Isolation Forest
iTRAK	Intelligent Track Systems
BPFI	Ball Pass Frequency Inner
BPFO	Ball Pass Frequency Outer
BSF	Ball Spin Frequency
FTF	Fundamental Train Frequency
RCM	Reliability centered Maintenance
RUL	Remaining Useful Life
CWT	Continuous Wavelet Transform
EMD	Empirical Mode Decomposition
SST	Synchrosqueezing Transform
HHT	Hilbert–Huang transform
EEMD	Ensemble Empirical Mode Decomposition
CEEMDAN	Complete Ensemble EMD With Adaptive Noise

VMD	Variational Mode Decomposition
ML	Machine Learning
K-NN	K-Nearest Neighbors
SVM	Support Vector Machine
CNN	convolutional Neural Network
DBN	Deep Belief Network
RNN	Recurrent Neural Network
GAN	Generative Adversarial Network
DAN	Deep Adaptation Networks
TCA	Transfer Component Analysis
ADA	Adversarial Domain Adaptation
DTL	Deep Transfer Learning
UDTL	Unsupervised Deep Transfer Learning
GOA-IDF	Grasshopper Optimization Algorithm–SVM
DRSN	Deep Residual Shrinkage Networks
GTI	Gray Texture Image
MSST	Multisynchrosqueezing Transform
FGD	Flexible Generalized Demodulation
FPGA	Field-Programmable Gate Array
ORBIG	Optimized Rotation-Based Iterative Gaussianization
CDF	Cumulative Distribution Function
HVAC	Heating, Ventilation, and Air Conditioning
ICA	Independent Component Analysis
SVDD	Support Vector Data Description
AC	Alternative Current
LSTM	Long Short-Term Memory
PWM	Pulse Width Modulation
GFX	Guidance System for Beckhoff Linear Transport Systems
XTS	Extended Transport System
IEPE	Integrated Electronics Piezo-Electric
D.O.F	Depth of Field
IR-Top	Inner Race Fault of the Top Bearing
OR-Top	Outer Race Fault of the Top Bearing
IR-Bottom	Inner Race Fault of the Bottom Bearing
OR-Bottom	Outer Race Fault of the Bottom Bearing
ANOVA	One-Way Analysis of Variance
KW	Kruskal–Wallis
LS	Laplacian Score
Var	Variance
Mono	Monotonicity
SNR	Signal-to-Noise Ratio
THD	Total Harmonic Distortion
SINAD	Signal-to-Noise And Distortion Ratio
VDR	Variance Discrepancy Representation
AWGN	Additive White Gaussian Noise
NB	Narrow Band
SN	Structural Noise
FRF	Frequency Response Function
PFRF	Pseudo Frequency Response Function
FFT	Fast Fourier Transform
IFFT	Inverse Fast Fourier Transform

FFT	Fast Fourier Transform
IFFT	Inverse Fast Fourier Transform
FIR025	Fault-Inner Race with Width = 0.25mm
FIR05	Fault-Inner Race with Width = 0.5mm
FIR10	Fault-Inner Race with Width = 1.0mm
FIR15	Fault-Inner Race with Width = 1.5mm
FOR025	Fault-Outer Race with Width = 0.25mm
FOR05	Fault-Outer Race with Width = 0.5mm
FOR20	Fault-Outer Race with Width = 2.0mm
HPT	Hyperbolic Power Transform
CDF–TS	Cumulative Distribution Function Transform and Shift

Chapter 1

Introduction

1.1 Context and Motivation

The evolution of industrial maintenance strategies has transitioned from run-to-break and time-based preventive maintenance to a more sophisticated predictive maintenance paradigm, known as condition-based maintenance (CBM). Under the run-to-break strategy, machinery is operated until failure, often resulting in unplanned downtime, collateral component damage, and significant production losses. Preventive maintenance introduces scheduled servicing at fixed intervals shorter than the anticipated time to failure, thereby reducing catastrophic breakdowns but leading to unnecessary component replacement and excessive maintenance efforts.

CBM, also referred to as Condition Monitoring (CM), represents the most advanced and efficient strategy in this evolutionary process. CBM anticipates potential failures by continuously evaluating a machine's health through measurable physical quantities, such as vibration, current, temperature, acoustic emission, lubricant analysis, and strain. By identifying the early signs of mechanical degradation and scheduling interventions at the optimal time, CBM facilitates predictive maintenance rather than reactive repair, thereby minimizing downtime and extending the equipment lifespan. It effectively bridges operational reliability and economic efficiency by enabling maintenance decisions based on actual conditions rather than arbitrary time intervals (Randall, 2011).

In recent decades, CM has become an essential element of industrial reliability engineering and a crucial facilitator of data-driven maintenance and digital-twin technology. In traditional rotary machinery, such as motors, gearboxes, turbines, and pumps, CM has made significant progress, supported by standardized sensing technologies, diagnostic algorithms, and widely utilized public datasets, including those from Case Western Reserve University (CWRU) (CWRU, 2005), Intelligent Maintenance Systems (IMS) (Lee et al., 2007), Politecnico di Torino (Daga et al., 2019), Xi'an Jiaotong University, and Sumyoung Technology (XJTU-SY) (B. Wang et al., 2020). Additional benchmarks, such as the Southeast University (SEU) gearbox fault dataset (Shao et al., 2018) and Prognostics and Health Management (PHM) Challenge dataset (Nectoux et al., 2012), have further facilitated comparative research on fault diagnosis methods. Collectively, these resources have established vibration analysis, envelope detection, and frequency-domain feature extraction as standard practices for fault detection in stationary or quasi-stationary machine components.

As industrial automation progresses towards greater flexibility and precision, manufacturing systems are increasingly adopting modular, magnetically actuated motion-control platforms, such as independent cart systems (ICS) (GFX-Guidance, 2025; iTRAK, 2025; XTS, 2025). Enabled by linear motor propulsion, ICS represents a new generation of intelligent transport solutions that replace conventional conveyors, rollers, and chain-driven mechanisms. Their operation relies on electromagnetic forces that propel carts with friction-free, high-speed, and individually controlled motions along customizable guide rails. This contactless actuation eliminates mechanical transmission components, thereby reducing wear, friction losses, and maintenance requirements while achieving remarkable precision and configurability.

Despite these advantages, ICS presents unique diagnostic challenges (Cavalaglio Camargo Molano et al., 2020; Jabbar, D'Elia, & Cocconcelli, 2024; Jabbar, Fonte, et al., 2024; Jabbar et al., 2023). Each cart is mechanically supported and guided by bearings in continuous contact with the rail, forming the primary path for vibration transmission and fault propagation. Unlike conventional rotary machinery, the bearings in ICS experience coupled translational and rotational motions, generating time-varying load distributions and non-stationary vibration signatures. Furthermore, carts operate under highly variable speeds, ranging from a few millimeters to several meters per second, and may accelerate, decelerate, or reverse direction rapidly. These hybrid motion patterns invalidate the traditional constant-speed assumptions, making fault signatures transient, speed-dependent, and position-specific. In large-scale installations, where dozens or even hundreds of carts move simultaneously, each equipped with multiple bearings, monitoring and isolating faults across interacting components become increasingly complex.

While condition monitoring for rotary systems has been extensively studied, ICS remains underexplored owing to the proprietary nature of industrial platforms and the lack of open, reproducible datasets. To address this gap, the *MOIRA–UNIMORE* dataset (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a) offers the first publicly accessible collection of vibration and system-variable data from a real ICS, establishing a critical experimental foundation for developing diagnostic frameworks that can adapt to non-stationary environments and ensure reliable fault detection under realistic, variable-speed regimes. Building on this foundation, the present study aims to advance data-driven adaptive diagnostic methods capable of addressing the complex hybrid translational–rotational behaviors characteristic of ICS. To contextualize these objectives, the principal industrial and diagnostic challenges motivating this research are summarized below:

1.1.1 Key Industrial and Diagnostic Challenges

Despite their growing industrial relevance, ICS introduces a unique set of technical and operational challenges that motivate the present research.

- **Economic impact and uptime sensitivity:** ICS installations are capital-intensive, and unplanned downtime directly affects throughput and profitability. Therefore, continuous condition monitoring is essential to avoid costly stoppages and sustain operational efficiency.

- **Data scarcity and confidentiality:** Although ICS are increasingly replacing conventional conveyors, their diagnostic data are rarely made public because of proprietary restrictions and non-disclosure agreements. This scarcity of open data has hindered repro-

ducible research, a challenge addressed in this study through newly acquired experimental data.

- **Non-stationary dynamics:** Cart velocities can vary from a few millimeters to several meters per second (up to 4m/s) with frequent accelerations and reversals. These rapid changes induce highly nonstationary vibration responses that violate the assumptions of most conventional diagnostic algorithms and destabilize deep models optimized for fixed-speed conditions.

- **Translational–rotational interplay:** Unlike conventional rotary machinery, where bearings are mounted around a fixed shaft, ICS bearings undergo both translation and rotation, altering the structure and evolution of fault signatures compared with those of standard test rigs.

- **System scale and modularity:** Each cart within an ICS typically incorporates multiple bearings, often numbering three or more, and industrial installations frequently consist of hundreds of independently moving carts. Consequently, a substantial number of bearings simultaneously experience coupled translational and rotational motions along the shared guide rails. This mechanical concurrency results in intricate vibration interactions, wherein signals from multiple sources overlap in both the time and frequency domains, thereby complicating the precise identification and localization of individual fault sources.

- **Mismatch with conventional datasets:** Existing bearing datasets, such as CWRU, Paderborn, IMS, and XJTU–SY, originate from stationary rotary machinery and fail to capture the coupled translational–rotational dynamics and speed-dependent behaviors characteristic of ICS. The recently published *MOIRA–UNIMORE* (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a) dataset provides the first realistic representation of these operating conditions, offering a sound basis for the development and validation of new diagnostic algorithms.

- **Need for label-free, real-time solutions:** Given the label scarcity and computational constraints in industrial settings, there is a growing need for lightweight, deterministic approaches that learn from healthy data and autonomously flag deviations in real time.

1.2 Problem Statement

The challenges outlined above highlight a fundamental discrepancy between the analytical assumptions inherent in traditional fault diagnosis methodologies and the dynamic operational characteristics of ICS. Although the broader literature on machine condition monitoring has made substantial advancements in addressing nonstationary behavior by utilizing time–frequency transforms, order-tracking techniques, empirical and variational decompositions, and deep learning architectures, these frameworks were primarily developed for rotating machinery operating around a fixed rotational axis. In such systems, the motion is typically periodic, the kinematics are consistent across cycles, and the spectral evolution remains coherent with the rotational speed under steady or gradually varying conditions. These assumptions facilitate the effective isolation of fault-related modulations in the angular or order domains. However, such periodicity and spectral coherence are entirely disrupted in ICS environments, where motion is not governed by a single rotational reference but by coupled translational–rotational

dynamics that continuously vary over space and time.

In ICS, carts can move along the guide rail following arbitrary speed profiles defined by the geometry of the linear motor stators and control logic, provided that collisions between carts are avoided. These profiles can be individually programmed to execute a wide range of trajectories suited for production or transport tasks. One of the most representative motion profiles is the oscillatory mode, in which a cart travels back and forth between two endpoints. This operation involves alternating phases of acceleration, deceleration, and direction reversal, resulting in complex motion patterns that combine translational and rotational dynamics of the cell. An example of this oscillatory trajectory is illustrated in Figure 1.5. Because the sensors are mounted on a fixed rail, the measured vibration signals vary not only with the cart's instantaneous speed but also with its spatial position and the mechanical characteristics of the rail section being traversed. Therefore, the signal amplitude and frequency content depend on the localized contact conditions, rail discontinuities, and dynamic coupling among multiple simultaneously moving carts. In industrial implementations, such variations are further influenced by changing payloads or process-induced forces, whereas in a laboratory setting, the excitation is limited to the cart's own weight. Consequently, the recorded signals are non-periodic, position-dependent, and non-stationary, lacking a consistent spectral reference.

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

Figure 1.1 presents representative raw vibration signals acquired from a healthy bearing and an inner-race fault of 0.25mm during a bidirectional traversal along a curved motor module. Impulsive events are intermittently observed in *both* healthy and faulty conditions, primarily arising from repeated acceleration-deceleration phases and transient interactions between the cart and guide rail. However, the overall waveforms were dominated by large-amplitude modulation effects induced by the motion dynamics. Consequently, fault-related components are neither temporally isolated nor consistently repeatable across cycles, rendering a reliable diagnosis based solely on direct time-domain inspection infeasible. To further assess the frequency-domain separability, Figure 1.2 depicts the power spectral density (PSD) estimates for multiple bearing conditions. Although broad energy redistribution is observable, the dominant spectral content is primarily governed by structural resonances and motion-induced excitation. Fault-related variations appear as subtle deviations rather than distinct spectral lines, and a strong overlap persists between healthy and faulty conditions, even when averaged spectra are considered.

Envelope analysis is commonly adopted to enhance impulsive fault signatures; however, as shown in Figure 1.3, under nonstationary ICS operation, envelope signals and their corresponding spectra remain heavily influenced by speed fluctuations and transient impacts. Although certain amplitude differences emerge, they are intermittent, broadband, and highly sensitive to the selected demodulation band, limiting their reliability as stand-alone diagnostic indicators. Finally, Figure 1.4 illustrates the time-frequency representations of the same signals.

The spectrograms revealed pronounced nonstationarity, with energy concentrations evolving continuously over time and frequency as the cart accelerated, decelerated, and reversed direction. Although localized transient patterns can be observed, these structures do not yield clear, manually separable fault signatures and exhibit significant overlap between healthy and faulty cases.

Collectively, Figures 1.1–1.4 demonstrate that under realistic ICS operating conditions, neither time-domain inspection nor conventional frequency or time–frequency domain representations produce stable or manually separable fault signatures. Consequently, a reliable diagnosis cannot be achieved solely at the signal level. This necessitates the extraction of higher-level representations through windowed feature computation, which facilitates the aggregation of information across time, frequency, and motion phases. However, as discussed subsequently, the challenges posed by ICS nonstationarity persist even after feature extraction, manifesting as overlapping and poorly structured feature distributions.

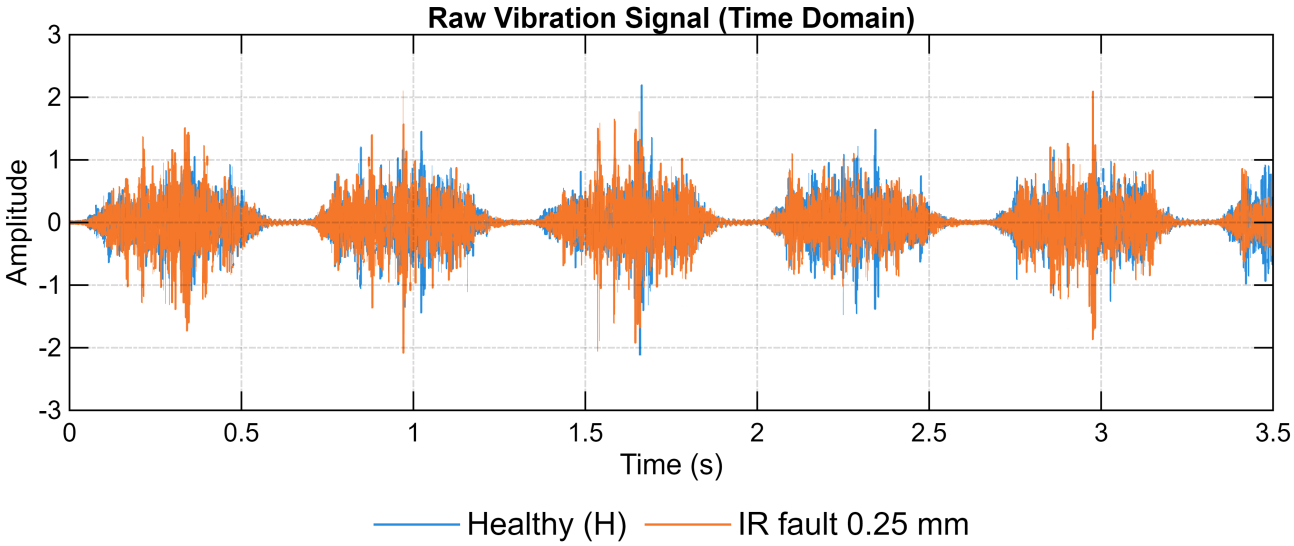


Figure 1.1: Vibration signals were obtained from an accelerometer affixed to the guide rail during a single bidirectional traversal of a curved module under both healthy (H) and inner-race fault (IR 0.25 mm) conditions. These signals demonstrated pronounced amplitude modulation owing to the acceleration, steady motion, and deceleration phases of the wearer. The impulses associated with faults are obscured by variability induced by motion and are not consistently discernible in the time domain.

In light of these signal-level constraints, the analysis was expanded beyond direct waveform interpretation to encompass a structured feature domain representation. The focus has shifted from exclusive reliance on spectral localization to capturing the statistical organization and variability patterns induced by motion-dependent dynamics. Consequently, windowed feature extraction was conducted over position-synchronized segments (see Figure 1.5), facilitating the aggregation of information across the motion phases. Statistical descriptors and Hjorth parameters were chosen for their capacity to characterize the amplitude modulation, signal power distribution, and dynamic complexity without presupposing strict periodicity or spectral coherence. When visualized in reduced-dimensional spaces using principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), these features exhibited skewed, heavy-tailed, and occasionally bimodal marginal distributions (Figure 1.6, Figure 1.7). Notably, such irregularities do not necessarily stem from intrinsic vibration randomness but arise from the

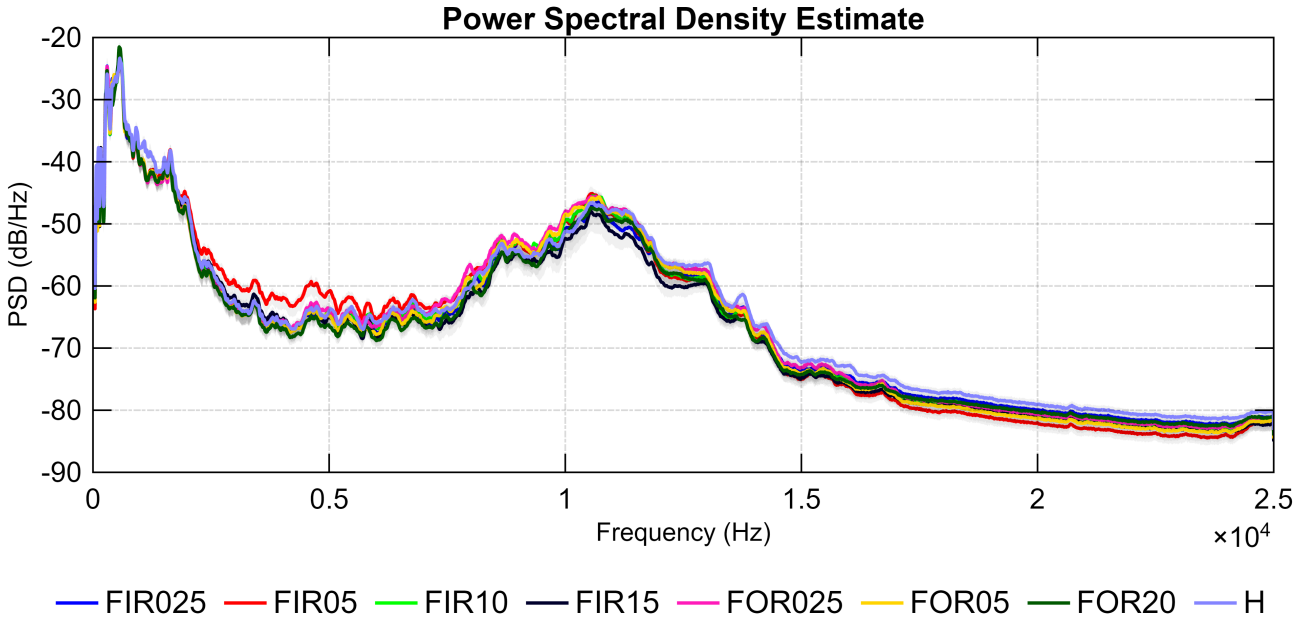


Figure 1.2: Power spectral density (PSD) estimates for both healthy and faulty bearing conditions under nonstationary ICS operation reveal energy redistribution across frequency bands. However, the dominant spectral components are predominantly influenced by the structural and motion-related effects of the target. The significant overlap among conditions constrains the ability to distinguish faults within the frequency domain.

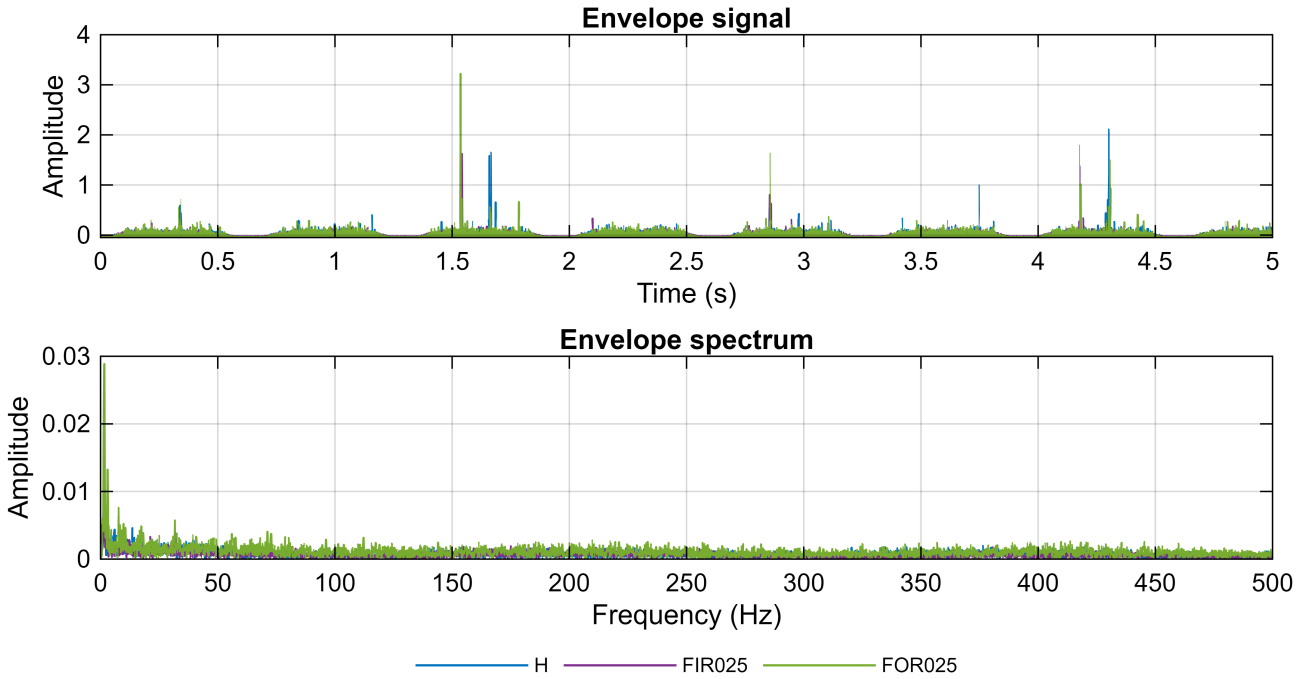


Figure 1.3: The envelope signals and their corresponding envelope spectra, derived through band-limited Hilbert demodulation, are analyzed for various bearing conditions. Although envelope analysis is effective in highlighting impulsive behavior, the resultant representations are significantly influenced by nonstationary speed variations and transient impacts, leading to broadband and overlapping spectral characteristics.

interaction between feature computation and spatially varying motion dynamics. Furthermore, the distributional behavior was context-dependent and varied across trajectories and load con-

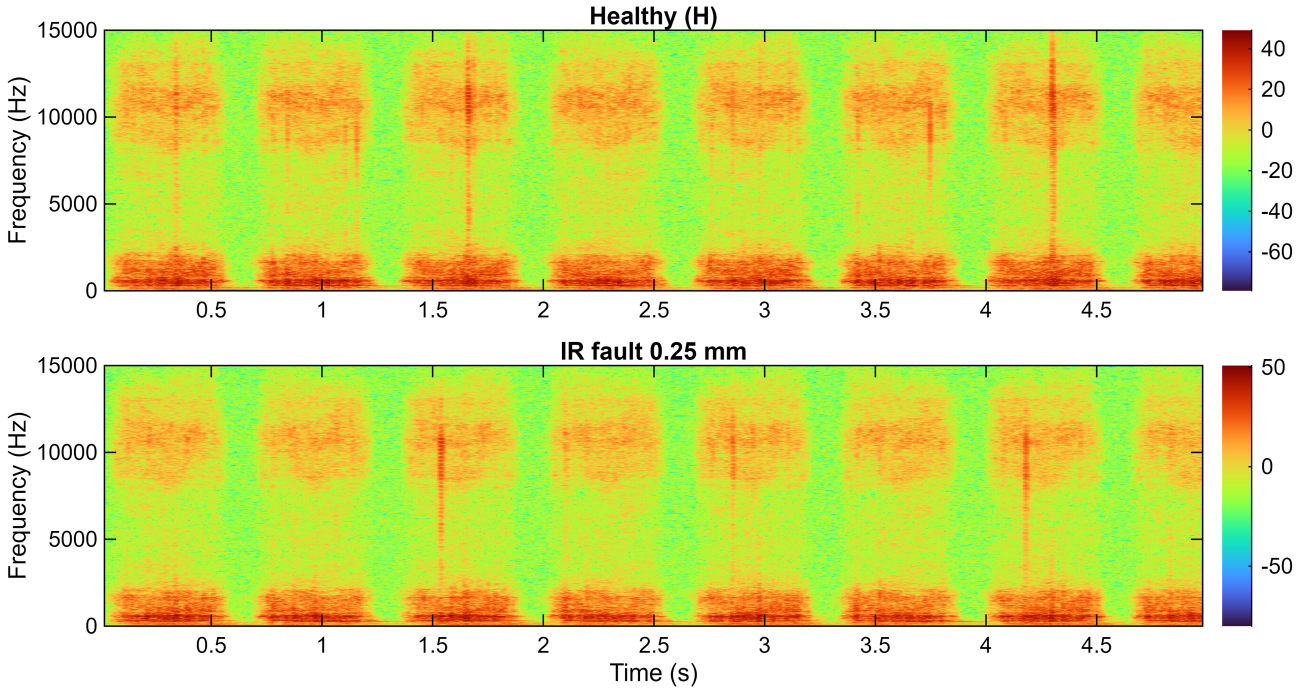


Figure 1.4: The time-frequency spectrograms of vibration signals from both healthy and inner-race faulty conditions during bidirectional cart motion along a curved module reveal significant nonstationarity and dynamic energy concentrations associated with phases of acceleration, deceleration, and direction reversal. Patterns indicative of faults are subtle and exhibit substantial overlap with those of healthy behavior.

ditions. Thus, while feature extraction alleviates some signal-level ambiguity, non-stationarity persists in the statistical structure of the feature space. The challenge is to transition from transient detection to distributional regularization and geometric stabilization.

Collectively, these challenges delineate the central research problem of this thesis as follows:

> *How can features from complex, nonstationary, and label-scarce motion-control systems be transformed into statistically regular and discriminative representations that enable reliable, scalable, and interpretable fault detection?*

Addressing this problem necessitates the development of adaptive, label-free transformation methodologies capable of reshaping irregular feature distributions into compact, Gaussian-like structures while preserving the intrinsic geometric relationships among samples. Such transformations should enhance separability and stability across operating conditions without relying on labeled data or extensive parameter-tuning. The subsequent chapters introduce and evaluate transformation-based diagnostic frameworks that are specifically designed to meet the requirements of the ICS platform.

1.3 Research Gaps

The preceding discussion underscores several unresolved challenges that motivate this study. Although significant progress has been achieved in vibration analysis and fault diagnosis for rotary machinery, ICS remain in the early stages of diagnostic development. A critical review

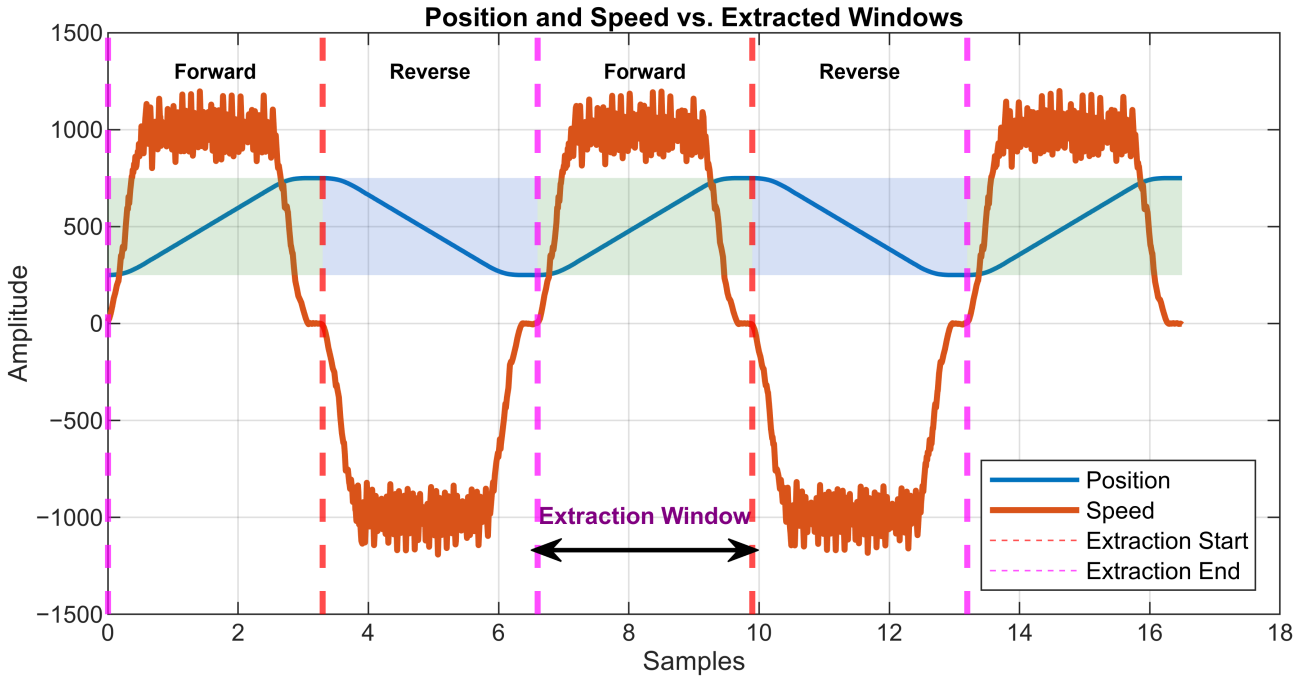


Figure 1.5: Window segmentation based on cart position. The motion phases are divided into forward and reverse intervals to facilitate feature extraction and ensure temporal alignment across motion cycles.

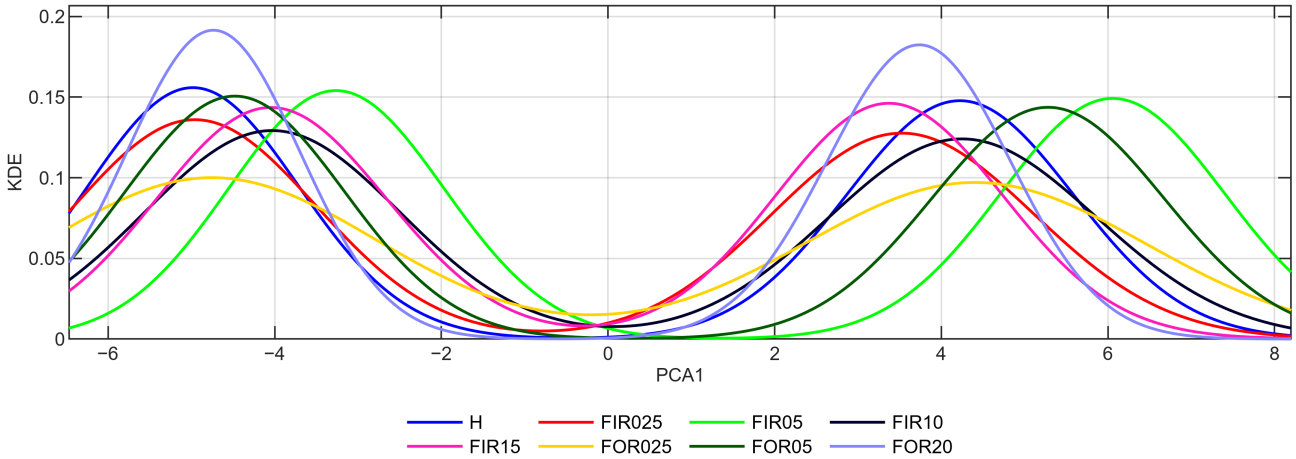


Figure 1.6: Marginal Kernel Density Estimate (KDE) along the first principal component for various fault classes. Several features exhibit overlapping and multimodal distributions, illustrating the challenge of achieving discriminative separation.

of the existing methods and literature reveals the following research gaps:

1. **Lack of open access datasets for ICS.** Most publicly available vibration datasets, such as CWRU, IMS, Paderborn, and XJTU–SY, are derived from stationary rotary rigs operating at constant or quasi-constant speeds. In contrast, no open and reproducible dataset has previously existed for linear-motor-driven ICSs, where the bearings experience coupled translational–rotational motion under variable-speed regimes. This absence hinders benchmark development, algorithm comparison, and reproducible research in the motion control domain.

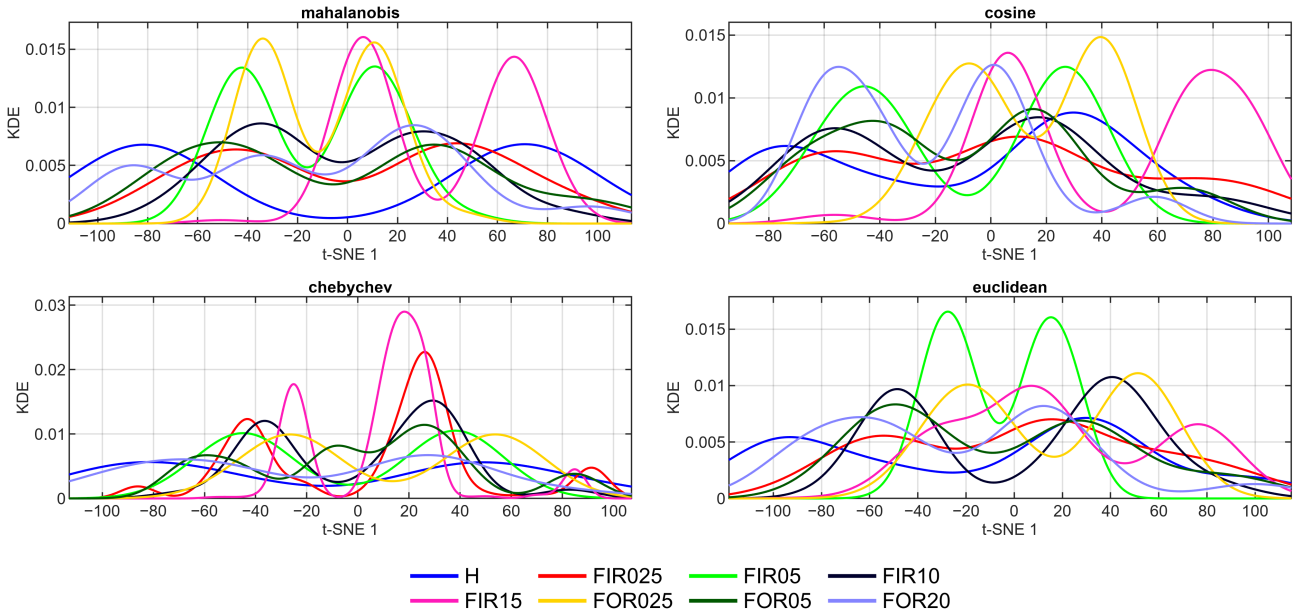


Figure 1.7: Marginal KDEs along the first t-SNE dimension using different distance metrics. Class overlap and scattered distributions are evident even within nonlinear embeddings.

2. **Poor feature separability due to distribution irregularities.** Conventional handcrafted and statistical features derived from ICS vibration signals frequently exhibit skewed, multimodal, or heavy-tailed distributions. These attributes diminish the interclass separability and undermine unsupervised clustering and anomaly detection. Although numerous studies have proposed Gaussianization or normalization techniques, such as zero-mean scaling, power transformation (Box & Cox, 1964; L.-X. Qin & Self, 2006; Yeo & Johnson, 2000), quantile transformation, hyperbolic power transformation (Tsai et al., 2017), Sinh-arcsinh transformation (Jones & Pewsey, 2009), and rotation-based iterative Gaussianization (RBIG) (Laparra et al., 2011), to regularize non-Gaussian features, these methods typically homogenize all features around a common mean, resulting in overlapping distinct classes. Consequently, while they enhance statistical regularity, they do not improve diagnostic discriminability, thereby limiting their effectiveness in label-free fault detection.

3. **Lack of adaptive, label-efficient transformations.** Current feature learning and signal processing methodologies predominantly depend on supervised training or presuppose quasi-stationary input conditions. In scenarios characterized by non-stationary ICS conditions, fluctuations in speed and motion state result in distortions of the statistical geometry of the features, thereby compromising class separability. Currently, there is no established framework capable of executing adaptive label-free transformations that maintain class-consistent geometry while standardizing feature densities for unsupervised fault detection.

4. **Absence of intelligent time–frequency window selection under variable-speed operation.** Classical time–frequency analyses (STFT, wavelet transforms) employ fixed or heuristically chosen window lengths that fail to adapt to the local signal complexity or instantaneous frequency content. This limitation leads to information loss in high-speed transients and excessive redundancy during the low-speed intervals. An automated criterion for optimal window selection that is responsive to signal entropy or non-stationarity

remains largely unexplored in ICS diagnostics.

5. **Limited frameworks for scalable, label-free fault detection in multi-cart systems.** In industrial installations, hundreds of independently moving carts, each containing multiple bearings, operate simultaneously. Current diagnostic strategies do not scale effectively to such modular architectures, where labeling is infeasible and dynamic coupling between carts obscures fault localization.

These research gaps collectively define the central objective of this thesis: to develop and validate adaptive, data-driven diagnostic methodologies that address non-stationary behavior, distribution irregularities, and scalability constraints in Independent Cart Systems, thereby advancing the frontier of condition monitoring in motion-control applications.

1.4 Research Objectives

Building on the research gaps identified above, this thesis aims to advance the field of condition monitoring for ICS through the development of data-driven, adaptive, and reproducible diagnostic frameworks. The primary objective of this study was to enhance fault separability and diagnostic reliability under variable-speed, nonstationary operating conditions. The specific research objectives are as follows.

1. **Develop a public benchmark dataset for ICS condition monitoring.** The first open-access dataset (*MOIRA–UNIMORE*) was established, comprising synchronized vibration and system-variable measurements from a linear-motor-based ICS, thereby facilitating reproducible research and benchmarking.

2. **Design transformation-based frameworks for enhanced fault separability.** Adaptive and label-efficient transformations have been developed, including bimodal distribution detection and transformation, Iterative Weighted Compact Clustering Transformation (IWCCT), and Distribution Reshaping via Bimodal Transformation (DRBT), to compact class distributions, reduce overlap, and preserve the discriminative structure in non-Gaussian feature spaces.

3. **Formulate intelligent time–frequency strategies for Intelligent windowing.** The Theil-Aware Intelligent Windowing (TAIW) and entropy-based frameworks have been devised to facilitate the automatic selection of optimal analysis windows. This selection process is guided by the spectrogram Theil index and Shannon entropy. Consequently, these frameworks enhance the time-frequency resolution in variable-speed regimes.

4. **Integrate unsupervised clustering and anomaly detection for robust diagnostics.** The proposed transformations and adaptive windowing are combined with unsupervised diagnostic algorithms—including One-Class Support Vector Machines (OC-SVM) and Isolation Forests (IF)—to enable scalable, real-time, label-free fault detection and monitoring in multi-cart ICS environments.

Collectively, these objectives define a comprehensive research agenda aimed at bridging the methodological gap between stationary diagnostic techniques and the emerging demands of motion control systems operating under hybrid translational–rotational dynamics.

1.5 Scope and Contributions

The research undertaken in this doctoral study significantly enhances the condition monitoring of linear-motor-based motion control platforms, with a specific focus on Independent Cart Systems (ICS). The contributions encompass a range of areas, including experimental data acquisition, algorithmic advancements in adaptive feature transformation, nonstationary signal analysis, and unsupervised fault diagnosis. Each of these contributions was disseminated through peer-reviewed publications and conference proceedings, as outlined below.

Published and Under Review Works

The following works were developed during the Ph.D. candidacy and collectively form the scientific basis of this thesis:

- **A. Jabbar, M. Cocconcelli, and G. D'Elia**, "*Adaptive Distribution Transformation for Robust Bearing Fault Diagnosis under Noise and Nonstationarity*," ***ISA Transactions*** 2026.
  Contribution: Proposed a robust, adaptive transformation framework for reshaping skewed and noisy feature distributions while preserving discriminative structure.

- **A. Jabbar, M. Cocconcelli, and G. D'Elia**, "*Theil-Aware Intelligent Windowing and Distribution Reshaping for Enhanced Spectrogram Feature Learning*," ***Mechanical Systems and Signal Processing (MSSP)***, 2026.
  Contribution: Combined theil-index guided window optimization (TAIW) with Distribution Reshaping via Bimodal Transformation (DRBT) for adaptive time–frequency analysis in variable-speed regimes.

- **A. Jabbar, M. Cocconcelli, G. D'Elia, D. Borghi, L. Capelli, J. Cavalaglio Camargo Molano, M. Strozzi, and R. Rubini**, "*MOIRA–UNIMORE Bearing Dataset for Independent Cart Systems*," ***Applied Sciences***, vol. 15, no. 7, p. 3691, 2025.
  Contribution: Established the first open-access vibration and system-variable dataset for ICS, providing a reproducible benchmark for motion-control diagnostics.

- **A. Jabbar, G. D'Elia, and M. Cocconcelli**, "*Distribution Reshaping Transformation for Bearing Fault Diagnosis in Independent Cart Systems*," in ***IEEE Access***, doi: 10.1109/ACCESS.2025.3636190, 2025.
  Contribution: Introduced the IWCCT algorithm for compacting class distributions and enhancing feature separability through iterative weighted compaction.

- **A. Jabbar, M. Cocconcelli, and G. D'Elia**, "*Bimodal Distribution Detection and Transformation for Gaussian Merging for Bearing Fault Classification of Independent Cart Systems under Nonstationary conditions*," Accepted, ***Measurement***, 2025.
  Contribution: Developed a bimodal Gaussian-merging approach to mitigate feature overlap and improve intra-class compactness in ICS fault dataset.

- **A. Jabbar, M. Mazzonetto, L. Orazi, and M. Cocconcelli**, "*Ultrafast Laser Damaging of Ball Bearings for the Condition Monitoring of a Fleet of Linear Motors*," in ***Proceedings of the PHM Society European Conference***, vol. 8, Prague, Czech Republic, 3–July 5, 2024, p. 10.
  Contribution: Demonstrated a controlled and repeatable method of fault introduction in

bearings using ultrafast laser machining to simulate progressive degradation in a fleet of linear motors.

- **A. Jabbar, C. Fonte, G. D'Elia, and M. Cocconcelli** *Ball-bearings fault detection for an independent cart system: experimental campaign and preliminary results*, (2024), ***Proceedings of the 31st International Conference on Noise and Vibration Engineering, ISMA 2024 and 10th International Conference on Uncertainty in Structural Dynamics, USD 2024.*** Lueven, Belgium, September 9-11, 2024, pp. 1698-1711.
  Contribution: Presented autoencoder based anomaly detection framework.

- **A. Jabbar, G. D'Elia, and M. Cocconcelli**, "*Experimental Setup for Non-Stationary Condition Monitoring of Independent Cart Systems*," in U. Kumar, R. Karim, D. Galar and R. Kour (Eds.), ***Industrial AI and eMaintenance 2023*** Lecture Notes in Mechanical Engineering. Springer, Cham, 2023. doi:10.1007/978-3-031-39619-9_38.
  Contribution: Presented the design and instrumentation of the MOIRA testbed, introducing a modular architecture for multi-cart data acquisition and synchronized system-variable measurement.

- **A. Jabbar, M. Cocconcelli, G. D'Elia, M. Strozzi, and R. Rubini**, "*Results on Experimental Data Analysis of Independent Cart Systems in Non-Stationary Conditions*," in ***Surveillance, Vibrations, Shock and Noise***, ISAE-SUPAERO, Toulouse, France, 2023. (HAL Id: hal-04165905)
  Contribution: Reported the first vibration-based analysis of ICS under variable-speed operation, establishing baseline spectral characteristics and non-stationary behaviors.

Summary of Research Scope

Table 1.1 summarizes the research stages and associated scientific outputs, linking each paper to its methodological and experimental contributions.

1.6 Thesis Structure

This thesis is systematically organized to present a coherent progression from foundational concepts to methodological advancements and experimental validation. The structure includes introductory material, background theory, the development of a comprehensive vibration dataset for Independent Cart Systems, and a series of research contributions that collectively address the broader challenge of unsupervised bearing fault diagnosis under nonstationary operating conditions. These contributions correspond to the published and submitted works summarized in Table 1.1. Overall, the organization ensures a clear and logical flow from problem formulation to proposed solutions, culminating in integrated insights and future research directions.

- **Chapter 1 – Introduction:** This chapter presents the research context, motivation, and problem definition, outlines the key industrial challenges in the condition monitoring of ICS, identifies existing research gaps, and states the objectives and contributions of the thesis.

- **Chapter 2 – Background and Literature Review:** This section provides the theoretical and technical foundations of the study, covering maintenance philosophies, condition

Table 1.1: Summary of research scope and publication-based contributions.

Paper/Venue	Core Contribution
ISA Transactions (2026)	Adaptive Distribution Transformation (ADT) robust to noise.
MSSP (2026)	Theil-Aware Intelligent Windowing (TAIW) + feature transformation for adaptive time–frequency learning.
Applied Sciences (2025)	Open-access ICS dataset (MOIRA–UNIMORE) for reproducible research.
IEEE Access (2025)	Distribution Reshaping Transformation for Bearing Fault Diagnosis in Independent Cart Systems.
Measurement (2025)	Gaussian merging framework for bimodal feature overlap reduction.
PHM Europe (2024)	Laser-based fault introduction for controlled bearing damage.
ISMA (2024)	Autoencoder-based fault detection for controlled bearing damage.
Industrial AI and eMaintenance (2023)	Experimental setup and data acquisition architecture for multi-cart ICS monitoring.
Surveillance, Vibrations, Shock and Noise (2023)	First vibration analysis of ICS under non-stationary conditions.

monitoring principles, signal processing methods for non-stationary analysis, and recent developments in machine learning and distribution-based transformations relevant to fault diagnosis.

- **Chapter 3 – Experimental Setup and Data Acquisition:** This chapter provides a detailed account of the experimental testbed, sensor specifications, experimental configurations, and data acquisition processes for the ICS. This discussion aligns with the studies published in *Applied Sciences* (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a), *Industrial AI and eMaintenance* (Jabbar, D'Elia, & Cocconcelli, 2024), *Surveillance, Vibrations, Shock and Noise* (Jabbar et al., 2023), *International Conference on Noise and Vibration Engineering* (Jabbar, Fonte, et al., 2024), and *Annual Conference of the Prognostics and Health Management* (Jabbar, Fonte, et al., 2024).

- **Chapter 4 – Bimodal Detection and Distribution Transformation:** This chapter introduces the initial distribution-reshaping frameworks designed to improve feature separability. It presents the *Bimodal Distribution Detection and Transformation (BDD–T)* method (Jabbar et al., 2026), which identifies bimodal feature densities and transforms them to create compact, Gaussian-like representations, as elaborated in a study accepted for publication in *Measurement*. Furthermore, the chapter expands on this concept with the *Adaptive Distribution Transformation (ADT)*, a generalized and noise-resistant enhancement of the BDD–T framework. Collectively, these approaches address the issues of skewness and bimodality in feature spaces, establishing a methodological basis for unsupervised diagnostics under variable-speed and non-stationary conditions.

- **Chapter 5 – Iterative Weighted Compact Clustering Transformation:** This chapter presents the *Iterative Weighted Compact Clustering Transformation (IWCCT)*–a

feature distribution reshaping framework designed to enhance the compactness and separability of features for unsupervised fault diagnosis under non-stationary conditions. This method introduces iterative density weighting and adaptive contraction to transform skewed and dispersed feature distributions into more cohesive and discriminative representations without requiring class labels. The study corresponding to this work was published in (Jabbar, D'Elia, & Cocconcelli, 2025).

- **Chapter 6 – Intelligent Time-Frequency Feature Embedding and Reshaping:** This chapter expands upon distribution-based adaptation by integrating the *Distribution Reshaping via Bimodal Transformation (DRBT)* strategy with the *Theil-Aware Intelligent Windowing (TAIW)* framework. It presents a Theil-index-driven window-selection mechanism for adaptive time–frequency feature extraction in highly non-stationary regimes.

- **Chapter 7 – Entropy-Based Intelligent Spectrogram Window Search Strategies:** This chapter presents entropy-driven frameworks for the automatic selection of spectrogram window. An exhaustive entropy-based search is established as a statistically grounded baseline, and two accelerated strategies, namely ternary and bidirectional searches, are proposed to mitigate computational demands. These accelerated strategies consistently achieve near-optimal window lengths while significantly reducing the number of spectrogram evaluations, thereby facilitating scalable time-frequency analysis across datasets.

- **Chapter 8 – Conclusions and Future Work** This final chapter summarizes the main findings and contributions of this study, discusses its limitations, and outlines potential directions for future extensions in data-driven and physics-informed diagnostics for motion control systems.

Collectively, this thesis follows a coherent trajectory from experimental foundation to methodological innovation and validation, culminating in a unified framework for unsupervised, adaptive, and reproducible condition monitoring of ICS.

Chapter 2

Background and Literature Review

2.1 Independent Cart Systems and Motion Control

ICS represent an advanced category of motion-control architectures that incorporate linear motor propulsion, magnetic carts, and modular mechatronic design to facilitate high-throughput and high-precision transport in industrial automation (GFX-Guidance, 2025; iTRAK, 2025; XTS, 2025). In contrast to traditional conveyor or rotary drive mechanisms that depend on belts, chains, or gears, ICS employs electromagnetic actuation. This involves a stationary stator composed of distributed coil segments that generate controlled magnetic fields that interact with permanent magnets embedded in each *cart* or *mover*. This interaction generates Lorentz forces that propel the cart along a customizable path, allowing fully independent acceleration, deceleration, synchronization, and reversal of motion. Typical movers can achieve speeds of up to 4 m/s with micrometer-level positioning accuracy, making ICS particularly suitable for packaging, pharmaceutical, and semiconductor applications.

2.1.1 Mechanical Layout and Bearing Configuration

Although the propulsion is contactless, the cart is mechanically guided by rolling-element bearings that ensure alignment and load transfer along the guide rail. The cart can be uniquely designed with a bearing configuration tailored to the application, including the type and material of the bearings. Additionally, independent carts may have guide rails attached directly to the linear motors or in parallel (Cocconcelli et al., 2019; Jabbar, D'Elia, & Cocconcelli, 2024; Jabbar, Fonte, et al., 2024; Jabbar, Mazzonetto, et al., 2024; Jabbar et al., 2023; Molano et al., 2017). These bearings constitute the sole mechanical interface between moving and stationary structures, serving as both the principal vibration transmission path and primary source of mechanical wear. Consequently, the rail–bearing interaction governs the dynamic response of the entire system and provides a foundation for vibration-based fault diagnostics (Cavalaglio Camargo Molano et al., 2020; Cocconcelli et al., 2019; Molano et al., 2017).

2.1.2 Coupled Translational–Rotational Dynamics

A distinctive feature of ICS motion is the coupling between the translational movement of the cart and the rotational motion of its bearings. During translation, the outer races of the bearings rotate owing to the rolling contact with the rail, whereas the inner races remain fixed relative

to the mover frame. This dual motion introduces complex time-varying vibration signatures that depend on the instantaneous position, velocity, and acceleration. In multi-cart operations, simultaneous movements along the shared track produce overlapping vibration components that can interfere with and obscure individual fault signatures (Cavalaglio Camargo Molano et al., 2020; Cocconcelli et al., 2019).

2.1.3 Dynamic Modeling and Fault Characterization

A few studies have investigated the dynamic and vibrational behavior of ICS to elucidate these coupled effects. A comprehensive dynamic model of an iTRAK-based ICS encompassing electromagnetic propulsion, cart kinematics, and track geometry was presented in (Molano et al., 2017). Complementary analytical formulations for bearing-fault vibration generation specific to ICS, incorporating variable-speed profiles, rail curvature, load modulation, and slip phenomena, have been developed and experimentally validated (Cavalaglio Camargo Molano et al., 2020; Cocconcelli et al., 2019; Molano et al., 2017). Furthermore, a multibody simulation environment was proposed in (Molano et al., 2020), which facilitated the generation of synthetic vibration data for training and validating machine-learning algorithms under controlled fault conditions. These modeling approaches provide valuable insights into the relationship between rail geometry, bearing dynamics, and resulting vibration signatures.

2.1.4 Contrast with Conventional Rotary Machinery

The operating principles of ICS fundamentally differ from those of traditional rotary systems and test rigs, such as those used in the CWRU (CWRU, 2005), IMS (Lee et al., 2007), Politecnico di Torino (Daga et al., 2019), SEU (Shao et al., 2018), and XJTU–SY (B. Wang et al., 2020) datasets. In these rotary setups, the bearing is fixed around a rotating shaft operating at a constant speed, resulting in stationary or quasi-stationary vibration signals characterized by periodic fault frequencies (BPFI, BPFO, BSF, and FTF). These conditions favor frequency-domain methods, such as envelope analysis, order tracking, and cepstral techniques. In contrast, ICS bearings operate under translational–rotational coupling, variable velocity, and frequent acceleration phases. Consequently, their vibration responses are non-periodic, position-dependent, and highly non-stationary, which invalidates many of the assumptions underlying conventional diagnostic techniques.

In summary, although the ICS architecture offers unparalleled flexibility and performance in motion control, it simultaneously introduces diagnostic complexities arising from its multi-cart, nonsynchronous, and position-dependent vibration dynamics. These characteristics necessitate advanced signal processing and data-driven frameworks capable of extracting consistent fault-related features under continuously varying operating conditions.

2.2 Condition Monitoring and Predictive Maintenance

Condition monitoring (CM) is a fundamental component of predictive maintenance strategies in contemporary industrial systems. It involves the continuous evaluation of machine health through quantifiable physical parameters, such as vibration, current, temperature, acoustic emission, or lubricant condition, to identify degradation prior to functional failure. Within the predictive maintenance framework, CM acts as a pivotal enabler for the shift from time-based

interventions to data-driven maintenance, where decisions are informed by real-time diagnostic and prognostic insights (Randall, 2011; Rao, 2019).

2.2.1 Theoretical Basis and Maintenance Hierarchy

The theoretical underpinning of condition monitoring is rooted in reliability-centered maintenance (RCM), which aims to optimize system availability by balancing preventive and corrective actions to minimize downtime. Traditional time-based maintenance presupposes that degradation follows predictable lifetimes, whereas condition-based maintenance uses sensory data to detect deviations from nominal operations. By analyzing the temporal and spectral variations in the monitored signals, CM facilitates the early detection of incipient faults, estimation of the remaining useful life (RUL), and scheduling of maintenance actions at optimal times. This closed loop between sensing, analysis, and decision-making forms the foundation of modern predictive maintenance frameworks (Mobley, 2002).

2.2.2 Modalities of Condition Monitoring

Various sensing modalities have been developed to observe the mechanical, thermal, electrical, and tribological phenomena associated with component degradation.

- **Vibration analysis:** Employed for bearings, gears, shafts, and structural faults through time-, frequency-, or time–frequency-domain techniques.
- **Acoustic emission:** Detects transient stress waves from crack initiation and surface defects.
- **Motor current signature analysis:**Monitors electrical and mechanical anomalies via current spectra.
- **Lubricant and oil-debris analysis:** Identifies wear particles and chemical degradation in lubrication systems.
- **Thermal and infrared monitoring:** Captures frictional heating and thermal imbalance.

These modalities complement each other in multi-sensor maintenance architectures; however, vibration-based monitoring remains the most prevalent and information-rich approach for rotating and reciprocating machinery.

2.2.3 Dominance of Vibration Analysis

Vibration analysis is widely recognized as the cornerstone of condition monitoring because of its high sensitivity to dynamic changes in mechanical systems and its capability to detect both localized and distributed faults. This enables the direct assessment of the mechanical interactions among the rolling elements, races, and supporting structures, thereby allowing precise fault detection, trending, and prognosis. Compared with other modalities, such as thermal, acoustic, or lubricant monitoring, vibration analysis offers superior temporal resolution, well-established diagnostic models, and mature sensing technologies that facilitate early fault identification across a wide range of operating conditions (Chu et al., 2024; Randall, 2011; Romanssini et al., 2023; Saidin et al., 2022; Scheffer & Girdhar, 2015; Tiboni et al., 2022). Furthermore, publicly

available datasets including CWRU (CWRU, 2005), IMS (Lee et al., 2007), Paderborn (Lessmeier et al., 2016), and XJTU–SY (B. Wang et al., 2020) have standardized vibration-based fault benchmarking, accelerating progress in both signal-processing and data-driven machine learning approaches.

2.2.4 Transition to Complex Motion-Control Applications

Numerous methodologies for vibration analysis have been developed to address non-stationary conditions, including the short-time Fourier transform (STFT), continuous wavelet transform (CWT), empirical mode decomposition (EMD), and synchrosqueezing transform (SST). However, these techniques were primarily designed for systems with controlled speed variations or single-axis rotational motions. In contrast, linear-motor-driven platforms, such as ICS, present fundamentally different operating dynamics characterized by hybrid translational–rotational motion, frequent accelerations and reversals, and a strong coupling between mechanical and electromagnetic excitations. The resulting vibration signals are nonstationary and spatially dependent, with time–frequency structures that evolve as a function of position and velocity rather than rotational order. Although existing non-stationary signal processing frameworks can capture some of these effects, they often struggle to maintain diagnostic consistency when the excitation mechanisms change with the trajectory or when multiple carts interact along shared guide rails. This complexity necessitates the development of adaptive data-driven diagnostic frameworks that integrate statistical distribution transformation, intelligent time–frequency analysis, and unsupervised learning to ensure robust fault detection under actual operating conditions. The subsequent sections of this chapter, along with the methodological chapters that follow, address these challenges by introducing adaptive distribution-reshaping and Theil-aware time–frequency frameworks that are specifically tailored for ICS environments.

2.3 Intelligent Fault Diagnosis under Nonstationary Conditions

Overview and Motivation

The field of fault diagnosis has increasingly embraced intelligent frameworks that incorporate signal processing and machine learning to automate feature extraction and minimize the reliance on expert knowledge. This advancement is particularly crucial under non-stationary conditions characterized by unpredictable fluctuations in speed and load. Traditional vibration-based methods, such as Fourier analysis, envelope detection, and spectral analysis, often struggle to maintain diagnostic reliability under such variability owing to their inherent assumption of stationarity (Farhat et al., 2021; Tayyab et al., 2022). Consequently, there has been a necessary shift towards adaptive and data-driven techniques to ensure reliable fault characterization in real-world industrial systems.

Advanced Signal-Processing Techniques

Robust signal processing tools have been developed to address the limitations of classical stationary domain methods. Angle domain techniques, such as order tracking and rotational resampling, provide quasi-stationary representations for variable-speed signals (Farhat et al., 2021). Localized spectral approaches, including the STFT, wavelet transform (WT), and Hilbert–Huang transform (HHT), effectively capture transient and non-stationary dynamics

within the time–frequency domain (Jiang et al., 2025; Z. Liu et al., 2022; Wu et al., 2012). Beyond fixed-resolution methods, adaptive decomposition methods, such as the EMD, ensemble EMD (EEMD), complete ensemble EMD with adaptive noise (CEEMDAN), and variational mode decomposition (VMD), further isolate fault-relevant nonlinear modes (B. Peng et al., 2022). Recent advancements, including bandwidth-division VMD (BDVMD) and SST, offer enhanced spectral resolutions in noisy and transient regimes (W. Huang et al., 2019; J. Ma et al., 2025).

Integration with Machine Learning

These signal processing advancements are frequently integrated with machine learning (ML) techniques to automate detection and classification. Traditional classifiers, such as k-nearest neighbors (k-NN), support vector machines (SVM), and decision trees, rely on handcrafted statistical features, whereas deep learning models, particularly convolutional neural networks (CNNs) and autoencoders, learn hierarchical representations directly from raw signals or spectrograms (B. Peng et al., 2022; Saufi et al., 2019; Tayyab et al., 2022). Comprehensive reviews have outlined deep architectures, including CNNs, deep belief networks (DBNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), as well as transfer-learning frameworks such as deep adaptation networks (DAN), transfer component analysis (TCA), and adversarial domain adaptation (ADA) (Hakim et al., 2023; Lei et al., 2020), which have shown strong potential in rotating machinery diagnostics. Despite significant diagnostic improvements, challenges persist, including noise sensitivity, data imbalance, interpretability, and generalization across the operating regimes.

Unsupervised learning techniques, such as k-means, self-organizing maps, and Isolation Forests, extend their applicability to unlabeled datasets (Brito et al., 2021; Islam et al., 2019). These methods are often complemented by autoencoder-based representation learning for feature compression and anomaly detection (Das et al., 2023; Spina et al., 2024).

Deep, Transfer, and Meta-Learning Approaches

Deep transfer learning (DTL) has emerged as a fundamental approach for fault diagnosis in scenarios characterized by limited label availability and domain shift conditions (C. Li et al., 2020). Instance-, feature-, and parameter-based DTL models address challenges related to domain mismatch and data scarcity; however, they encounter issues such as computational complexity and sensitivity to data distributions. Comprehensive taxonomies of DTL and deep architectures underscore their potential, while highlighting concerns regarding interpretability and scalability (Hakim et al., 2023; Lei et al., 2020).

Unsupervised deep transfer learning (UDTL) advances these concepts by aligning source and target features in the absence of labeled target data (Zhao et al., 2021). Recent investigations have provided structured benchmarks for UDTL evaluation and underscored the risks associated with negative transfer. Complementary studies on cross-domain diagnosis and adversarial adaptation underscore the significance of domain-invariant feature extraction to ensure robustness under varying operational conditions (Zheng et al., 2019).

Hybrid models that integrate time–frequency analysis with deep neural architectures further enhance the robustness. Notable examples include wavelet-based denoising combined with machine learning classifiers (Fu et al., 2023), CNNs employing nuisance attribute projection

to mitigate speed-induced variability (H. Ma et al., 2019), and meta-heuristic optimization frameworks such as the Grasshopper Optimization Algorithm–SVM (GOA-IDF) hybrid, which improves accuracy and computational efficiency (Bristi et al., 2023).

Meta-learning and few-shot paradigms are gaining prominence in data-scarce environments. Noteworthy efforts encompass deep residual shrinkage networks (DRSN) with prototypical learning (Zeng et al., 2024), Gray Texture Image (GTI)-based lightweight CNNs tailored for IoT applications (Fan et al., 2024), and residual 1D-CNNs designed for high-speed train bearings (D. Peng et al., 2019). Complementary research has utilized acoustic emission and spectral imaging for low-data fault detection (Hasan et al., 2019), Mixup-augmented self-attention networks with contrastive adaptation (L. Li et al., 2025), and SST fused with multiscale CNNs for high-resolution classification (L. Zhang et al., 2022). Additional hybrid formulations include EEMD–CWT-based mutual attention architectures (Y. Li & Xia, 2024), FFT-based lightweight domain adaptation (DAFTL) (Tong et al., 2018), and low-latency CNNs for embedded acoustic applications.diagnostics (Pham et al., 2020).

Recent advancements in unsupervised frameworks have explored third-order tensor clustering (Wei et al., 2023), contrastive few-shot domain learning (T. Zhang et al., 2023), and interpretable classifiers that utilize decision trees, random forests, and XGBoost for edge analytics (Alhams et al., 2024). Additional contributions include hybrid RNN–CNN and VMD-based deep models for variable-speed operations (Z. Ma & Guo, 2024), the multisynchrosqueezing transform (MSST) for achieving high spectral concentration (Q. Liu & Wang, 2021), and flexible generalized demodulation (FGD) for frequency normalization under varying speeds (D. Liu et al., 2023). Further advancements, such as multi-scale convolutional domain adaptation with dynamic MMD loss (Q. Zhang et al., 2024), domain-adaptive residual networks (Yang et al., 2021), and meta-learning for rapid adaptation (C. Li et al., 2021), enhance resilience to domain shifts.

Domain Adaptation and Generalization Frameworks

The research landscape increasingly emphasizes domain generalization, wherein diagnostic models learn invariant representations without access to target-domain data (Sa'd et al., 2024). Multi-source and optimal-transport-based architectures (Shen et al., 2022) have achieved improved cross-domain performance, although they are parameter-sensitive. Other transfer learning extensions for rail and industrial systems employ adversarial networks and channel reconstruction units (Zou et al., 2024) or semi-supervised deep models utilizing complex envelope spectra (Sohaib & Kim, 2018). Hybrid LMD–SVD–ELM frameworks (Tian et al., 2015), ICEEMDAN–entropy–ESVM methods (R. Li et al., 2020), and adaptive PCA-based algorithms (Hamadache et al., 2015) demonstrate enhanced resilience to variable-speed operations, though their applicability to compound and real-world noise conditions remains limited. Fuzzy inference and probabilistic reasoning approaches, which combine the PWVD, RCI, ACO, and possibility theory (K. Li et al., 2013), have also shown potential but remain computationally complex.

Recent investigations have focused on multiscale fractional indicators (Y. Huang et al., 2022), multispectral preprocessing (Song et al., 2023), and ConceFT-based time–frequency feature extraction (Feng et al., 2017). These frameworks significantly enhance diagnostic precision but impose high computational costs and often rely on geometry-specific prior knowledge.

Remaining Challenges and Research Gaps

Despite significant advancements in signal processing, machine learning, and domain-adaptive frameworks, a critical methodological gap persists. Most contemporary techniques emphasize feature extraction or classifier enhancement, often overlooking the intrinsic statistical structure of feature distributions. In practical industrial settings, features derived from vibration data frequently exhibit skewed, multimodal, or heavy-tailed distributions that deviate from Gaussian assumptions, thereby impairing fault-class separability (An et al., 2022; Y. Chen et al., 2025). Although deep models can approximate these distributions through hierarchical nonlinear mappings, their reliability diminishes in noisy, nonstationary, or low-data conditions, where overfitting and poor generalization are prevalent (Lei et al., 2020; Saufi et al., 2019; Zhao et al., 2021). Therefore, explicit distribution transformation methodologies that can reshape skewed or bimodal features into more compact, Gaussian-like representations are urgently needed. Such transformations have the potential to stabilize feature variance, enhance clustering separability, and improve interpretability in unsupervised settings. Addressing this limitation is particularly pertinent for complex motion control systems, such as ICS, where inherent speed variability, coupled translational–rotational motion, and nonlinear interactions necessitate adaptive and statistically consistent diagnostic frameworks.

2.4 Related work

To address the limitations inherent in Gaussian-based performance metrics within industrial control systems, a novel framework was proposed that transforms non-Gaussian system outputs into Gaussian-like distributions through a quantile-based transformation (Meng et al., 2024). This methodology enables consistent benchmarking and tuning of controllers without the need for deep or black-box models. In the domain of high-dimensional remote sensing, a rotation-based iterative Gaussianization algorithm was introduced to convert multivariate data into a joint Gaussian form, utilizing marginal Gaussianization and orthonormal rotations (Johnson et al., 2021). This concept has been further extended by incorporating convolutional layers to scale the method for large image datasets, thereby facilitating statistical image analysis tasks, such as texture synthesis and entropy estimation (Laparra et al., 2022). The Gaussianization process has been optimized for real-time hyperspectral anomaly detection using field-programmable gate array (FPGA) hardware in a pipeline termed Optimized Rotation-Based Iterative Gaussianization (ORBIG), which bypasses traditional PCA and eigen-decomposition by employing fast correlation-based rotations and parallelized cumulative distribution function (CDF) computation (K. Yu et al., 2024). Complementing these advancements, an unsupervised invertible transformation was proposed for anomaly and change detection in multispectral imagery, where pixel likelihoods under a learned Gaussian model identify anomalies without the need for prior labels or parameter tuning (Padrón-Hidalgo et al., 2022).

In fault detection, data preprocessing and transformation have shown considerable potential for enhancing detection robustness under non-Gaussian conditions. Box–Cox transformations have been applied to facilitate Gaussian-based monitoring statistics within a multi-regime canonical correlation framework, resulting in improved generalization across operating modes (P. Wang et al., 2019). This transformation has also been employed to reinforce the assumptions of Gaussian-based classifiers in Heating, Ventilation, and Air Conditioning (HVAC) fault detection, thereby enhancing model reliability and interpretability (Tang et al., 2025). The efficacy of the Yeo–Johnson transformation, when combined with denoising resampling techniques, has

been highlighted in balancing skewed fault datasets in transformer diagnosis (Yusoff et al., 2025). Feature transformation has also been integrated into domain-specific sparsity metrics. A generalized Box–Cox-based sparsity measure was introduced to increase the sensitivity to weak fault signatures in early stage rotating machinery faults (T. Chen et al., 2025). In a related study, a generalized envelope analysis using multiple Box–Cox spectra was proposed and combined to enhance consistent fault peaks through a visualization technique termed PESOgram (B. Chen et al., 2023). These techniques improve fault localization even in noisy environments.

Unsupervised and one-class classification methodologies are employed when fault data are scarce or exhibit non-Gaussian characteristics. A hybrid fault detection framework was introduced, integrating Independent Component Analysis (ICA), autocorrelation-based component selection, and Support Vector Data Description (SVDD) modeling to characterize nominal behavior using only healthy data (M.-C. Chen et al., 2016). This methodology facilitates the interpretable localization of anomalies. Similarly, PCA has been utilized for dimensionality reduction, followed by kernel density estimation to enable fault detection through divergence measurement in alternating current (AC) motor currents (Giantomassi et al., 2015). A kernel density estimation (KDE)-based anomaly detector, trained exclusively on normal time-series data, has demonstrated superior performance compared to deep learning and statistical baselines while remaining lightweight and interpretable (Lang et al., 2022). This approach incorporates mechanisms for gradual concept drift and transfer learning, making it suitable for dynamic industrial environments.

The transfer of fault diagnosis models across domains poses a considerable challenge because of distributional discrepancies. This issue has been addressed by integrating Convolutional Neural Network (CNN)–Long Short-Term Memory (LSTM) architectures with adversarial training and semantic alignment layers to learn domain-invariant yet fault-discriminative features (Jang & Cho, 2021). In contrast, a statistical alignment approach has been introduced that evaluates not only the distribution means but also the variances using a metric termed Variance Discrepancy Representation (Qian et al., 2024). This method has proven to be more effective than traditional mean-matching strategies, particularly for heavy-tailed distributions, and incorporates a robust kernel to enhance resistance to outliers. Both approaches underscore that aligning deeper distributional structures across domains enhances the generalization of fault classification. Signal processing plays a pivotal role in extracting informative features from noisy data sets. The band-focused cyclic BCSM (CB-BCSM) was introduced to specifically target frequency ranges where bearing fault signatures are typically observed (D. Peng et al., 2023). By assuming α-stable noise and confining the sparsity analysis to pertinent bands, this approach more effectively isolates periodic impulsive components than full-spectrum methods. In ultralow-speed scenarios, where fault signals are weak and intermittent, iterative nonlinear filtering is employed to suppress periodic noise and amplify transients, followed by morphological envelope detection to extract characteristic fault frequencies (Z. Liu & Zhang, 2020). Frequency band selection was optimized using Gaussian mixture models to identify and exclude noise-dominated regions, resulting in interpretable features for reliable classification (Maliuk et al., 2021). A complementary strategy introduces a generalized Shannon entropy sparse wavelet packet transform that promotes the sparse representation of impulsive events, coupled with cyclostationary weighting to emphasize fault-related periodicities (L. Qin et al., 2024). A comprehensive comparative study demonstrated that classical time, frequency, and time-frequency features, combined with band filtering and simple classifiers, such as random forests, outperform deep models when tuned with domain expertise (Bienefeld et al., 2023). This underscores

the value of transparent handcrafted pipelines for rotating machinery diagnostics.

In power and energy systems, data-driven methodologies are increasingly used to enhance fault localization and classification. A combination of statistical feature engineering and rapid ensemble models has been employed to enable the real-time detection and classification of transmission line faults, with XAI techniques integrated to bolster trust and transparency (Noura et al., 2025). In power electronic converters, a slope-based current feature transformation was introduced to mitigate the impact of load variations in pulse-width modulation (PWM) rectifier fault diagnosis (Kou et al., 2020). This method utilizes random forests to consistently classify open-circuit faults across diverse operating conditions, demonstrating the benefits of targeted feature engineering and robust ensemble learning.

Chapter 3

The MOIRA–UNIMORE Bearing Dataset

The experimental study presented in this thesis employs the *MOIRA–UNIMORE Bearing Dataset for Independent Cart Systems*, an open-access benchmark specifically designed to address the scarcity of publicly available data that support reproducible and benchmarkable research on vibration-based diagnostics for ICS under controlled yet realistic operating conditions. This dataset constitutes the first publicly accessible collection of synchronized vibration and system variable measurements. It was curated at the *Lombardini Laboratory* of the Department of Engineering Sciences and Methods (DISMI) at the University of Modena and Reggio Emilia (UNIMORE), Italy, to assist the scientific community in advancing fault diagnosis and condition monitoring for nonstationary motion-control applications. The research was conducted within the framework of the European project Monitoring of Large-Scale Complex Technological Systems (MOIRA), from which the dataset derives its name.

The MOIRA–UNIMORE dataset was designed with three primary objectives:

1. to provide a controlled yet realistic experimental data for investigating bearing faults under coupled rotaional-translational motion;

2. to enable reproducible validation of advanced diagnostic algorithms, including time–frequency analysis, statistical transformation, and unsupervised learning methods; and

3. to promote open scientific exchange through structured data, comprehensive metadata, and transparent acquisition procedures.

The dataset includes vibration measurements from both *healthy* and *faulted* bearings recorded under various cart speeds, directions, and experimental configurations. Multiple test campaigns were conducted to replicate different industrial conditions, including inner race (IR) and outer race (OR) defects of varying severity, normal operations, and multi-cart interaction effects. Each experiment also included system variables such as cart position, velocity, and drive current, which enabled precise temporal and spatial synchronization of the vibration signals with the motion cycle. A summary of the dataset structure, acquisition setup, and variable specifications is provided in the following sections. Collectively, these data form the foundation for the diagnostic frameworks and transformation-based analyses developed in subsequent chapters.

3.1 Description of the Experimental Setup

This study employed an ICS based on the *GFX – Guidance System for Beckhoff Linear Transport Systems*, utilizing Beckhoff's *Extended Transport System (XTS)* technology (GFX-Guidance, 2025; XTS, 2025). The system consists of several essential components (Figure 3.1 and Figure 3.2): a fixed array of linear motors (stators), movable carts/movers equipped with permanent magnets and a set of bearings, a guide rail running parallel to the linear motors that facilitates the carts' movement, a set of five strategically positioned vibration sensors, an analog input acquisition system, a control circuit for programming and motion profiling of the movable carts, and software for real-time control and data management. This study primarily focuses on the mechanical aspects of the system that directly affect the vibration signatures and system variables.

3.1.1 Linear Motor Modules

The experimental setup employed a closed-loop track configuration comprising ten linear motor modules: eight straight modules and two 180-degree clothoid (curved) modules, as shown in Figure 3.1. Each straight module measured 250 mm in length, whereas each curved module extended 500 mm. This arrangement resulted in a total track length of 3000 mm, facilitating the carts' traversal of a complete loop under various test conditions. The track was segmented into four primary sections, each fulfilling a distinct function and contributing uniquely to the system dynamics.

3.1.1.1 Upper Straight Modules Section

This section extended between absolute positions of 750 mm and 1750 mm and was situated in the upper part of the system. A notable feature of this module is the presence of a lock that permits the addition or removal of carts from the guide rail. This lock introduces a physical discontinuity in the track, resulting in distinctive impulsive vibration signals caused by the bearing impacts as the cart traverses this area. Consequently, the vibration signals captured in this section differed significantly from those in the lower straight section, providing valuable insights into the response of the system to irregularities in the track.

3.1.1.2 Curved Modules

The track incorporates two curved modules located on the left and right. The first curved module connected the absolute positions of 250 and 750 mm, whereas the second spanned the positions between 1750 and 2250 mm. These modules introduce additional complexities to the cart motion, such as varying inertial forces owing to changes in direction and speed. The curved sections play a critical role in evaluating the performance of the system under nonlinear motion profiles.

3.1.1.3 Lower Straight Modules Section

Situated in the lower segment of the system, this straight motor module encompasses absolute positions ranging from 2250 to 3000 mm. In contrast to the upper section, this portion of the track is devoid of physical discontinuities, resulting in smooth vibration signals. A comparative analysis of the signals between the upper and lower straight sections underscores the influence of track features, such as locks, on the system dynamics.

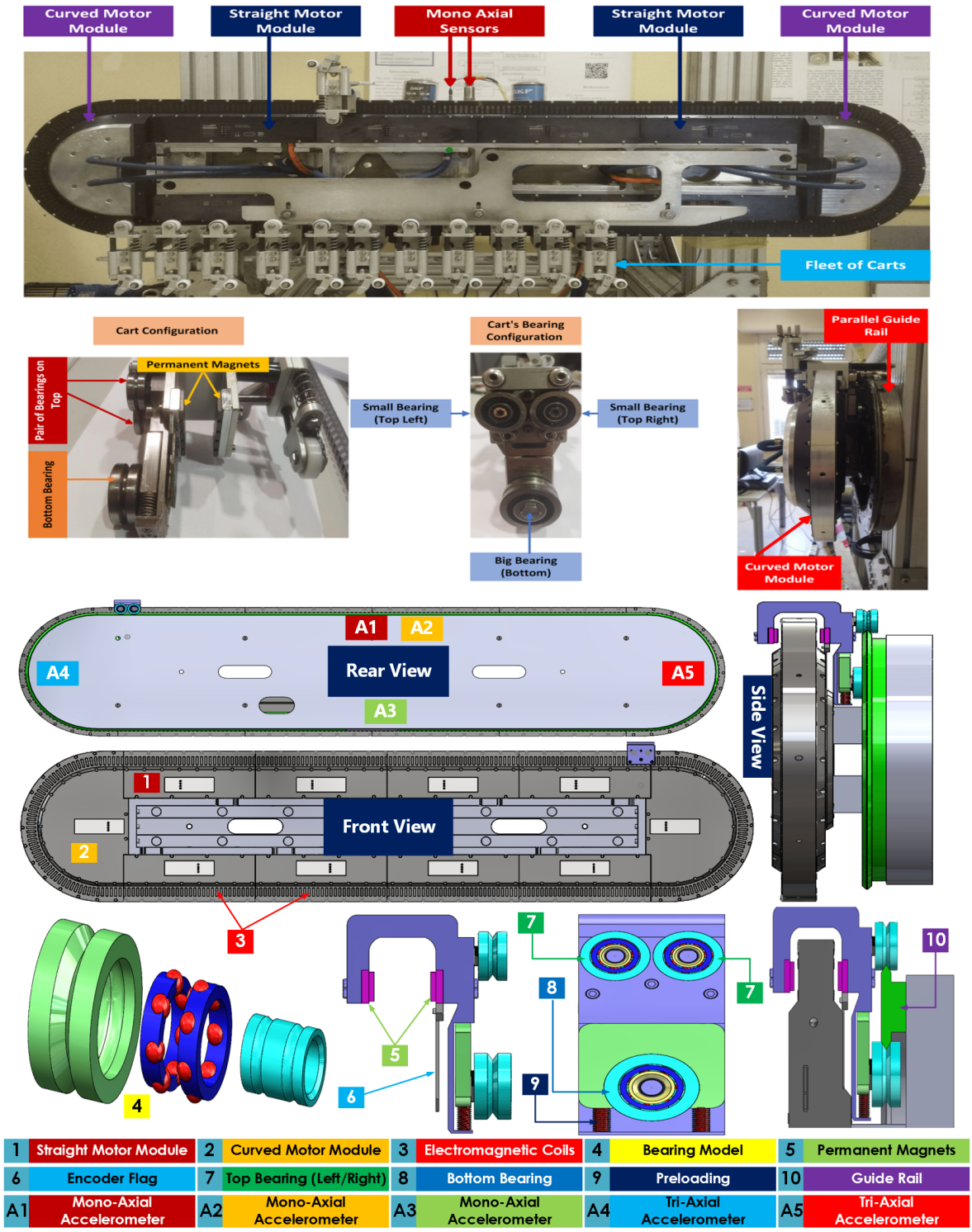


Figure 3.1: The experimental test rig consists of eight motor modules. Vibrations were assessed using mono-axial (A1–A3) and tri-axial (A4–A5) accelerometers, which were strategically positioned on the rail frame in proximity to the bearing–rail interfaces. The CAD model provides a detailed depiction of the cart configuration and its interaction with the guide rail.

3.1.2 Cart/Mover Configuration

Each cart within the system is equipped with four principal components that facilitate its functionality and ensure precise motion control:

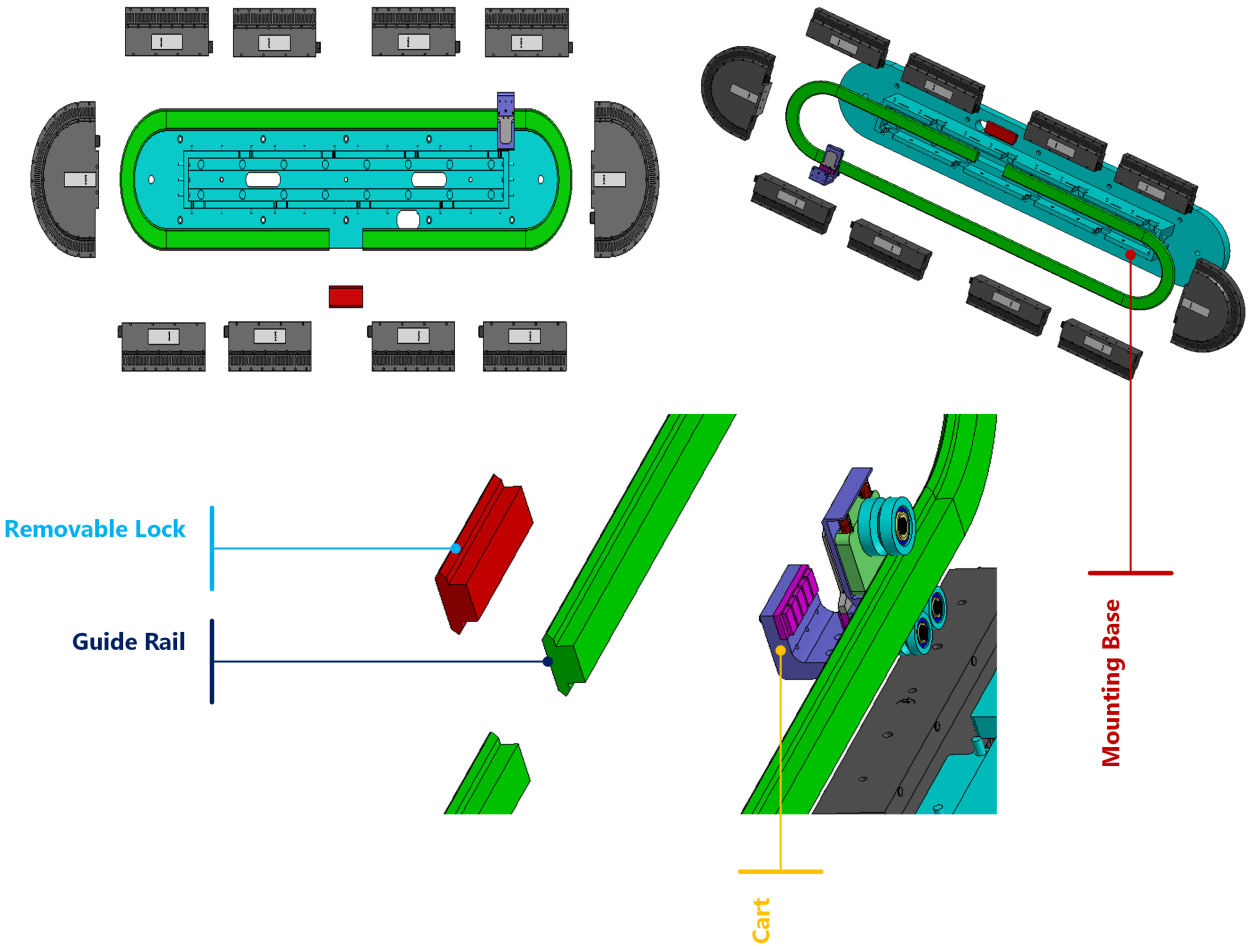


Figure 3.2: The CAD model of the experimental setup offers a modular perspective of the system. The removable lock, depicted in red, facilitates the addition and removal of carts. The contact between the bearing and rail, shown in green, which includes two upper bearings and one lower bearing, constitutes the primary source of vibration.

3.1.2.1 Magnetic Plates

Each cart was equipped with a pair of magnetic plates, each containing five magnets. These magnets define the cart's length along the track as 50 mm and interact with the magnetic field of the linear motor to facilitate controlled motion.

3.1.2.2 Encoder Flag

The cart was fitted with an encoder flag that provided real-time feedback on its actual position along the track. This ensures precise tracking and synchronization of cart movement with the overall system.

3.1.2.3 Rolling Element Bearings

The cart was supported by three rolling element bearings arranged to maintain stability and proper alignment with the guide rail. The configuration included two smaller, identical bearings positioned at the top and one larger bearing at the bottom. This arrangement secures the cart to the guide rail, thereby ensuring smooth and stable movement. All bearings shared the same

design characteristics: they featured two rows of deep grooves on the inside and a V-shaped groove on the outer race that matched the profile of the guide rail. In this setup, the outer race of the bearing rotates, whereas the inner race remains stationary. The dimensions and parameters of the bearings used in the experimental setup are detailed in Table 3.1, providing essential information for understanding the mechanical characteristics of the system. The bearings used in this study were part of the GFX Hepco guidance systems for the Beckhoff XTS. Detailed specifications regarding the bearing materials, including the types of materials in contact (e.g., steel-on-steel), are provided in the manufacturer's documentation (GFX-Guidance, 2025). This information is critical for understanding the material interactions and operational conditions of bearings during experiments.

3.1.2.4 Preloading Mechanism

A spring-like preload mechanism was incorporated to adjust the grip and evenly distribute the forces across the top and bottom bearings. This mechanism enhances stability and provides the necessary radial forces for the bearings, ensuring consistent contact with the guide rail under varying loads.

Table 3.1: Specifications for bearings are provided herein. All measurements were expressed in millimeters (mm). The abbreviations used are as follows: OR denotes the Outer Race, IR signifies the Inner Race, and Ø represents the Diameter.

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

Moreover, the cart can be tailored to incorporate an end-effector for payload transportation, provided that the weight remains within the specified maximum limits. Additionally, as the magnets of the cart do not directly contact the motor coils but instead interact via a magnetic field, friction and wear are substantially reduced during operation. Collectively, these attributes ensure that the cart delivers reliable, efficient, and precise motion while minimizing maintenance requirements, thereby rendering the system highly suitable for industrial applications. Although some defects, such as chipped magnets, have been reported, the primary focus of condition monitoring in this study was on bearing faults, as they constitute the most stressed components in this experimental setup.

3.1.3 Accelerometers

The experimental configuration incorporated five vibration sensors (refer to Table 3.2) strategically positioned to monitor the system's vibration signature. Two triaxial accelerometers were affixed to sections of the guide rail parallel to each of the two curved modules (left: 356B21, right: 356A02). In addition, two monoaxial accelerometers (353B18) were installed on the guide rail sections adjacent to the upper and lower straight modules, approximately midway along the track. Another monoaxial accelerometer (VSP001) was mounted on an aluminum rod directly attached to the guide rail on the upper side of the system and located midway along the straight section. This sensor arrangement ensured comprehensive coverage for capturing the vibration signals across different sections of the track.

3.1.4 Acquisition System

For data acquisition, we employed the Beckhoff ELM3602-0002 module (Beckhoff Automation GmbH & Co., Verl, Germany), a high-performance analog input terminal specifically designed for IEPE (Integrated Electronics Piezo-Electric) sensors and accelerometers. This module features a 24-bit resolution, ensuring precise and accurate measurements, and supports sampling rates of up to 50 kilo samples per second (ksps). Its high resolution and rapid sampling capability make it ideal for capturing detailed vibration signals and dynamic responses, which are critical for condition monitoring and fault diagnosis in industrial applications. This advanced hardware ensures that even subtle changes in the vibration signatures are effectively recorded, providing a robust foundation for data analysis and modeling.

Table 3.2: The accelerometers utilized in the ICS test rig are detailed with respect to their part number, manufacturer, sensing type, installation location, and a concise description.

Part Number	Manufacturer	Type	Location	Description
356B21	PCB Piezotronics, Depew, NY, USA	Tri-axial	Left curved module	Mounted on the left curved module of the guide rail
356A02	PCB Piezotronics, Depew, NY, USA	Tri-axial	Right curved module	Mounted on the right curved module of the guide rail
353B18	PCB Piezotronics, Depew, NY, USA	Mono-axial	Upper straight section	Mounted midway along the upper straight section of the guide rail
353B18	PCB Piezotronics, Depew, NY, USA	Mono-axial	Lower straight section	Mounted midway along the lower straight section of the guide rail
VSP001	ifm efector, Malvern, PA, USA	Mono-axial	Aluminum rod (upper straight)	Mounted on an aluminum rod attached to the guide rail, midway along the upper straight section

3.1.5 Motion Control and System Programming

The programming for motion control and data acquisition was conducted using the TwinCAT 3 programming environment, which offers advanced tools for real-time automation and control. However, the technical details of the programming process are beyond the scope of this study, as the primary focus is on the recorded dataset itself. This study aims to present the data alongside the experimental conditions under which they were collected, rather than delving into the complexities of motion control programming or data acquisition configurations. The emphasis is placed on the distinctive dataset and the methodologies employed to create it, ensuring that it serves as a valuable resource for researchers and practitioners.

3.2 Experimental Campaign

In this study, data collection was conducted over several months and encompassed multiple iterations of the experiments. Initially, various test runs were executed, and certain datasets were discarded owing to inadequate fault injection, insufficient resolution of the data acquisition system, or weak vibration signals that failed to yield meaningful insights. Furthermore, different experimental conditions were explored, including variations in the number of carts and the loading mechanisms. Based on these observations, the data acquisition system was upgraded to a 24-bit resolution with a 50 kHz sampling frequency. Faults were subsequently introduced

using laser-based precision techniques to ensure repeatability and accuracy of the fault injection process. The experimental campaign can be categorized into two primary aspects: fault injection and data acquisition under varying conditions, including different cart configurations, speed profiles, fault sizes, and fault locations. Fault injection experiments were further classified into two types: noninvasive fault injection, where the bearings remained intact, and invasive fault injection, where the bearings were disassembled before introducing faults. A summary of the experimental phases, challenges encountered, and improvements implemented is presented in Table 3.3. The fault injection methodology is discussed in detail in the following section.

Table 3.3: Phased experimental campaign: timeline, key issues, and implemented improvements.

Phase	**Description**	**Issues encountered**	**Improvements implemented**
Initial experiments	Faults injected manually; data collected with early acquisition system	Poor fault injection; low-resolution data; low sampling frequency	Baseline experiments with initial setup
Refined experiments	Different cart numbers and loading mechanisms tested	Variability due to inconsistent loads and cart configurations	Standardized loading mechanisms
Data acquisition upgrade	Upgraded to 24-bit resolution at 50 kHz sampling	Higher resolution but inconsistent manual fault sizes	Adopted laser-based fault injection
Laser fault injection	Laser fault injection without dismantling the bearings	Result variability due to partial fault coverage (inner–outer race proximity)	Bearings disassembled to ensure full coverage
Final experiments	Laser-based fault injection with bearing disassembly	—	Final dataset used for analysis

3.3 Fault Injection Methodology

Fault injection was conducted using two distinct methodologies: nondisassembled and fully dismantled bearings. Initially, faults were introduced using a laboratory drill mill without dismantling the bearings. However, the limited maneuverability of the drill-mill head, proximity of the bearing races, and inherent rigidity of the stainless-steel material posed challenges in creating significant faults. Consequently, the faults were irregular in shape and depth and did not fully encompass the grooves. To address this limitation, a picosecond laser (EKSPLA Atlantic 50) was employed to create controlled faults without dismantling the structure. The EKSPLA Atlantic 50 laser generated a Gaussian beam profile at an infrared (IR) wavelength of 1064 nm (see Figure 3.3). This laser source produces ultrashort pulses in the picosecond range, enabling the precise ablation of both the inner and outer races by creating grooves with accurate shapes. Notably, this process is not constrained by the hardness of the processed material. The diameter of the laser beam at the focal point was set to $\phi \approx 10\,\mu\text{m}$, measured at the $1/e^2$ intensity level. A dedicated optical path for the IR wavelength ensured the efficient delivery of the laser beam from the source to the scanning head. The system employed a Raylase SuperScan IV galvanometric scanner coupled with an 80 mm F-theta lens, providing a square working area with a side length of 39 mm. Movements outside this scanning area are facilitated by the translation of a stage in the X and Y directions, on which the bearings to

be processed are placed. Additionally, the Z-axis translation ensured that the bearings were processed at the correct focal height. The entire setup, located in the BrightLab laboratory of the DISMI Department, is illustrated in Figure 3.4. Preliminary tests were conducted to assess the response of the material to infrared radiation. These tests were crucial for identifying and defining the optimal process parameters to achieve damage with a specific geometry, dimensions, and prescribed depth. The laser parameters used in the process are summarized in Table 3.4. Although effective, this method presents challenges, particularly in damaging the inner and outer races. The first challenge stemmed from the shielding effect of the bearing components, which obstructed the laser radiation and limited defect creation along the full axial extension. The second difficulty was the formation of grooves with predetermined depths on inclined surfaces (Figure 3.5).

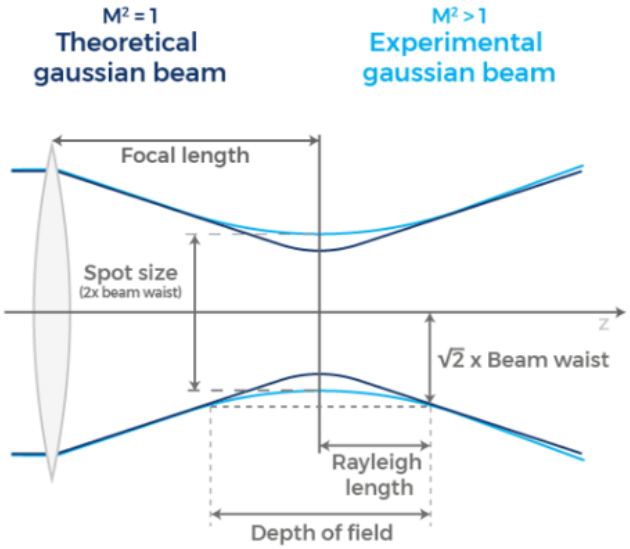


Figure 3.3: The profile of a picosecond laser beam is analyzed through a comparison between theoretical ($M^2 = 1$) and experimental ($M^2 > 1$) Gaussian beams. The theoretical beam demonstrated an ideal diffraction-limited focus, whereas the experimental beam exhibited increased divergence and a broader waist, attributable to the higher M^2 values. Essential parameters such as focal length, spot size, Rayleigh length, and depth of field are identified.

Table 3.4: Laser parameters used during experiments.

Parameter	**Unit of measure**	**Value**
Wavelength, λ	nm	1064
Average output power, P	W	13.32
Pulse frequency, f	kHz	300
Pulse energy, E	μJ	44.4
Pulse duration, τ	ps	10
Pulse fluence, F	J/cm^2	56.56
Line spacing, s	μm	5
Marking speed, v_s	m/s	1
Number of passes on each groove, p	—	150

To address these limitations and ensure consistent damage results across various bearing componentscomparable to those achieved in preliminary planar tests, an assessment of the allowable

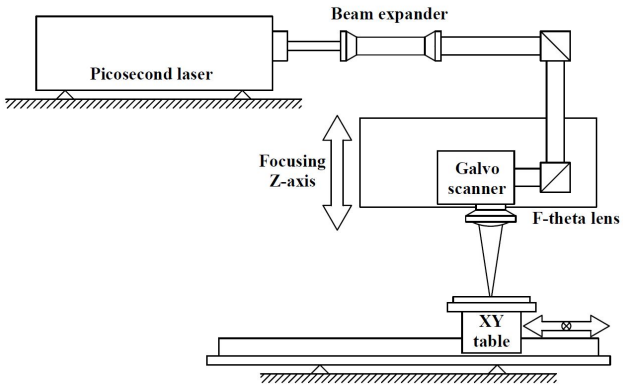


Figure 3.4: The schematic configuration of the picosecond laser ablation system is depicted. The arrangement included a picosecond laser, beam expander, galvanometer scanner equipped with an F-theta lens, and motorized XY table. The Z-axis facilitates precise focusing, while the galvanometer scanner directs the beam to achieve controlled ablation.

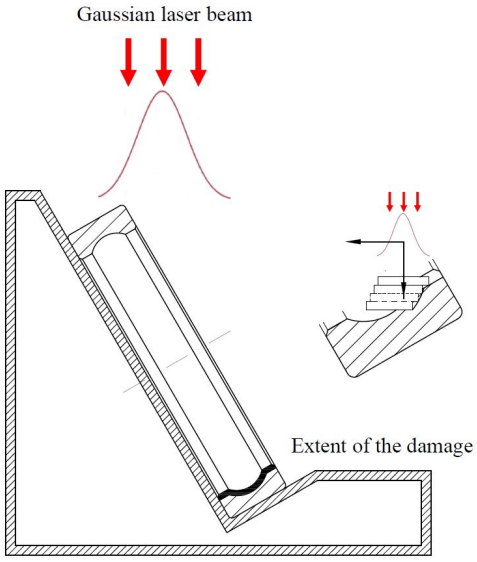


Figure 3.5: The arrangement of the bearings on a provisional support is depicted. The enlarged illustration elucidates the operational principle of the multi-altitude approach employed to achieve progressive focusing during the process of laser ablation.

depth of focus was necessary. Specifically, by assuming a beam quality factor M^2 of 1.5 and a laser beam diameter (measured at $1/e^2$ intensity) of $D_0 = 14\,\text{mm}$ on the lens, the following calculations were performed:

- Rayleigh length z_f, as the distance from the beam waist, where the beam radius is increased by a factor of $\sqrt{2}$:

$$z_f = \frac{\pi \left(\frac{D_0}{2}\right)^2}{M^2 \lambda} \approx 100\ \mu\text{m} \tag{3.1}$$

- Depth of field (D.O.F.) as the distance on either side of the beam waist, D_0, over which the beam diameter grows by 5%

$$\text{D.O.F.} = \pm 0.08\pi D_0^2 M^2 \lambda \approx 118 \ \mu\text{m} \tag{3.2}$$

The relatively small depth of field (D.O.F.) of our laser, calculated at 118 µm, rendered it impossible to damage the bearing components in a single laser pass. Consequently, a multipass approach was adopted. This method involves focusing a laser beam on a specific target area of the component to be damaged. The process is defined by achieving the desired depth in the focused region, while areas outside the focal zone remain untreated owing to insufficient energy deposition for proper material ablation. Consequently, the processing area was confined to a rectangle with a height equal to the previously calculated depth of the field. Regions outside this height are unprocessed because the laser beam is out of focus. To achieve the required depth incrementally, the height of the galvanometric head was adjusted, and the X position was compensated for by advancing the XY table. By refocusing the laser on a lower area, additional ablation can be performed. This process was repeated continuously until the desired damage geometry was achieved, with a depth tolerance of ±0.15 mm from the nominal shape of the panel.

To address the limitations discussed earlier, the bearings were disassembled completely. This allowed the individual components to be exposed, enabling precise and controlled fault injection. Disassembly eliminated obstructions, such as balls between the inner and outer races, which hindered access to the target areas. Consequently, more accurate and well-defined flat-bottom grooves were created using the multipass strategy described above. The final fault dimensions and geometry were characterized using an optical microscope (Nikon LV100ND). Detailed measurements were obtained by adjusting the depth of focus. Further methodological details on fault characterization can be found in Jabbar et al. (Jabbar, Mazzonetto, et al., 2024). Figure 3.6 shows a microscopic view of some faults created without disassembling the bearings of the test rig. As can be observed, the depth of the faults varies inconsistently owing to the challenges previously discussed.

3.4 Experimental Data Acquisition

The acquisition system utilized in this study incorporated five accelerometers, consisting of both monoaxial and triaxial types, which were categorized into three groups based on the number of channels and sensitivity levels. A monoaxial accelerometer (Sensitivity S1) with a sensitivity of 100 mV/g was installed midway along the upper straight section of the system. This sensor was not directly affixed to the guide rail; rather, it was mounted on an aluminum rod that was securely fastened to the guide rail. Two monoaxial accelerometers (Sensitivity S2), each with a sensitivity of 10 mV/g, were positioned midway along the upper and lower sections of the track. In addition, two triaxial accelerometers were mounted on the left and right curved modules, with each channel offering a sensitivity of approximately 10 mV/g. These sensors are strategically placed at critical points along the track to effectively capture the vibration signals. Along with the vibration data, system variables such as the cart position, speed, following error, and set current were recorded for each cart, thereby enriching the dataset for advanced condition monitoring and fault diagnostics. Detailed specifications for these accelerometers can be accessed on the manufacturer's website using the part numbers provided in the System description section.

The data acquisition strategy was meticulously designed to collect comprehensive datasets

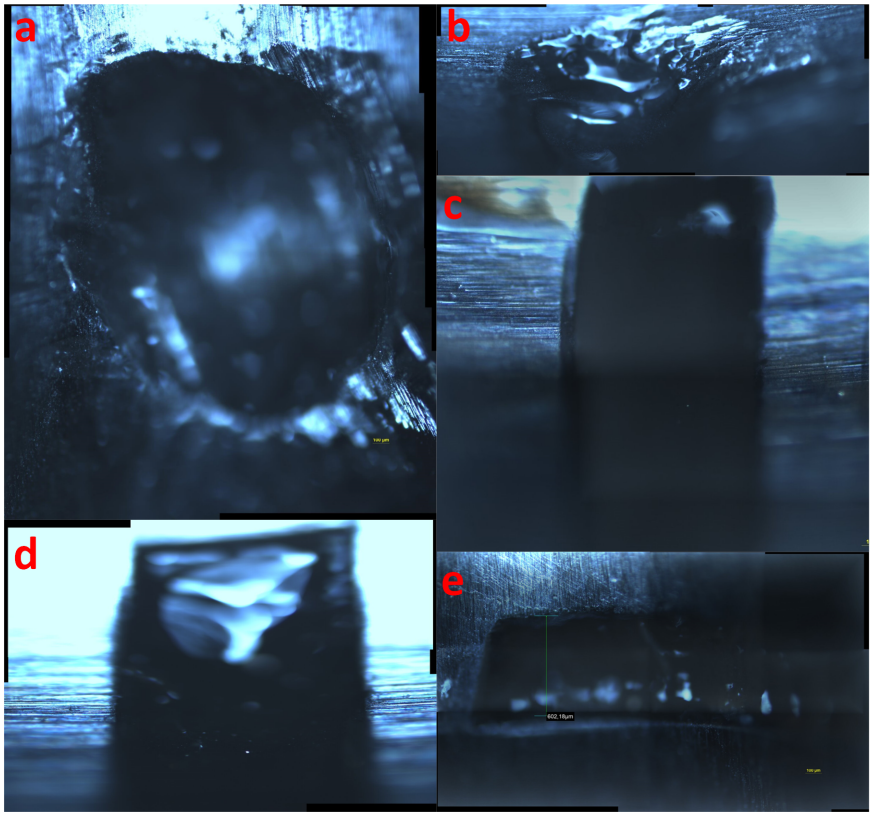


Figure 3.6: Microscopic images of artificially induced bearing faults: (**a**) inner race (IR) fault on the bottom bearing created using a handheld drill mill, (**b**) outer race (OR) fault on the bottom bearing created using a handheld drill mill, (**c**) laser-induced 0.5 mm OR fault on the top bearing, (**d**) laser-induced 1.0 mm OR fault on the bottom bearing, and (**e**) laser-induced 0.5 mm IR fault on the bottom bearing.

from an independent cart transport system operating under various fault conditions, speed profiles and cart configurations. The experiments were organized into two primary categories: single-cart and three-cart fleet experiments. Each category was further subdivided into four distinct experimental types, encompassing variations in the fault location, fault orientation, fault severity, motion pattern, and operating speed. These experimental classifications aimed to thoroughly capture the dynamic behavior of the system across diverse operational scenarios, as illustrated in Figure 3.7. This structured approach ensured that the data accurately represented the interplay between the system variables and fault characteristics in different configurations. The data were acquired at two cart speeds of 1000 and 2000 mm/s. A speed of 1000 mm/s was defined as the minimum speed at which the vibration signals became clearly observable and measurable. Below this threshold, the vibration signals were too weak or inconsistent for a reliable analysis. A speed of 2000 mm/s was selected as it represents half of the system's maximum operational speed. This allowed for a comparative analysis of how the vibration signals changed as the speed doubled, providing insights into the relationship between speed and vibration behavior. Both speeds were within the operational limits of the experimental setup, ensuring that the system could reliably achieve and maintain these speeds without compromising data quality or experimental integrity.

The experiments were systematically conducted under conditions with and without bearing

faults to ensure a comprehensive analysis. For small bearings, faults were consistently introduced on the top-right bearing, with only one bearing being faulty at any given time. Furthermore, across all experimental conditions, either the top or bottom bearing exhibited an inner race (IR) or outer race (OR) fault, thereby ensuring controlled and consistent fault conditions. The same cart was used for all tests, irrespective of the presence of faults. To maintain experimental consistency, identical speed profiles and program settings were applied to each type of experiment. The development of the test setup was an iterative process spanning several months, incorporating refinements such as enhanced fault injection techniques and improvements in the data acquisition. These aspects are elaborated upon in the experimental campaign section, which delineates the phases of data collection, improvements implemented, and rationale underpinning the experimental design.

3.4.1 Single-Cart Experiments

Experiments 1–4 were categorized as single-cart experiments, each meticulously designed to investigate distinct movement patterns along the track. Experiment 1 involved a back-and-forth motion between absolute positions of 500 and 2000 mm. Experiment 2 focused on the back-and-forth movement between absolute positions 250 mm and 750 mm, which corresponded to the right curved module. Experiment 3 encompassed back-and-forth movement between absolute positions 750 mm and 1750 mm, whereas Experiment 4 involved continuous circular motion along the entire track. All experiments were conducted under two speed profiles: 1000 and 2000 mm/s.

3.4.2 Three-Cart Fleet Experiments

Experiments 5–8 were classified as three-cart fleet experiments, each designed to capture the vibration signatures and system behavior when multiple carts operated simultaneously. Experiment 5 involved back-and-forth movement between absolute positions of 750 and 1750 mm, with each cart separated by 150 mm. In this configuration, Cart 1 initiates its motion from an initial position of 750 mm and proceeds to 1400 mm, whereas the third cart, originally positioned at 1050 mm, reaches 1750 mm. Experiment 6 featured continuous circular motion along the entire track, with carts separated by 250 mm. Similarly, Experiments 7 and 8 involved circular motion along the full track, with cart separations of 500 mm and 750 mm, respectively. All experiments were conducted at two speed profiles: 1000 and 2000 mm/s.

3.4.3 Bearing Defects

Bearing defects are classified into four categories: inner race fault of the top bearing (IR-Top), outer race fault of the top bearing (OR-Top), inner race fault of the bottom bearing (IR-Bottom), and outer race fault of the bottom bearing (OR-Bottom). Each fault category is further subdivided into specific fault severities. For instance, IR-Top faults include severities with widths of 0.25, 0.5, 1.0, and 1.5 mm, whereas OR-Top faults have severities of 0.25, 0.5, and 2.0 mm. Similarly, IR-Bottom faults have severities of 0.25 mm, 0.5 mm, and 1.0 mm, and OR-Bottom faults include severities of 0.25 mm, 0.5 mm, and 1.0 mm. Notably, the IR-Top and IR-Bottom fault categories were further classified based on the orientation of the fault with respect to the radial axis. For IR-Top, the fault orientations were categorized as 0°, 180°, and 270°. Similarly, for the IR-Bottom, the orientations were classified as 0°, 90°, and 270°. The fault signature analysis indicated that for IR-Top, the 270° orientation exhibited the strongest

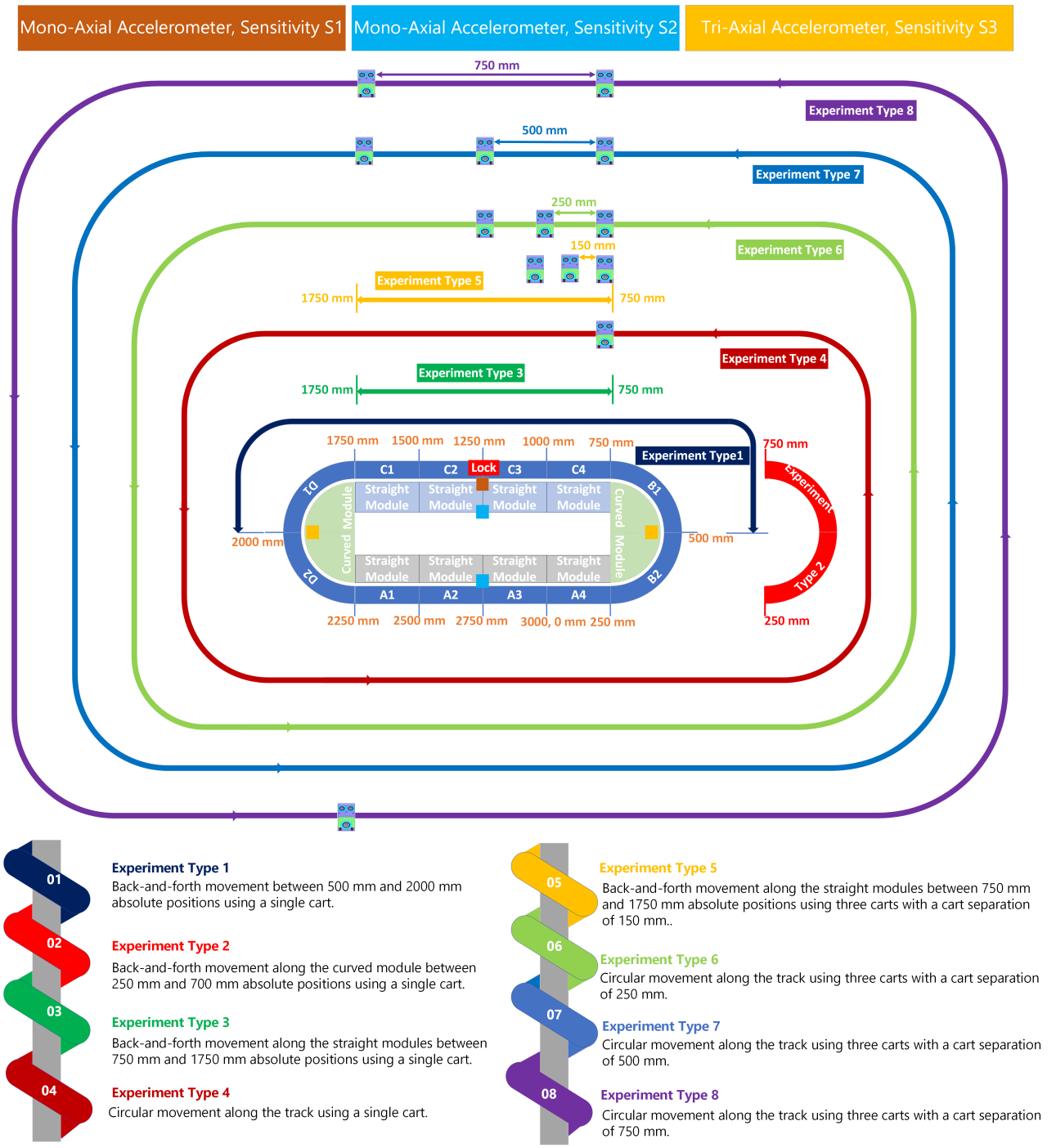


Figure 3.7: The schematic representation of the experimental configurations utilized in this study is depicted. Eight distinct experiments were conducted, involving back-and-forth movements on both straight and curved modules, as well as circular movements using one to three carts with varying relative separations between the carts (150, 250, 500, and 750 mm). The positions of mono-axial and tri-axial accelerometers with sensitivities S1, S2, and S3, respectively, along the rail are also indicated.

vibration signature, highlighting that this section of the inner race was subjected to greater stress and was therefore more prone to damage and wear. Likewise, for the IR-Bottom, the fault signature was strongest in the 90° orientation, indicating that this region experienced the highest stress levels and was the most vulnerable to damage. These observations underscore the

importance of fault orientation in determining the stress distribution and potential failure points within bearings. The variation in fault behavior between the top inner race (IR) and bottom IR, contingent on the fault orientation, is likely influenced by the spring-like preload mechanism integrated into the cart design. This mechanism exerts radial forces that draw the top bearings downward and propel the bottom bearings upward against the V-shaped edges of the guide rail, on which the bearings traverse. This force distribution ensures a stable connection between the bearings and rail, while inducing varying stress concentrations at different orientations. Consequently, the top bearing experienced increased stress at the 270° position, whereas the bottom bearing was subjected to greater stress at the 90° position. Comprehending these dynamics is essential for accurately diagnosing and predicting faults based on the vibration signatures associated with these orientations.

Remark 1. *It should be noted that fault orientation is not applicable to the outer race (OR) faults of the top and bottom bearings. This is due to the fact that the outer race rotates in accordance with the operational speed profile and cannot be maintained in a stationary position.*

Remark 2. *It is crucial to recognize that the data set includes folder names with inner race (IR) fault orientation labels such as 3-O'Clock, 6-O'Clock, 9-O'Clock, and 12-O'Clock. Specifically, 3-O'Clock corresponds to the* 0° *orientation, 12-O'Clock to* 90°*, 9-O'Clock to* 180°*, and 6-O'Clock to* 270°*,*

Remark 3. *In the context of three-cart experiments, the middle cart, designated as Mover M2, is consistently identified as the faulty one. Regardless of the experimental configuration, whether involving a single cart or three carts, only one of the three bearings on the faulty cart exhibits a defect at any given time, which may be either an inner race (IR) or an outer race (OR) fault.*

Figure 3.8 and Figure 3.9 provide a general overview of the experimental configuration, encompassing fault types, sizes, orientations, and speed profiles. Table 3.5 presents a comprehensive overview of the experimental configurations, detailing the fault types, sizes, orientations, and corresponding speed profiles across various experimental classes. This table also elucidates the types of experiments conducted under different conditions, thereby facilitating a clear understanding of the dataset structure and experimental conditions.

3.5 Data Set Repository

The MOIRA-Unimore Bearing dataset for Independent Cart Systems has been archived in Zenodo. Owing to upload limitations, the dataset was divided into five separate uploads, each with a distinct subtitle: *Experiment Types 1 and 2*, *Experiment Types 3 and 4*, *Experiment Types 5 and 6*, *Experiment Type 7*, and *Experiment Type 8*. These resources are freely available to the research community and can be accessed at **Experiment Type 1 and 2** (Jabbar, Cocconcelli, D'Elia, et al., 2025), **Experiment Type 3 and 4** (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025e), **Experiment Type 5 and 6** (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025b), **Experiment Type 7** (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025c), and **Experiment Type 8** (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025d). These datasets enable the replication of analyses or the extension of the findings presented in this study.

Remark 4. *The vibration and system signals were recorded using the TwinCAT 3 Measurement System, with each experiment archived as a distinct measurement project file bearing the* ***".svdx"*** *extension. These files encompass raw time-domain data, acquisition configurations,*

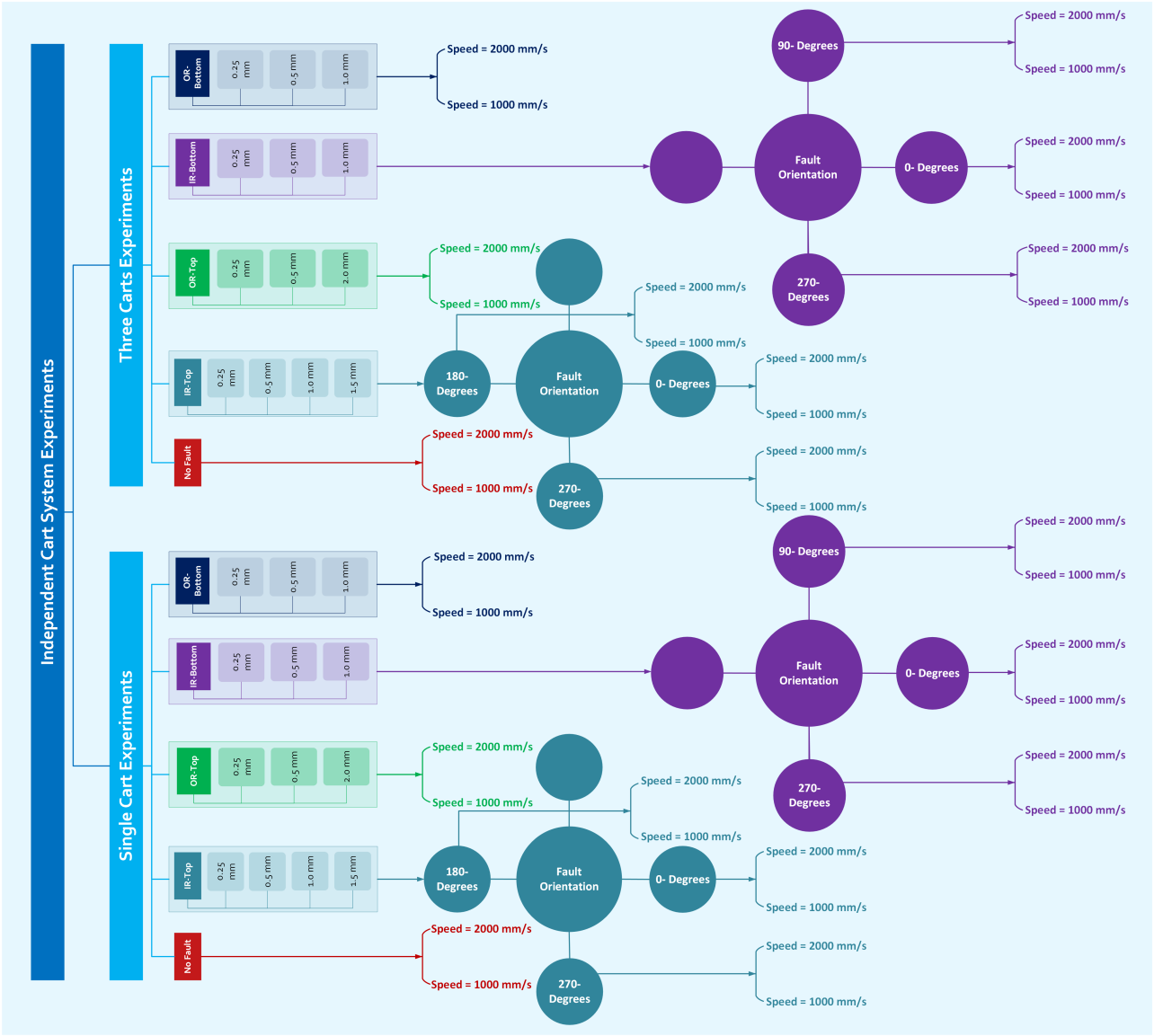


Figure 3.8: The graphical summary illustrates the experimental configurations of the ICS. The experiments were conducted under both single-cart and three-cart scenarios, encompassing various fault types (inner race, outer race, and no fault), fault sizes (0.25 mm, 0.5 mm, 1.0 mm, and 1.5 mm), four fault orientations (0°, 90°, 180°, and 270°), and two linear speeds (1000 mm/s and 2000 mm/s).

and the synchronized channel metadata from all the sensors. Appendix A offers comprehensive instructions on accessing these TwinCAT 3 projects, interpreting the variable naming conventions, and exporting the ***".svdx"*** *files into the preferred formats for subsequent analysis.*

3.6 Preprocessing

Given the extensive nature of the dataset, it is impractical to provide a comprehensive analysis of each experiment in this article. Instead, a selection of representative experiments is presented in this section to illustrate the behavior of the system under various experimental conditions and fault magnitudes.

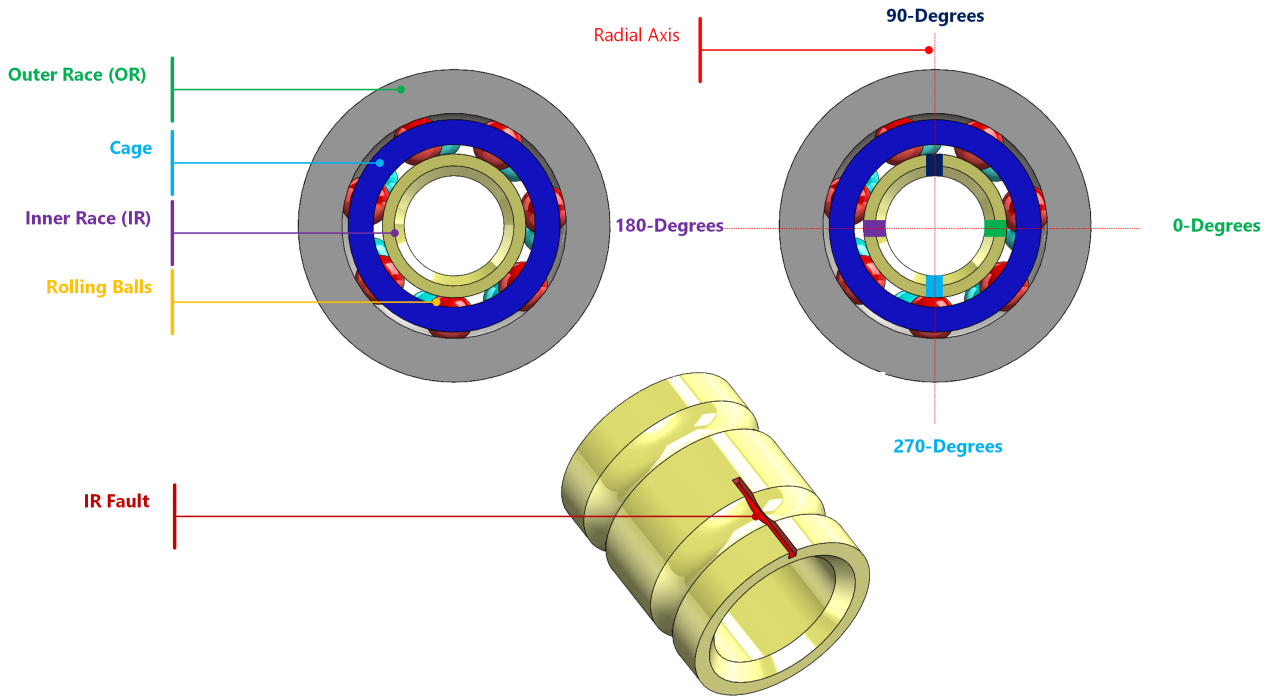


Figure 3.9: The schematic representation illustrates the bearing structure and the orientation of faults. The bearing components comprise the outer race (OR), cage, inner race (IR), and rolling elements of the bearing. The four reference orientations, specified as 0°, 90°, 180°, and 270°, are aligned along the radial axis, with the fault position on the inner race depicted in red.

.

The statistical features listed in Table 3.6 were extracted for each experimental realization, encompassing both fault-free and faulty conditions with varying fault sizes (refer to Table 3.7). For single-cart experiments, the dataset comprised five system variables: cart position, following error, actual velocity, velocity error, and set current of the cart. In addition, nine vibration channels were recorded from five different accelerometers, resulting in 156 features per realization (excluding cart speed and velocity, as these channels were not considered for feature extraction). In the context of three-cart experiments, the number of vibration channels remains constant, but the system variables increase threefold (one set per cart), leading to a higher-dimensional feature set than that in single-cart experiments. Given the substantial dimensionality of the dataset for each experiment, visualizing the impact of individual features is challenging. To address this, a feature ranking method was employed prior to the application of feature reduction techniques. Feature ranking aids in identifying the most significant features that contribute to the detection and characterization of faults, thereby reducing the computational complexity and enhancing the interpretability of the results. This approach ensures the retention of the most pertinent features for further analysis, thereby enabling efficient and accurate fault diagnosis.

3.6.1 Feature Ranking Methods

A total of five feature-ranking methods, encompassing both supervised and unsupervised approaches, were employed to assess the significance of features in distinguishing patterns and trends within the dataset, while simultaneously reducing computational costs for large datasets.

Table 3.5: The configuration matrix for the ICS experiments is detailed as follows. The rows delineate the experimental class, which includes either single-cart or three-cart configurations, and the speed profile, specified as either 1000 mm/s or 2000 mm/s. These are combined with the fault location, identified as either the inner race (IR) or outer race (OR) on the top or bottom bearings. The columns specify the fault size in millimeters, fault orientation (0°, 90°, 180°, and 270°; applicable only to IR), and movement protocol, categorized as Experiment Types 1–8. A checkmark (✓) denotes a configuration that was executed, whereas a cross (✗) indicates that it was not. The term "Not applicable" is used where orientation is undefined for OR faults.

Experiment Class	**Speed Profile**	**Fault Type**	**Fault Size (mm)**					**Fault Orientation**				**Experiment Type**							
	(mm/s)		**0.025**	**0.5**	**1.0**	**1.5**	**2.0**	**0°**	**90°**	**180°**	**270°**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Single Cart	1000	IR-Top	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
		OR-Top	✓	✓	✗	✗	✓		Not Applicable			✓	✓	✓	✓	✗	✗	✗	✗
		IR-Bottom	✓	✓	✓	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗	✗	✗
		OR-Bottom	✓	✓	✓	✗	✗		Not Applicable			✓	✓	✓	✓	✗	✗	✗	✗
	2000	IR-Top	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
		OR-Top	✓	✓	✗	✗	✓		Not Applicable			✓	✓	✓	✓	✗	✗	✗	✗
		IR-Bottom	✓	✓	✓	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗	✗	✗
		OR-Bottom	✓	✓	✓	✗	✗		Not Applicable			✓	✓	✓	✓	✗	✗	✗	✗
Three Carts	1000	IR-Top	✓	✓	✓	✓	✗	✓	✗	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
		OR-Top	✓	✓	✗	✗	✓		Not Applicable			✗	✗	✗	✗	✓	✓	✓	✓
		IR-Bottom	✓	✓	✓	✗	✗	✓	✓	✗	✓	✗	✗	✗	✗	✓	✓	✓	✓
		OR-Bottom	✓	✓	✓	✗	✗		Not Applicable			✗	✗	✗	✗	✓	✓	✓	✓
	2000	IR-Top	✓	✓	✓	✓	✗	✓	✗	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓
		OR-Top	✓	✓	✗	✗	✓		Not Applicable			✗	✗	✗	✗	✓	✓	✓	✓
		IR-Bottom	✓	✓	✓	✗	✗	✓	✓	✗	✓	✗	✗	✗	✗	✓	✓	✓	✓
		OR-Bottom	✓	✓	✓	✗	✗		Not Applicable			✗	✗	✗	✗	✓	✓	✓	✓

3.6.1.1 One-Way ANOVA

The One-Way Analysis of Variance (ANOVA) (Fisher, 1925; Montgomery, 2017; Neter et al., 1996) is a statistical technique utilized to identify significant differences among the means of two or more groups. In the context of feature ranking, ANOVA examines the relationship between a feature and the target variable and ranks features based on their capacity to differentiate between various classes or groups. This presupposes that the data within each group follow a normal distribution and that the variances are homogeneous across groups. Mathematically, the test statistic for One-Way ANOVA is derived by partitioning the total variance of the data into variance between groups and variance within groups. For a feature X divided into k groups with n total samples, the statistic F is calculated as

$$F = \frac{\text{Variance Between Groups}}{\text{Variance Within Groups}} = \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2} \tag{3.3}$$

where $\bar{X}_i$ is the mean of group i, $\bar{X}$ is the overall mean, n_i is the number of samples in group i, k is the number of groups, and X_{ij} is the j-th sample in group i. A higher F-statistic indicates that the feature effectively discriminates between groups, thereby enhancing its relevance for classification and regression tasks. The features are ranked based on their corresponding F-statistics.

Table 3.6: Classical time-domain statistical descriptors. Closed-form expressions for features computed per analysis window.

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

Notes: SNR: Signal-to-Noise Ratio; THD: Total Harmonic Distortion; SINAD: Signal-to-Noise And Distortion Ratio. x_i: ith sample in a window of length N; x_p: peak magnitude. Spectral powers P_1 (fundamental), P_h ($h \geq 2$), H (highest harmonic index); $P_{\text{dist}} = \sum_{h=2}^{H} P_h$; P_{noise} is the in-band residual excluding harmonics. Logarithmic ratios use $10\log_{10}$ (power quantities).

3.6.1.2 Kruskal–Wallis

The Kruskal–Wallis test (Kruskal & Wallis, 1952; Sheskin, 2003) is a nonparametric statistical method employed to determine whether multiple independent samples originate from the same distribution. This test is particularly useful for feature ranking in datasets where the assumptions of parametric tests, such as normality and homogeneity of variance, are not met. It evaluates whether there are statistically significant differences between the medians of the groups for a given feature. Mathematically, for a feature X with k groups and n total observations, the Kruskal–Wallis test calculates a test statistic H as

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{\left(\sum R_i\right)^2}{n_i} - 3(n+1) \tag{3.4}$$

where n_i represents the number of observations in group i, R_i denotes the mean rank of group i, and n is the total number of observations, given by $n = \sum_{i=1}^{k} n_i$. To compute H, the data

Table 3.7: Class labels and fault specifications. Codes denote bearing defect type and notch width in millimeters; IR = inner race, OR = outer race.

Label	Location	Width (mm)	Description
H	–	–	Healthy (no induced defect)
FIR025	Inner race	0.25	Narrow inner-race notch
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
FOR05	Outer race	0.5	Outer-race notch; moderate width
FOR20	Outer race	2.0	Outer-race notch; severe width

Notes: Class codes follow the pattern `F{IR/OR}{width}`. Defect depth was not specified because of the deep-groove geometry, and severity were compared solely based on the notch width.

were ranked across all groups, and these ranks were used to determine the mean rank R_i for each group. Under the null hypothesis that all groups are derived from the same distribution, the test statistic H approximately follows a chi-square (χ^2) distribution with $k-1$ degrees of freedom, assuming a large sample size. A higher H statistic suggests that the feature effectively distinguishes between groups, thereby enhancing its relevance for the feature selection. The features are ranked based on their corresponding H statistics, with higher values indicating greater discriminatory power.

3.6.1.3 Laplacian Score

The Laplacian score (He et al., 2005) is a feature ranking method designed to assess the relevance of individual features based on their capacity to preserve the intrinsic geometric structure of data. In contrast to methods that depend on labeled data, the Laplacian score is particularly effective for unsupervised learning because it identifies features that maintain locality information within the feature space. This characteristic renders it highly suitable for applications such as clustering, dimensionality reduction, and other tasks in which preserving the manifold structure of the data is essential. Given a dataset $X \in \mathbb{R}^{n \times d}$, where n represents the number of samples and d denotes the number of features, the method is initiated by constructing a similarity graph that represents the relationships between the data points. For two data points x_i and x_j, the similarity W_{ij} is computed using a Gaussian kernel, defined as

$$W_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & \text{if } x_i \in N(x_j) \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

where t is a scaling parameter, and $N(x_i)$ denotes the neighbors of x_i. The diagonal degree matrix D is then calculated as $D_{ii} = \sum_j W_{ij}$, and the graph Laplacian matrix L is derived as $L = D - W$. For a given feature F_j, the Laplacian score is mathematically expressed as

$$S_j = \frac{\sum_{i,j}(x_{j,i} - \bar{x}_j)(x_{j,j} - \bar{x}_j)W_{ij}}{\sum_i (x_{j,i} - \bar{x}_j)^2 D_{ii}} \tag{3.6}$$

where $x_{j,i}$ is the value of the j-th feature for the i-th sample and $\bar{x}_j$ is the mean of the j-th feature. Features with lower Laplacian scores are more effective in preserving the locality of

the data and are ranked higher, making them more suitable for tasks such as clustering and dimensionality reduction.

3.6.1.4 Variance

The variance ranking method constitutes a straightforward yet efficacious approach to feature selection, predominantly employed in unsupervised learning contexts. This method ranks features according to their variance, operating under the premise that features exhibiting higher variance encapsulate more information and are more likely to differentiate between data points. Variance is frequently used as a baseline method for feature selection because of its computational simplicity and intuitive appeal. For a given dataset $X \in \mathbb{R}^{n \times d}$, where n represents the number of samples and d denotes the number of features, the variance of the j-th feature F_j is calculated as

$$\sigma_j^2 = \frac{1}{n} \sum_{i=1}^{n} (x_{j,i} - \mu_j)^2 \tag{3.7}$$

where σ_j^2 denotes the variance of the jth feature, $x_{j,i}$ is the ith sample of the jth feature, and μ_j is the mean of the jth feature. Variance quantifies the dispersion of data points around the mean of each feature. Features with higher variance are ranked more favorably because they suggest greater variability and potential discriminatory capacity. Despite its simplicity, the variance ranking method has demonstrated utility across various applications, including dimensionality reduction, clustering, and preprocessing stages for more sophisticated feature-selection techniques. Nonetheless, it presupposes that high variance correlates with high relevance, a condition that may not invariably hold, particularly in datasets characterized by nonlinear relationships or noisy features.

3.6.1.5 Monotonicity

The monotonicity ranking method (Coble & Hines, 2009) assesses the extent to which a feature exhibits consistent changes over time or during degradation progression, making it particularly advantageous for fault diagnosis and prognostics. Monotonicity can be determined using two distinct approaches, contingent upon the chosen method: the Signum method and Spearman's rank correlation coefficient method. The Signum method evaluates monotonicity based on the consistency of the directional changes (increases or decreases) in a feature over time. Mathematically, it is expressed as

$$Monotonicity = \frac{1}{M} \sum_{j=1}^{M} \left| \frac{1}{N_j - 1} \sum_{k=1}^{N_j - 1} \operatorname{sgn}(x_j(k+1) - x_j(k)) \right| \tag{3.8}$$

where x_j denotes the vector of measurements for a feature on the j-th system, M represents the total number of systems monitored, N_j is the number of measurements on the j-th system, and $\operatorname{sgn}(\cdot)$ is the signum function. This method evaluates the proportion of consistent changes in features across all monitored systems, with higher values indicating stronger monotonicity. The Spearman rank method calculates monotonicity based on the correlation between the ranks of feature measurements and their corresponding time points. The formula for monotonicity using this method is

$$Monotonicity = \frac{1}{M} \sum_{j=1}^{M} |\operatorname{corr}(\operatorname{rank}(x_j), \operatorname{rank}(t_j))| \tag{3.9}$$

where x_j represents the feature measurement vector on the jth system, t_j is the vector of time points corresponding to the measurements x_j, M is the total number of systems monitored, and rank$(\cdot)$ denotes the ranking of elements in a vector.

3.6.2 Dimensionality Reduction

To address the high dimensionality inherent in the dataset, PCA and t-SNE were employed because of their complementary strengths and prevalent use in the scholarly literature. PCA is a well-established, computationally efficient technique for linear dimensionality reduction, particularly effective in capturing global structures within high-dimensional data by identifying orthogonal directions of maximum variance (Jolliffe & Cadima, 2016; Pořízka et al., 2018). Mathematically, PCA identifies these orthogonal directions, or principal components, by solving the eigenvalue decomposition of the covariance matrix of the data, thereby ensuring the retention of most of the information in a reduced set of dimensions. In the context of condition monitoring and fault diagnosis, PCA has proven invaluable for noise reduction and highlighting dominant patterns within the data (Kitao, 2022). Conversely, t-SNE excels at preserving local structures and revealing clusters in high-dimensional data, making it highly suitable for visualizing complex relationships. t-SNE minimizes the Kullback–Leibler divergence between the high- and low-dimensional distributions of pairwise similarities. This approach is particularly effective for uncovering clusters, patterns, and relationships in complex datasets, rendering it well-suited for this study (Jung et al., 2024; van der Maaten & Hinton, 2008). Unlike PCA, t-SNE can capture nonlinear relationships, which are often critical for understanding fault patterns in mechanical systems (Gove et al., 2022). By employing both PCA and t-SNE separately on the same dataset, the unique strengths of each method are leveraged. PCA was utilized for efficient dimensionality and noise reduction, facilitating the identification of dominant patterns and global structures within the data. In contrast, t-SNE is applied for detailed visualization and cluster analysis because it excels at preserving local structures and revealing nonlinear relationships. This dual approach allows for a comprehensive exploration of the dataset from different perspectives, thereby enhancing the ability to robustly detect and diagnose faults.

3.6.3 Speed Profiles

The speed profiles observed in these experiments demonstrated non-synchronous behavior, particularly during the back-and-forth motion. This is characterized by rapid speed fluctuations as the cart decelerates to a stop before reversing its direction and subsequently accelerates until it regains its nominal speed. This phenomenon becomes increasingly pronounced as the nominal speed increases. Figure 1.5 illustrates the speed profiles recorded under various experimental configurations. Specifically, Figure 1.5 depicts a nominal speed profile of 1000 mm/s along the curved module of the system, representative of Experiment Type 2. This speed profile was consistent with other experiments involving back-and-forth movements, such as Experiment Types 1, 3, and 5. The primary distinction is that Experiment Type 5 involves three movers, whereas Experiment Types 1, 2, and 3 involve only one mover. Conversely, 3.10 represents the characteristic speed profiles of Experiment Types 4, 6, 7, and 8. In Type 4, a single mover continuously circulates along the track, whereas Types 6, 7, and 8 involve three movers executing similar continuous movements. The comparison across these profiles underscores the impact of varying the number of movers on the dynamic behavior of the system while maintaining consistency in the movement patterns.

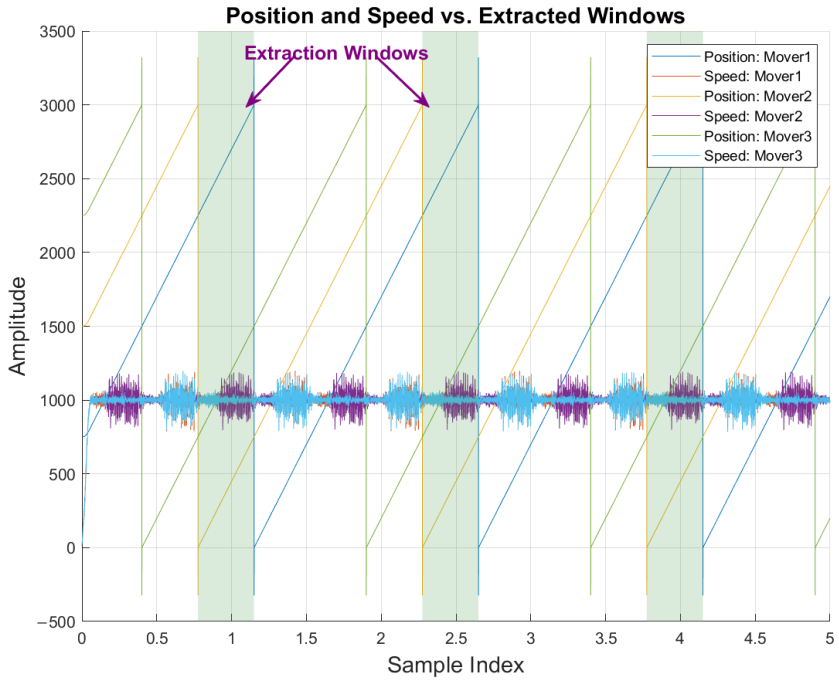


Figure 3.10: The position and velocity trajectories for three carts are depicted, with the selected extraction windows highlighted by green bands. The legends denote the position and velocity of each cart. These windows delineate the segments utilized for subsequent analysis.

3.7 Data Visualization and Analysis

The results of selected single-cart and three-cart experiments are presented, focusing on varying defect sizes and IR fault orientations. The analysis examined key patterns in the vibration signals and system variables, offering a comparative perspective across different experimental configurations. The original feature set, as summarized in Table 3.4, was computed using all channels, including the following error, velocity error, set current for the movers, and vibration channels, while excluding the position and speed of the cart. The exclusion of the position and speed channels is based on the assumption that they are periodic and deterministic in nature. Consequently, although the reduced feature set excludes these deterministic channels, it does not solely represent the vibration data, as it still incorporates other system-level variables. The defect labels corresponding to the experiments are detailed in Table 3.5, which provides clarity on the types of faults and their respective sizes.

3.7.1 Single Cart Experiments

Scatter plots derived from Experiment Types 1 and 2 are presented to analyze the distribution of features in single-cart configurations. All plots correspond to a nominal cart speed of 1000 mm. For Experiment Type 1, only the PCA scatter plots are included, as the no-fault profile (or normal profile) can be easily distinguished from the abnormal profile in most fault categories using PCA. Figure 3.11 and Figure 3.12 illustrate that the PCA features corresponding to the healthy bearing state form a distinct cluster that is clearly separated from the IR fault categories of all sizes. This distinction highlights the ability of PCA to reduce dimensionality while preserving the separability of fault data. However, the OR fault categories appeared closer to the no-fault boundary, indicating overlapping feature distributions. Although this overlap

suggests challenges in differentiating OR faults from the healthy state using PCA alone, further processing or advanced clustering techniques could help achieve a more compact and distinct representation of these fault classes. For Experiment Type 2 (Figure 3.13 and Figure 3.14), both PCA and t-SNE scatter plots were used to provide a more comprehensive visualization of the feature sets. These plots represent experimental data for which the IR fault orientation information was unknown because it was not recorded during the experiment. PCA applied to datasets ranked by one-way analysis of variance (ANOVA) and Kruskal–Wallis methods resulted in highly scattered features, with limited separability observed across all fault classes of the top bearing, including IR and OR faults. This outcome indicates the limitations of PCA when applied to specific feature rankings. In contrast, PCA applied to datasets ranked by unsupervised methods (Laplacian Score, Variance, and Monotonicity) yielded a bimodal distribution, suggesting a better representation of the underlying structure for these rankings. Similarly, the t-SNE-reduced feature sets consistently exhibited bimodal distributions across nearly all distance metrics, including Mahalanobis, Cosine, Chebyshev, and Euclidean. These visualizations demonstrate that t-SNE effectively captures the underlying patterns within the feature set, revealing separability across fault categories even when different distance metrics are applied. This highlights the importance of dimensionality reduction techniques, particularly nonlinear methods such as t-SNE, for the visualization and analysis of high-dimensional datasets. However, the bimodal nature of the distributions observed in both PCA and t-SNE suggests that further processing is required to transform feature distributions into Gaussian-like clusters. Such transformations could result in more compact clustering and enhanced classification performance, which remains an area for future research.

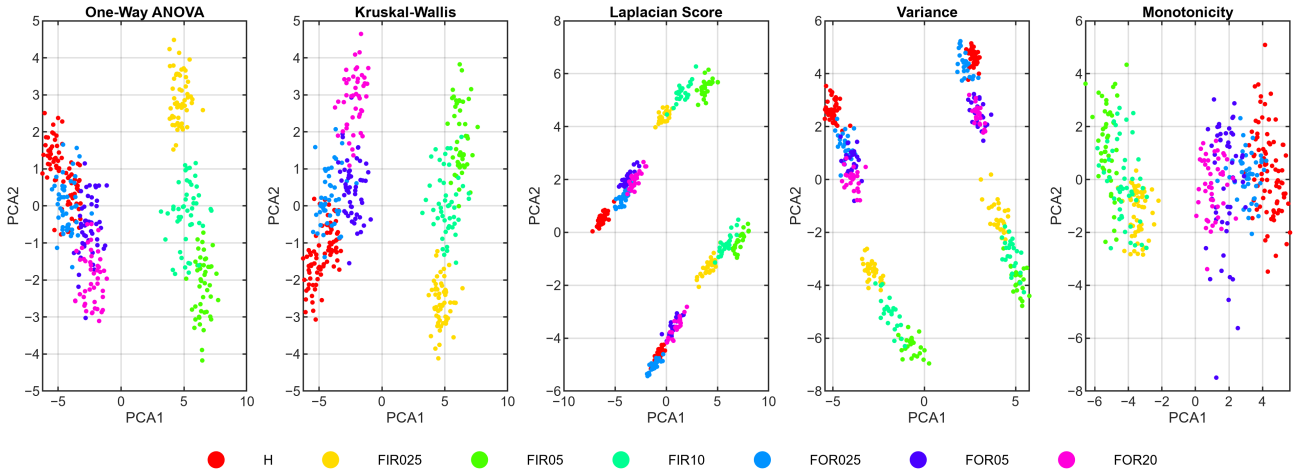


Figure 3.11: **Experiment Type1:** PCA scatter plots (PCA1 vs. PCA2) depict the IR and OR faults of the **top** bearing, with the IR fault positioned at 3 O'Clock. The feature rankings include One-Way ANOVA, Kruskal–Wallis, Laplacian Score, Variance, and Monotonicity.

3.7.2 Three Cart Experiments

In the experiments involving three movers, the plots (Figure 3.15 and Figure 3.16) pertain to Experiment Type 8, conducted at a nominal cart speed of 1000 mm/s, with a relative distance of 750 mm between the movers. The PCA-reduced feature set effectively differentiated the normal profile from the FIR025 and FIR05 fault categories of the top bearing across all ranking methods, except for the Laplacian score. In this instance, the Laplacian Score performed inadequately, failing to provide meaningful separability between the fault and no-fault states. This limitation may be attributed to the unsupervised nature of the Laplacian score, which

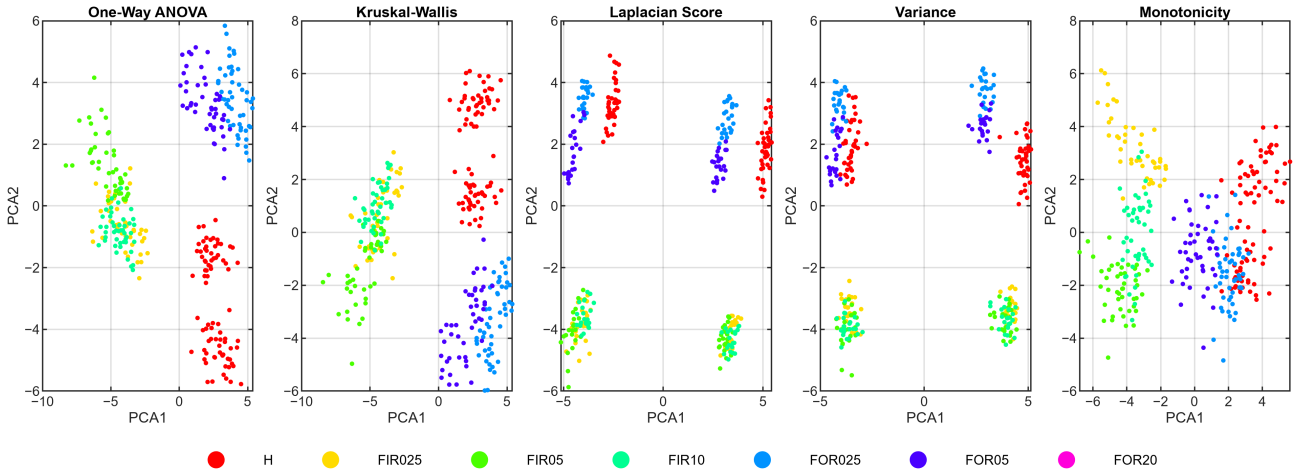


Figure 3.12: **Experiment Type1:** PCA scatter plots (PCA1 vs. PCA2) depict the IR and OR faults of the **bottom** bearing, with the IR fault positioned at 3 O'Clock. The feature rankings include One-Way ANOVA, Kruskal–Wallis, Laplacian Score, Variance, and Monotonicity.

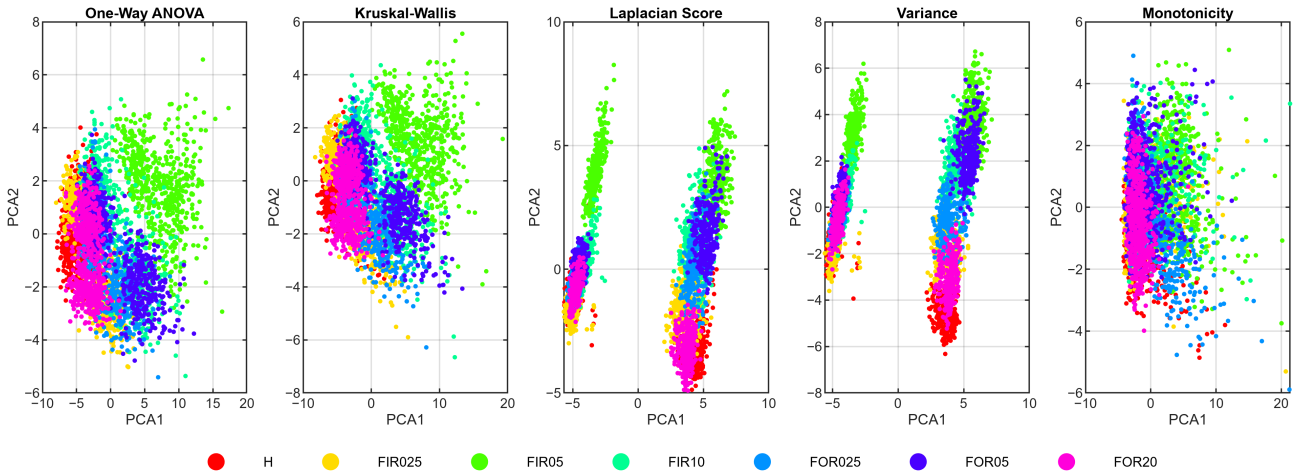


Figure 3.13: **Experiment Type 2:** PCA (PC1 vs. PC2) scatter plots are presented to illustrate the IR and OR faults of the top bearing at varying severities, without accounting for the orientation of the IR fault. The features are ranked using One-Way ANOVA, Kruskal–Wallis, Laplacian Score, Variance, and Monotonicity.

prioritizes the preservation of the local data structure but may struggle to differentiate fault classes in complex datasets. Furthermore, the FIR10 fault category appears closer to the OR fault classes in the PCA-reduced space, indicating some overlap in feature distributions. This overlap underscores the challenges associated with using PCA to separate fault types, particularly when the severity of the fault increases or when the faults share similarities in their vibration signatures. In contrast, the t-SNE scatter plots offer superior fault separability for Experiment Type 8, providing a clearer visualization of the distinctions between normal and fault conditions. The t-SNE-reduced feature set demonstrated robust clustering across nearly all distance metrics, including Mahalanobis, Cosine, Chebyshev, and Euclidean. This improved separation underscores the strength of t-SNE in capturing nonlinear relationships within high-dimensional feature sets. Notably, t-SNE resolves the overlap between the FIR10 and OR fault classes observed in the PCA plots, providing a more distinct representation of the underlying feature space. These results highlight the importance of selecting appropriate dimensionality reduction techniques and distance metrics for the effective analysis of fault data in multicart

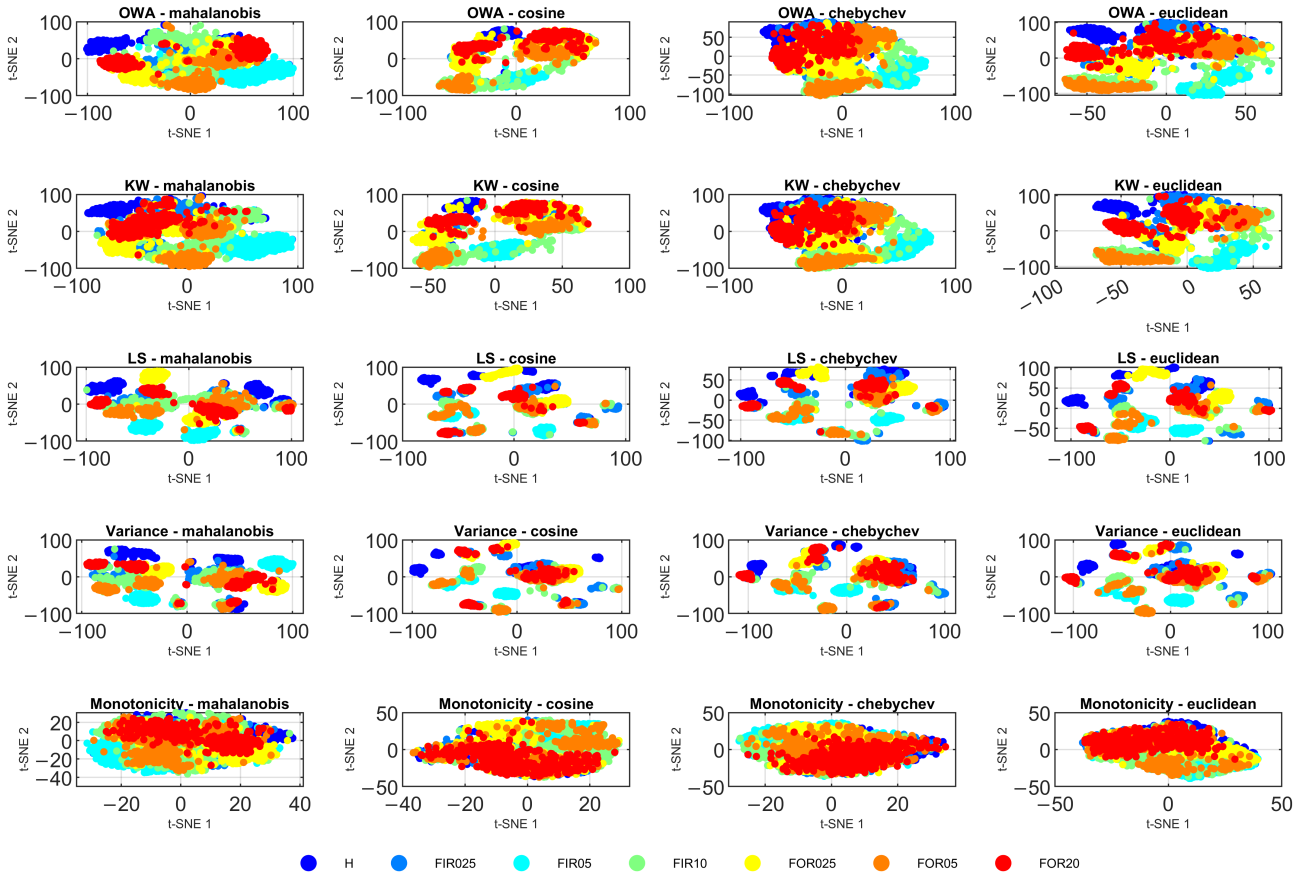


Figure 3.14: **Experiment Type 2:** t-SNE embeddings (2D) of the ranked feature sets were generated using four distance metrics: Mahalanobis, cosine, Chebyshev, and Euclidean. These embeddings illustrate the IR and OR faults of the top bearing without considering the orientation of the IR fault. The panels are organized according to the ranking method employed.

configurations. Further research should focus on refining clustering techniques to enhance the compactness and separability of fault categories, particularly in cases where fault distributions overlap.

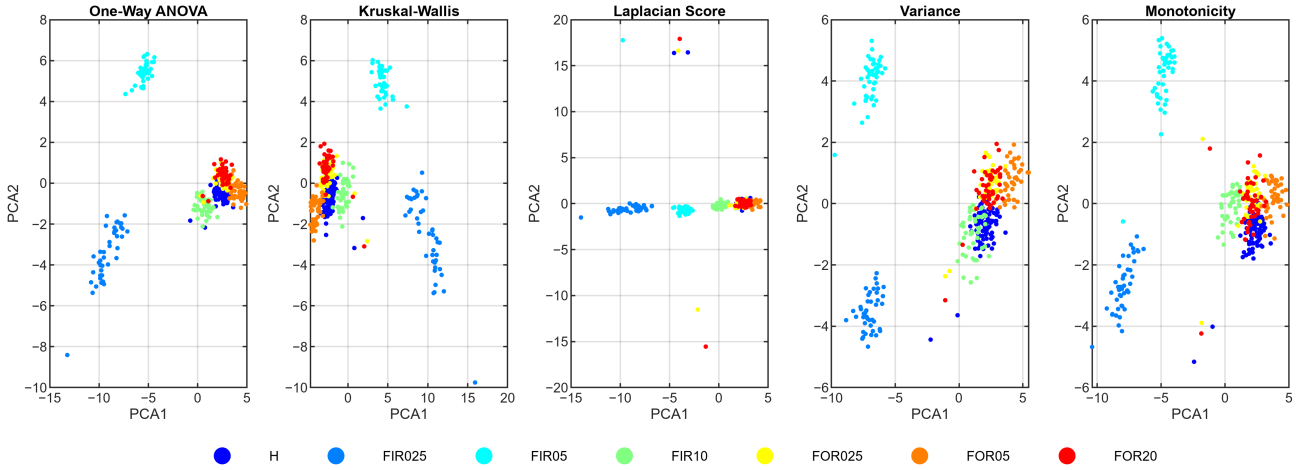


Figure 3.15: **Experiment Type 8:** Principal Component Analysis (PCA) score plots (PC1 vs. PC2) for ranked features are presented, illustrating the Inner Race (IR) and Outer Race (OR) faults of the top bearing, with the IR fault oriented at 3 O'Clock. The features are computed for each complete cycle of the M1 (first) mover/cart.

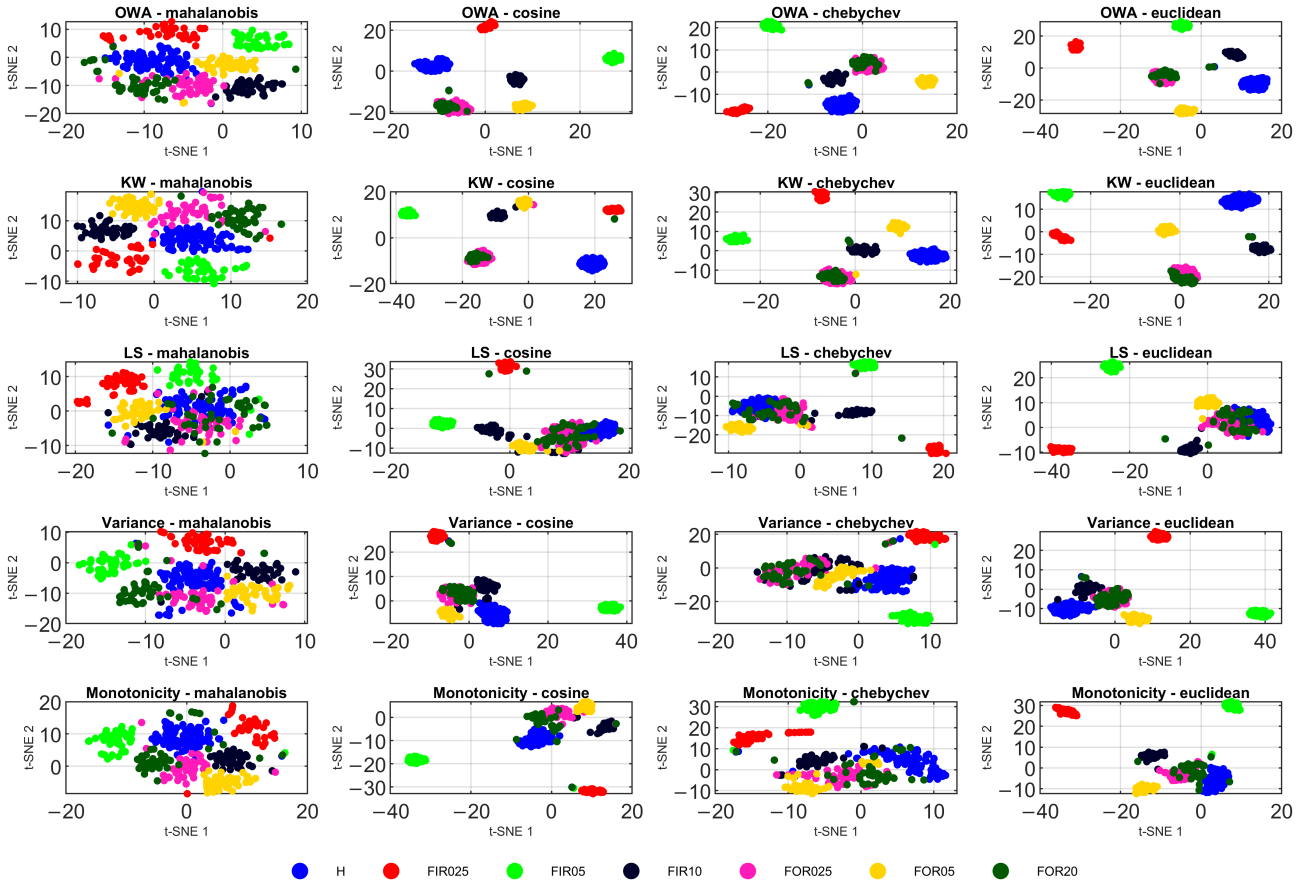


Figure 3.16: **Experiment Type 8:** The t-SNE embeddings (2D) for ranked features are presented, illustrating the IR and OR faults of the top bearing, with the IR fault oriented at 3 O'Clock. This analysis employed four distance metrics: Mahalanobis, cosine, Chebyshev, and Euclidean. The features are computed for each complete cycle of the M1 (first) mover/cart.

Chapter 4

Bimodal Detection and Distribution Transformation

The efficacy of any fault diagnosis framework is significantly influenced by the compactness and consistency of the extracted features that represent the underlying system state. In practical industrial settings, particularly under variable speed or load conditions, features derived from vibration or process signals frequently exhibit non-Gaussian, skewed, or multimodal distributions. These irregularities result in overlapping class regions, increased intra-class variance, and ambiguous decision boundaries for one-class or clustering-based anomaly detectors. While traditional normalization and scaling operations ensure numerical consistency, they do not alter the *statistical shape* of the distribution. This limitation necessitates the development of specialized *distribution transformation algorithms* that reshape marginal feature distributions into more compact, regular, and Gaussian-like forms, thereby enhancing separability and interpretability in unsupervised contexts.

The presence of non-Gaussian characteristics in process and vibration data poses challenges in assessing control performance, reliability analysis, and fault diagnosis. This has led to various distribution transformation strategies for reshaping and Gaussianizing data to enhance interpretability. Initial efforts focused on quantile-based and power-transformation methods to approximate Gaussianity. A quantile transformation framework converting non-Gaussian system outputs into Gaussian-like representations while preserving mutual information between inputs and outputs was introduced in (Meng et al., 2024).

Subsequent advances produced Gaussianisation techniques for high-dimensional datasets. The Rotation-Based Iterative Gaussianization (RBIG) algorithm in (Johnson et al., 2021) applies marginal Gaussianization and orthonormal rotations for joint Gaussian representation. This was expanded to convolutional RBIG in (Laparra et al., 2022), where learned convolutional rotations enabled scalable Gaussianization. An optimized FPGA implementation of RBIG (ORBIG) was presented in (K. Yu et al., 2024) for real-time hyperspectral anomaly detection. A fully unsupervised invertible Gaussian model for change detection was introduced in (Padrón-Hidalgo et al., 2022), transforming multivariate data via differentiable mappings. These studies demonstrate distribution Gaussianization's efficacy as an interpretable method for aligning data with Gaussian assumptions.

In industrial fault diagnosis, Box–Cox and Yeo–Johnson transformations remain prevalent for addressing skewed monitoring statistics. Box–Cox normalization has been used in multi-regime canonical correlation analysis (P. Wang et al., 2019) for fault detection across operating conditions. Box–Cox preprocessing improved classifier reliability in HVAC sensor fault detection (Tang et al., 2025), while Yeo–Johnson transformations with resampling addressed transformer fault datasets (Yusoff et al., 2025). An improved Box–Cox sparsity measure was proposed in (T. Chen et al., 2025) to increase sensitivity to fault signatures. Work in (B. Chen et al., 2023) generalized the envelope spectrum through Box–Cox transformation, consolidating spectra into a product envelope spectrum for amplifying fault frequencies. These studies show feature transformations can enhance fault indicators without complex models.

Unsupervised and one-class methodologies have investigated nonparametric density modeling to characterize nominal system behavior in non-Gaussian environments. The ICA–DW–SVDD framework in (M.-C. Chen et al., 2016) extracts independent components, selects decorrelated features using Durbin–Watson criterion, and encapsulates nominal data through one-class Support Vector Data Description. Kernel density estimation (KDE) was applied in (Giantomassi et al., 2015) to model healthy motor current signals and identify anomalies using divergence measures. The KDE-based anomaly detector in (Lang et al., 2022) extends this to multivariate time-series monitoring and learns from normal process data while adapting to concept drift. These methods aim to construct distributions approximating stable Gaussian-like structures, from which deviations indicate anomalies.

The Variance Discrepancy Representation (VDR) metric in (Qian et al., 2024) aligns both mean and variance of feature distributions between source and target domains using a robust kernel function. This approach shows greater efficacy than mean-matching techniques under heavy-tailed distributions, highlighting that controlling higher-order distributional properties is essential for fault diagnosis under variable conditions.

Overall, these contributions provide a conceptual foundation for the present study, which develops an adaptive, interpretable transformation for reshaping non-Gaussian feature distributions in vibration-based diagnostics. This chapter presents two novel and computationally efficient feature transformation frameworks: Bimodal Distribution Detection and Transformation (BDD–T) (Algorithm 1) and its adaptive extension, Adaptive Distribution Transformation (ADT). Both methodologies function in a feature-wise, geometry-agnostic manner and do not require supervised model training, making them suitable for real-time and label-scarce industrial applications.

The BDD–T framework is primarily concerned with detecting and reshaping bimodal marginal feature distributions using kernel density estimation (KDE), clustering, and distribution transformation. Its comprehensive workflow, from raw signal acquisition to transformed feature representation, is depicted in Figure 4.1. The pipeline includes window segmentation, feature extraction, feature ranking, and dimensionality reduction, followed by bimodal distribution reshaping.

Building on this foundation, the Adaptive Distribution Transformation (ADT) extends the distribution-aware reshaping strategy to address both bimodal and high-variance unimodal feature distributions in a fully adaptive manner. ADT is implemented through two complementary processing variants, each tailored to accommodate distinct analytical objectives. The first variant applies the transformation directly within the original ranked feature space with-

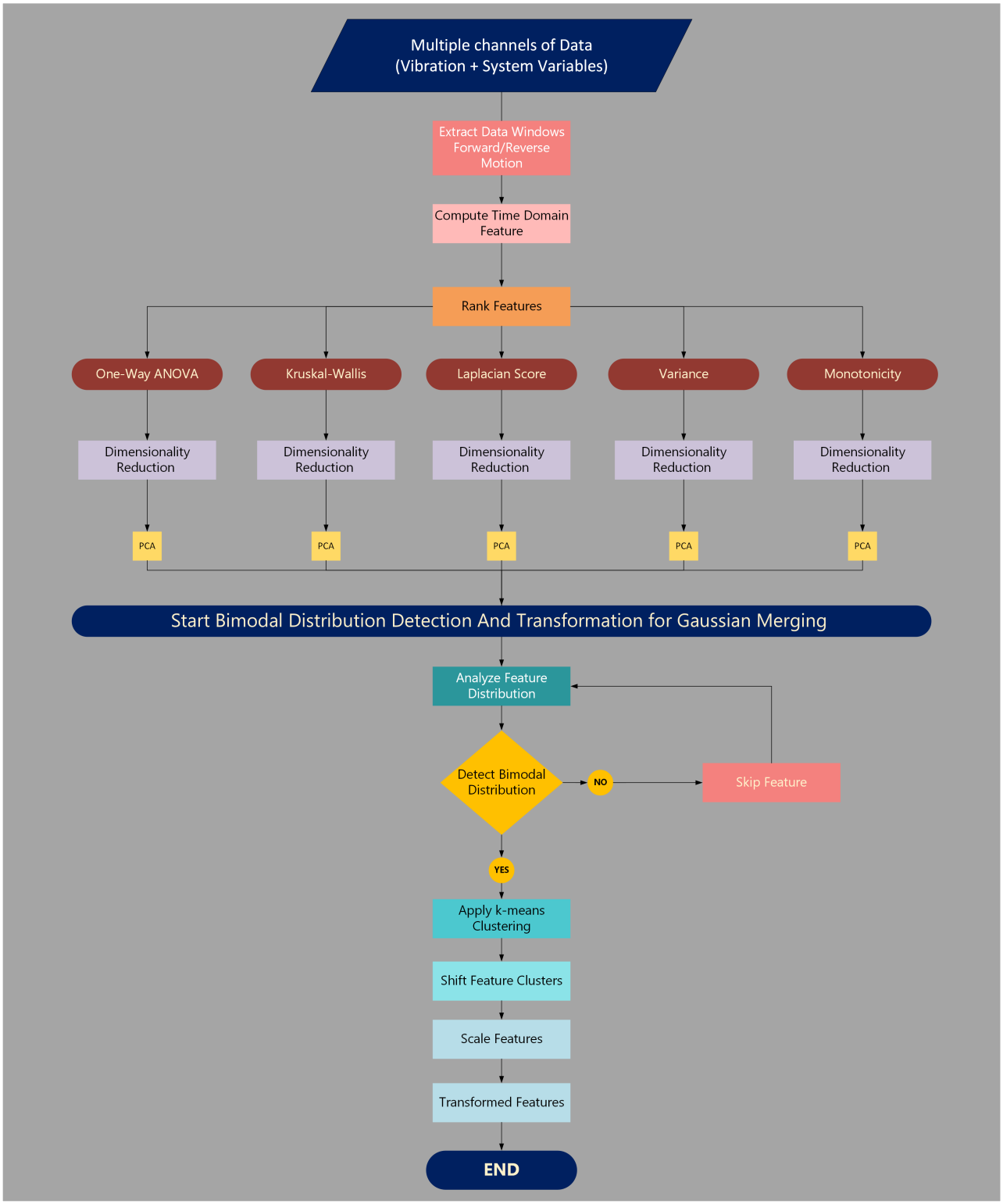


Figure 4.1: Workflow of the proposed *BDD–T* framework, illustrating the progression from raw vibration signals to the final transformed feature representation. The stages include segmentation, feature extraction, feature ranking, optional dimensionality reduction for analysis, and bimodal distribution detection and transformation.

out engaging in explicit dimensionality reduction. This configuration preserves the complete statistical structure of the feature set and is particularly advantageous in high-dimensional de-

ployment scenarios. The second variant incorporates dimensionality reduction, such as PCA or t-SNE, subsequent to feature ranking and prior to transformation. This facilitates the low-dimensional analysis, visualization, and quantitative assessment of separability both before and after reshaping. Both ADT variants shared identical preprocessing stages, including signal segmentation and time-domain feature extraction, followed by feature ranking using one-way ANOVA, Kruskal–Wallis statistics, Laplacian Score, Variance, and Monotonicity criteria. They differ solely in whether dimensionality reduction is applied before the adaptive transformation stage.

The direct feature space variant is illustrated in Figure 4.2, and the dimensionality reduction-based variant is depicted in Figure 4.3. In both instances, the core transformation mechanism analyzes marginal feature distributions via KDE and adaptively performs clustering, controlled shifting, and variance regularization to enhance class separability while preserving the intrinsic geometric relationships among samples. The complete adaptive procedure is described in Algorithm 2.

The following sections elucidate their conceptual evolution, mathematical formulation, and advantages over conventional Gaussianization techniques.

4.1 Algorithmic Evolution Overview

4.1.1 Bimodal Distribution Detection and Transformation (BDD–T)

The *Bimodal Distribution Detection and Transformation (BDD–T)* framework is designed to address features whose empirical distributions display two distinct modes. Initially, kernel density estimation (KDE) was used to detect multimodality by identifying significant local maxima in the estimated probability density function. For features exhibiting more than one mode, k-means clustering with $k = 2$ was employed to partition the feature samples into two sub-distributions. Subsequently, a controlled *shift–scale* operation is executed to consolidate the two modes into a single Gaussian-like distribution:

$$F' = \mu + \beta(F - \mu), \tag{4.1}$$

where β is a scaling factor determined by the intermodal distance and feature variance. This operation maintains the mean while smoothly aligns the clusters, thereby reducing the intra-class scatter without distorting the global feature geometry.

While BDD–T significantly enhances feature uniformity and improves the performance of unsupervised detectors such as One-Class SVM and Isolation Forest, its effectiveness is limited for features that are *unimodal but highly scattered*. Such scenarios frequently occur in non-stationary industrial environments, where load variations and stochastic fluctuations broaden the feature distribution, even in the absence of faults.

4.1.2 Adaptive Distribution Transformation (ADT)

The *Adaptive Distribution Transformation (ADT)* extends the BDD–T by incorporating a *variance-adaptive contraction* mechanism to address unimodal yet excessively dispersed features. When the variance of a feature surpasses a predefined threshold σ_t^2, ADT applies a controlled contraction toward the mean:

$$F_{j,\text{centered}} = \mu_{F_j} + \xi(F_j - \mu_{F_j}), \tag{4.2}$$

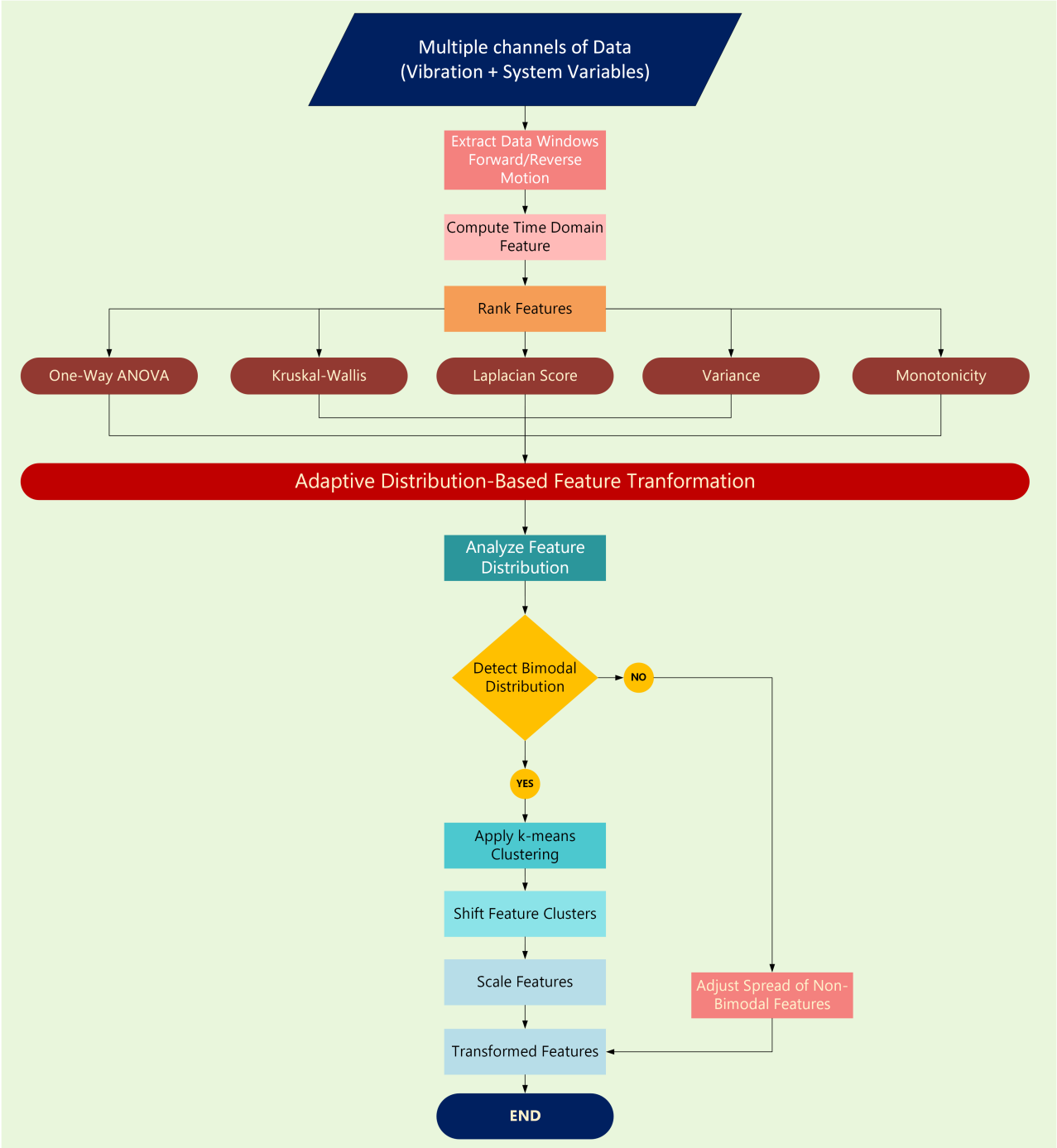


Figure 4.2: Overview of the ADT pipeline operating directly in the original ranked feature space. Multichannel vibration signals are segmented, time-domain features are extracted and ranked, and the adaptive distribution transformation is applied without intermediate dimensionality reduction.

where $\xi \in (0, 1)$ is a scaling coefficient determined by the degree of excess variance. This contraction reduces the intra-class spread while preserving the sample ordering and global topology.

Consequently, ADT integrates two complementary mechanisms.

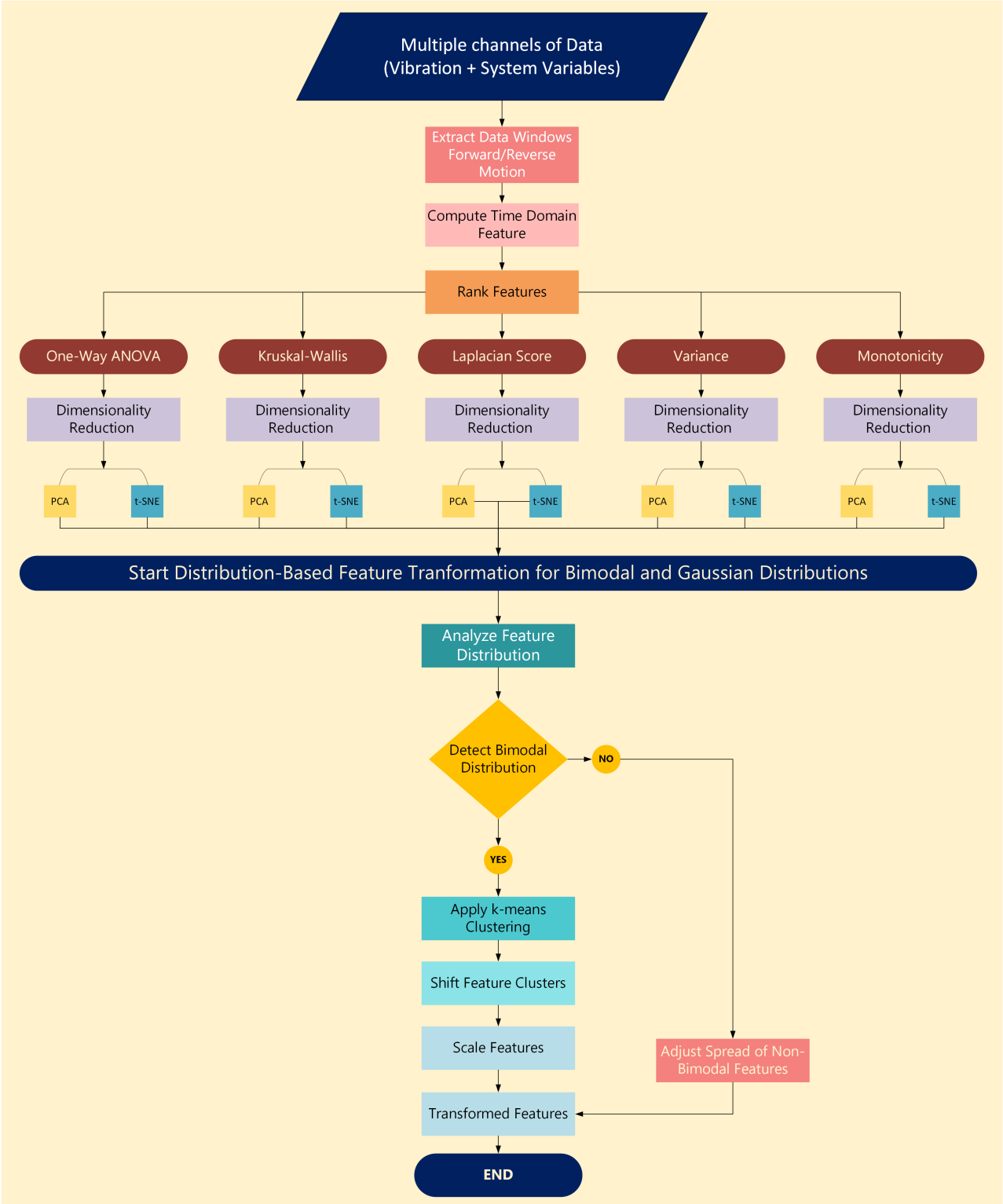


Figure 4.3: The proposed *ADT* framework is depicted in a two-part pipeline. The upper section demonstrates the process of multichannel feature extraction and ranking, which is followed by dimensionality reduction aimed at facilitating separability analysis. The lower section illustrates the adaptive distribution reshaping process, which serves to enhance both class compactness and separability.

1. *Mode merging* for multimodal features, inherited from the KDE-driven shift–scale process in BDD–T;

2. *Variance compression* for scattered unimodal features via adaptive contraction.

Together, these mechanisms provide a unified distribution regularization strategy that effectively addresses both multimodal and high-variance scenarios, enhancing cluster compactness and diagnostic reliability under various operating conditions.

4.2 Contributions and Advantages

The proposed BDD–T and ADT frameworks advance unsupervised distribution transformation through interpretable, mathematically principled approaches to Gaussianization. Their main contributions are summarized as follows.

- **Feature-wise statistical reshaping**: transforms marginal feature distributions directly without reliance on learned embeddings.
- **Bimodality-driven Gaussian merging**: detects and merges separated density modes without requiring supervision.
- **Variance-adaptive compression**: contracts high-variance unimodal features to reduce scatter and enhance compactness.

4.3 Methodology

This section introduces a distribution-aware diagnostic framework specifically designed for independent cart systems (ICS), addressing the dual challenges of high-dimensional feature spaces and overlapping class distributions. The proposed framework consists of four sequential stages: (i) feature extraction from multisensor data, (ii) statistical ranking of feature relevance, (iii) dimensionality reduction, and (iv) distribution-based transformation to enhance class separability in the projected space.

4.3.1 System Variables and Feature Extraction

The ICS platform is equipped with multiple sensor modalities, including two triaxial and three monoaxial accelerometers, as well as system-level signals such as the following error, set current, and velocity error. The raw signals were segmented using synchronized motion intervals (see Figure 1.5), and the time-domain statistical features presented in Table 3.6 were computed per segment, excluding the cart position and actual velocity to avoid redundancy.

This process results in a high-dimensional feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, where n represents the number of motion windows and d denotes the number of extracted features. To manage redundancy and enhance model tractability, the feature space was ranked and reduced using unsupervised and supervised metrics. As depicted in Figures 4.1 and 4.3, the statistical features were ranked using methodologies such as one-way analysis of variance (ANOVA), Kruskal–Wallis, Laplacian Score, Variance, and Monotonicity. Dimensionality reduction was applied using PCA and t-SNE to facilitate visual and quantitative analyses in a low-dimensional space. The core

of the framework is a distribution-based transformation strategy that adaptively modifies the feature distributions to improve class separability. The complete procedure is detailed in Algorithms 1 and 2, where both bimodal and high-variance unimodal features are addressed using KDE analysis, clustering, shifting, and scaling operations.

4.3.2 Feature Ranking and Dimensionality Reduction

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

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

The top-ranked features were retained for dimensionality reduction using Principal Component Analysis (PCA), facilitating projection into a compact subspace that captures the majority of the variance while preserving the class structure.

4.3.3 Notation and Preliminaries

This section delineates the notation, assumptions, and foundational concepts that form the basis of the proposed *Bimodal Distribution Detection and Transformation (BDD–T)* and its adaptive variant, the *Adaptive Distribution Transformation (ADT)*. As both algorithms operate on marginal feature distributions in a fully unsupervised manner, they are not dependent on class labels or geometric relationships. The notation established in this section is consistently employed throughout this chapter. Consider the dataset represented as

$$\mathbf{X} = [F_1, F_2, \ldots, F_d] \in \mathbb{R}^{n \times d}, \tag{4.3}$$

where n denotes the number of samples, d represents the number of extracted features, and $F_j = [F_{j,1}, F_{j,2}, \ldots, F_{j,n}]^\top \in \mathbb{R}^n$ denotes the jth feature column. For each feature F_j, the sample mean and standard deviation are defined as follows

$$\mu_{F_j} = \frac{1}{n} \sum_{i=1}^{n} F_{j,i}, \tag{4.4}$$

$$\sigma_{F_j} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (F_{j,i} - \mu_{F_j})^2}, \tag{4.5}$$

The overall mean after shifting, which is utilized during scaling, is denoted by μ_{overall}. Cluster assignments are indicated as

$$z_{j,i} \in \{1, 2\}, \tag{4.6}$$

which specifies the cluster index to which sample $F_{j,i}$ is assigned following k–means partitioning. Cluster centers are expressed as $\mu_{j,1}$ and $\mu_{j,2}$, and we define

$$\mu_{\text{higher}} = \max(\mu_{j,1}, \mu_{j,2}), \qquad \mu_{\text{lower}} = \min(\mu_{j,1}, \mu_{j,2}), \tag{4.7}$$

with the inter-cluster distance given by

$$d = \mu_{\text{higher}} - \mu_{\text{lower}} \tag{4.8}$$

4.3.4 Kernel Density Estimation

To delineate the distributional characteristics of each feature, a univariate kernel density estimate (KDE) is calculated as follows:

$$f_{\text{KDE}}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - F_{j,i}}{h}\right), \tag{4.9}$$

where $h > 0$ represents the bandwidth and $K(\cdot)$ denotes the Gaussian kernel, defined by

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \tag{4.10}$$

Potential modes, or local maxima, are identified at points where the following conditions are met:

$$f'_{\text{KDE}}(x_p) = 0, \qquad f''_{\text{KDE}}(x_p) < 0 \tag{4.11}$$

Let x_p denote a mode and x_{valley} denote the nearest valley between two modes. A mode is retained if its prominence surpasses a specified threshold.

$$f_{\text{KDE}}(x_p) - f_{\text{KDE}}(x_{\text{valley}}) \geq p_{\min} \tag{4.12}$$

The number of valid peaks is represented as k_j.

$$k_j = \sum_{x_p} \mathbb{I}[f_{\text{KDE}}(x_p) - f_{\text{KDE}}(x_{\text{valley}}) \geq p_{\min}] \tag{4.13}$$

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

4.3.5 Clustering Preliminaries

For bimodal features, the k–means clustering algorithm is employed with $k = 2$. The objective function to be minimized is given by:

$$J = \sum_{i=1}^{n} \sum_{k=1}^{2} \mathbb{I}[z_{j,i} = k] \, \|F_{j,i} - \mu_{j,k}\|^2 \tag{4.14}$$

The assignments and updates of the cluster centers are determined as follows.

$$z_{j,i} = \arg\min_{k\in\{1,2\}} \|F_{j,i} - \mu_{j,k}\|^2, \tag{4.15}$$

$$\mu_{j,k} = \frac{1}{|\{i : z_{j,i} = k\}|} \sum_{i:z_{j,i}=k} F_{j,i} \tag{4.16}$$

These procedures form the foundation for subsequent shift–scale adjustment.

4.3.6 Shift and Scaling Coefficients

The transformation is governed by two coefficients:

1. **Shift factor α:** This factor determines the extent to which the higher cluster is shifted towards the lower cluster:
$$\alpha = \min\left(\gamma,\ \gamma \cdot \frac{d}{\sigma_{F_j}}\right), \tag{4.17}$$
where $\gamma \in (0, 1]$ is an upper limit.

2. **Scale factor β:** This factor facilitates a seamless merging process:
$$\beta = \max\left(\delta,\ \epsilon - \eta \cdot \frac{d}{\sigma_{F_j}}\right), \tag{4.18}$$
where δ, ϵ, and η are user-defined constants.

The shifted values were calculated as follows:

$$F_{j,\text{shifted}} = F_{j,\text{higher}} - \alpha d, \tag{4.19}$$

The final scaling is executed as follows:

$$[F_{j,\text{scaled}} = \mu_{\text{overall}} + \beta(F_j - \mu_{\text{overall}}) \tag{4.20}$$

4.3.7 Variance Thresholding for ADT

The ADT algorithm further addresses the problem of unimodal yet excessively dispersed features. Let σ_t represent the variance threshold. If $k_j \leq 1$ (indicating no bimodality) and $\sigma_{F_j} > \sigma_t$, a controlled contraction is implemented as follows:

$$F_{j,\text{centered}} = \mu_{F_j} + \xi(F_j - \mu_{F_j}) \tag{4.21}$$

where $\xi \in (0, 1)$ is a contraction factor chosen to diminish the spread without altering its order. This procedure maintains global topology while reducing intra-feature variance.

4.4 Experimental Results and Discussion

This section provides a detailed explanation of the rationale behind the selection of hyperparameters and examines their transferability and stability. It also includes a sensitivity analysis, categorization of additive noise, and evaluation of transformation results with and without the presence of additive noise.

Algorithm 1 Bimodal Distribution Detection and Transformation (BDD–T)

Require: Feature set $\mathbf{X} \in \mathbb{R}^{n \times d}$, parameters $c_{\min}, \gamma, \delta, \epsilon, \eta$
Ensure: Transformed feature set $\widetilde{X}$

$\widetilde{X} \leftarrow \emptyset$
for each feature F_j in $\mathbf{X}$ **do**
    Compute KDE
    $f_{\text{KDE}}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - F_{j,i}}{h}\right)$
    $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$
    Identify peaks x_p such that $f'_{\text{KDE}}(x_p) = 0$ and $f''_{\text{KDE}}(x_p) < 0$
    Retain peaks satisfying: $f_{\text{KDE}}(x_p) - f_{\text{KDE}}(x_{\text{valley}}) \geq p_{\min}$
    $k_j \leftarrow$ number of valid peaks
    if $k_j > 1$ **then** ▷ Bimodal case
        Apply k-means with $k = 2$ clusters
        $z_{j,i} \leftarrow \arg\min_{k \in \{1,2\}} \|F_{j,i} - \mu_{j,k}\|^2$
        $\mu_{j,k} \leftarrow \frac{1}{|\{i : z_{j,i} = k\}|} \sum_{i : z_{j,i} = k} F_{j,i}$
        $\mu_{\text{higher}} \leftarrow \max(\mu_{C1}, \mu_{C2})$
        $\mu_{\text{lower}} \leftarrow \min(\mu_{C1}, \mu_{C2})$
        $d \leftarrow \mu_{\text{higher}} - \mu_{\text{lower}}$
        $\alpha \leftarrow \min\left(\gamma,\ \frac{d}{\sigma_{F_j}} \gamma\right)$
        $F_{j,\text{shifted}} \leftarrow F_{j,\text{higher}} - \alpha\, d$
        $\beta \leftarrow \max\left(\delta,\ \epsilon - \eta \frac{d}{\sigma_{F_j}}\right)$
        $\mu_{\text{overall}} \leftarrow \text{mean}(F_j)$
        $F_{j,\text{scaled}} \leftarrow \mu_{\text{overall}} + \beta\,(F_j - \mu_{\text{overall}})$
        Append $F_{j,\text{scaled}}$ to $\mathbf{Y}_{\text{transformed}}$
    else ▷ Not bimodal
        Append F_j to $\widetilde{X}$
    end if
end for
    return $\widetilde{X}$

4.4.1 Rationale for Hyperparameter Selection

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

Bimodality gate and peak prominence threshold The detection of bimodality is primarily governed by the minimum peak prominence parameter, $c_{\min}$, which is defined in relation to the maximum value of the estimated marginal density, as follows: By design, $c_{\min} \in (0, 1)$ specifies the minimum prominence necessary for a density peak to be deemed structurally significant. Interpreting $c_{\min}$ as a percentage of the maximum density ensures scale invariance across different features and datasets. The values of $c_{\min}$ ranging from approximately 10% to 20% of the maximum density are adequate for reliably identifying genuine bimodal structures while

Algorithm 2 Adaptive Distribution Transformation (ADT)

Require: $\mathbf{X} \in \mathbb{R}^{n \times d}$, $c_{\min}$, γ, δ, ϵ, η, σ_t, ξ
Ensure: $\widetilde{X}$

$\widetilde{X} \leftarrow \emptyset$
for each feature F_j in $\mathbf{X}$ **do**
  Compute KDE
  $f_{\text{KDE}}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - F_{j,i}}{h}\right)$
  $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$
  Identify peaks x_p such that $f'_{\text{KDE}}(x_p) = 0$ and $f''_{\text{KDE}}(x_p) < 0$
  Retain peaks where $f_{\text{KDE}}(x_p) - f_{\text{KDE}}(x_{\text{valley}}) \geq p_{\min}$
  $k_j \leftarrow$ number of valid peaks
  if $k_j > 1$ **then** ▷ Bimodal / multi-peak case
    Assign clusters via k-means with $k = 2$
    $z_{j,i} \leftarrow \arg\min_{k \in \{1,2\}} \|F_{j,i} - \mu_{j,k}\|^2$
    $\mu_{j,k} \leftarrow \frac{1}{|\{i : z_{j,i} = k\}|} \sum_{i : z_{j,i} = k} F_{j,i}$
    $\mu_{\text{higher}} \leftarrow \max(\mu_1, \mu_2)$
    $\mu_{\text{lower}} \leftarrow \min(\mu_1, \mu_2)$
    $d \leftarrow \mu_{\text{higher}} - \mu_{\text{lower}}$
    $\alpha \leftarrow \min\left(\gamma, \frac{d}{\sigma_{F_j}} \gamma\right)$
    $F_{j,\text{shifted}} \leftarrow F_{j,\text{higher}} - \alpha\, d$
    $\beta \leftarrow \max\left(\delta, \epsilon - \eta \frac{d}{\sigma_{F_j}}\right)$
    $\mu_{\text{overall}} \leftarrow \text{mean}(F_j)$
    $F_{j,\text{scaled}} \leftarrow \mu_{\text{overall}} + \beta\,(F_j - \mu_{\text{overall}})$
    Store $F_{j,\text{scaled}}$ in $\widetilde{X}$
  else if $\sigma_{F_j} > \sigma_t$ **then** ▷ High-variance unimodal case
    $\mu_{F_j} \leftarrow \text{mean}(F_j)$
    $F_{j,\text{centered}} \leftarrow \mu_{F_j} + \xi\,(F_j - \mu_{F_j})$
    Store $F_{j,\text{centered}}$ in $\widetilde{X}$
  else ▷ Already compact / low-variance case
    Store original F_j in $\widetilde{X}$
  end if
end for
  return $\widetilde{X}$

mitigating minor fluctuations attributable to sampling noise or finite sample effects. Selecting excessively low values of $c_{\min}$ increases the sensitivity to spurious local maxima, resulting in false bimodality detection and unwarranted cluster formation.

Upon detection of bimodality, the transformation parameters within the framework are activated conditionally. The shift coefficient γ regulates the maximum fraction of inter-cluster separation that can be applied during the cluster realignment. To prevent excessive displacement, the effective shift is adaptively constrained as $\min\left(\gamma, \gamma \frac{\Delta\mu}{\sigma}\right)$, where $\Delta\mu$ represents the distance between cluster centers and σ the feature standard deviation. This normalization ensures that the intensity of shifting is governed by relative rather than absolute separation, rendering γ applicable across the datasets. The subsequent scaling is controlled by parameters

δ, ϵ, and η, which collectively define a bounded contraction range. The scale factor is calculated as $\max\big(\delta,\ \epsilon - \eta\frac{\Delta\mu}{\sigma}\big)$, ensuring that contraction strength decreases smoothly as cluster separation increases, while remaining strictly lower-bounded by δ to prevent distribution collapse. Typical values of $\delta \in [0.01, 0.2]$, $\epsilon \in [0.2, 0.4]$, and $\eta \in [0.3, 0.6]$ have been found to provide stable reshaping across the feature types. Notably, these parameters regulate the extent of reshaping once bimodality is confirmed but do not influence the detection decision.

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

4.4.2 Transferability and Stability of Hyperparameter

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

or feature-normalization schemes. These adjustments correspond to coarse calibration rather than dataset-specific retuning and do not alter the underlying transformation. Consequently, the selected hyperparameter values serve as conservative defaults that regulate the distributional behavior rather than optimize the dataset-specific performance. Observations across multiple datasets indicate that moderate perturbations around the reported values do not lead to qualitative changes in the transformed feature space, thereby supporting the robustness and transferability of the proposed framework.

Table 4.1: Hyperparameters for Adaptive Distribution-Based Transformation. Values remain fixed across all experiments.

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

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

4.4.3 Hyperparameter Sensitivity Analysis

A one-at-a-time sensitivity analysis was conducted to assess the sensitivity of the hyperparameter values for the ADT. Each transformation parameter was varied over a broad interval (0.01–0.9), while all other parameters were maintained at their nominal values, and the entire detection pipeline was re-evaluated using the F1-score as the performance metric. The analysis was performed on two distinct feature families, classical statistical descriptors (Table 3.6) and Hjorth parameters (Table 4.2), to determine whether the parameter sensitivity was consistent across heterogeneous feature representations. In the reported sensitivity plots, the horizontal axis represents the varied parameter, the vertical axis indicates the mean F1-score, and each curve corresponds to a different baseline regime defined by either the unimodal dispersion threshold σ_t or the contraction factor ξ.

Across both feature families, most parameters exhibited wide stability regions (Figures 4.4 and 4.5), where the performance varied only marginally. Notably, the parameters associated with the bimodality gate ($c_{\min}$, γ, δ, ϵ, and η) exhibit relatively flat response curves over broad intervals, with performance degradation occurring only at extreme values. This behavior suggests that these parameters are not sharply tuned and that moderate deviations from the nominal values do not result in qualitative changes in detection performance.

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

Overall, the sensitivity analysis confirmed that the hyperparameter values adopted in Table 4.1 were well within the stable operating regions for both feature families. The observed trends

demonstrate that the proposed transformation is robust to moderate parameter perturbations and that performance degradation occurs only under extreme, practically irrelevant conditions. These results provide empirical support for fixing the hyperparameters in all the experiments.

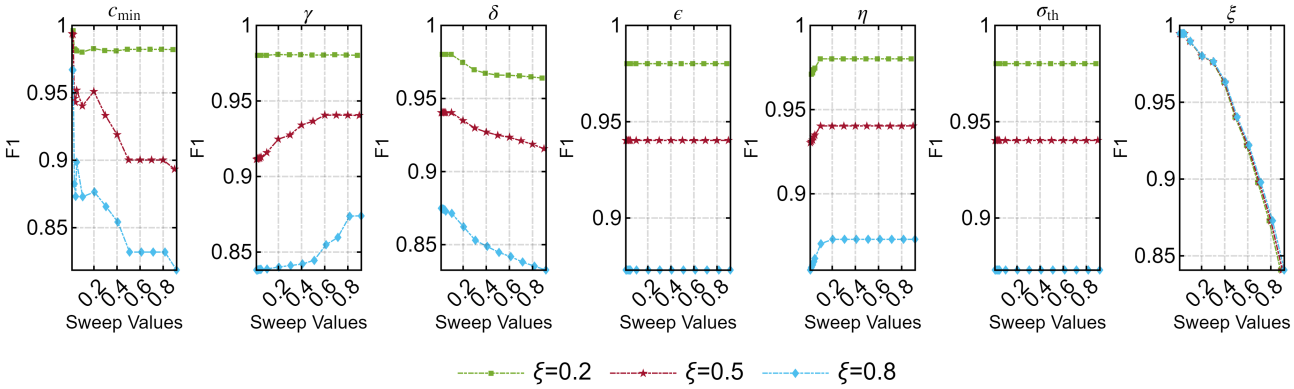


Figure 4.4: Sensitivity analysis of adaptive transformation hyperparameters on **classical statistical descriptors** is conducted. Each panel examines the variation of a single parameter while maintaining the others at their nominal value. The curves represent different baseline regimes of the unimodal-compression gate, with performance evaluated using the mean F1-score.

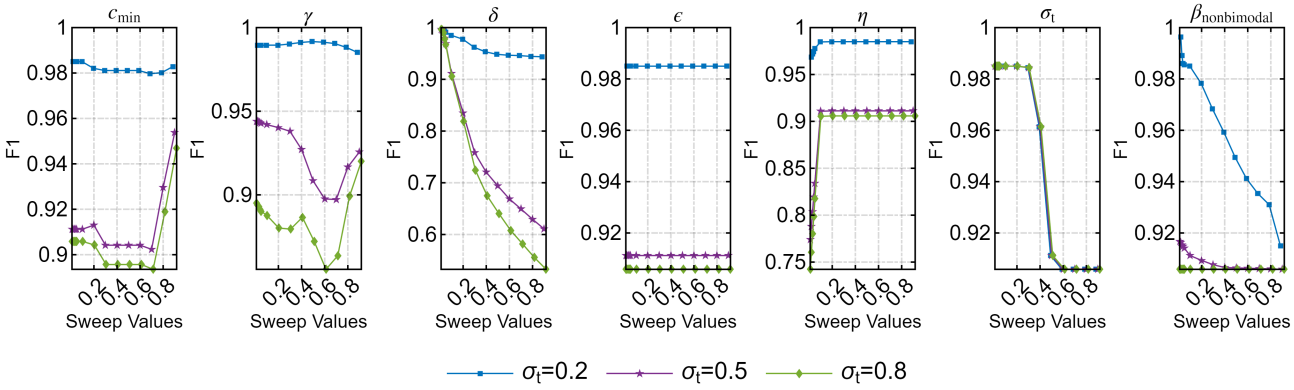


Figure 4.5: The sensitivity analysis of adaptive transformation hyperparameters on **Hjorth parameters** is presented. Each panel examines the variation of a single parameter while maintaining the constancy of the other parameters. The findings indicate extensive stability regions and affirm the function of gating parameters as activation thresholds, rather than as variables requiring dataset-specific tuning.

4.4.4 Additive Noise Categories

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

Table 4.2: Hjorth prameters. Closed-form expressions for activity, mobility, and complexity computed per analysis window.

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

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

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

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{n}_{\mathrm{AWGN}} + \mathbf{n}_{\mathrm{UNI}}, \quad \mathbf{n}_{\mathrm{AWGN}} \sim \mathcal{N}(0,\sigma^2), \quad \mathbf{n}_{\mathrm{UNI}} \sim \mathcal{U}(-a,a) \tag{4.22}$$

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

Narrowband Intermittent Interference To simulate periodic tonal interference resulting from mechanical resonances or rotating machinery, we utilized band-limited pulses generated via amplitude-modulated noise, which was shaped using a Butterworth bandpass filter:

$$n_{\mathrm{NB}}(t) = A\, e^{\kappa t}\left(w(t) * h_{[f_L,f_H]}(t)\right) \tag{4.23}$$

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

Synthetic Structural Noise via Pseudo-FRF Shaping To simulate the broadband excitation shaped by mechanical resonance profiles, a pseudo-frequency response function (PFRF) was generated, consisting of sinusoidal peaks with randomized phases:

$$|H[r]| \propto \mathrm{PFRF}[r] \cdot (\nu_1[r]\,\nu_2[r] + \zeta), \quad \angle H[r] \sim \mathcal{U}[0,2\pi) \tag{4.24}$$

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

$$\mathrm{SN}(t) = \Re\{\mathrm{IFFT}(H)\} \tag{4.25}$$

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

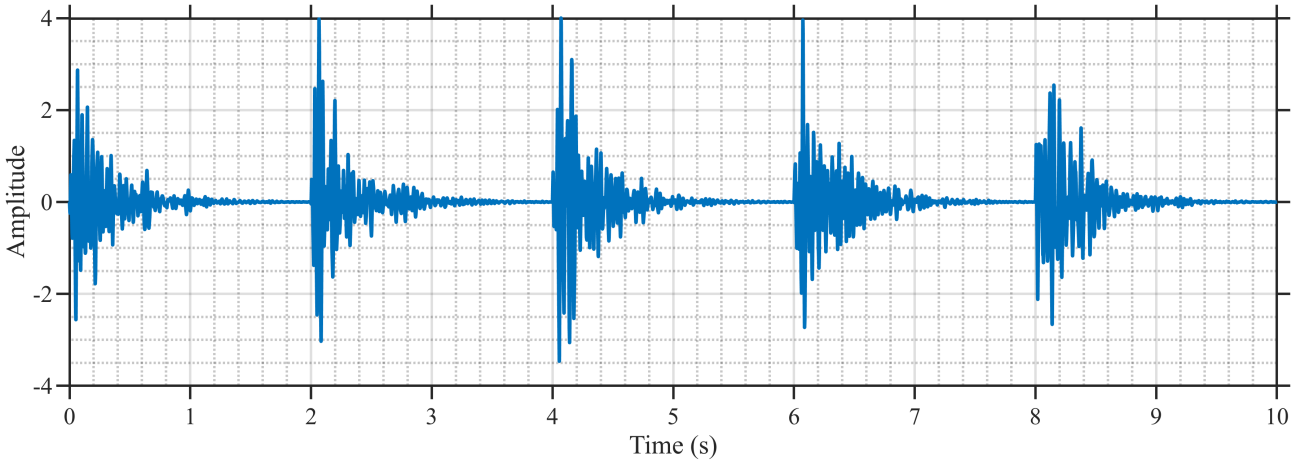


Figure 4.6: An illustration of the amplitude-modulated narrowband interference employed in Category 2 is presented (f_s=50 kHz, band-pass $[20, 40]$ Hz, pulse repetition f_p=0.5 Hz, exponential decay δ). The decaying bursts simulate intermittent tonal contamination that aligns with the mechanical cycle.

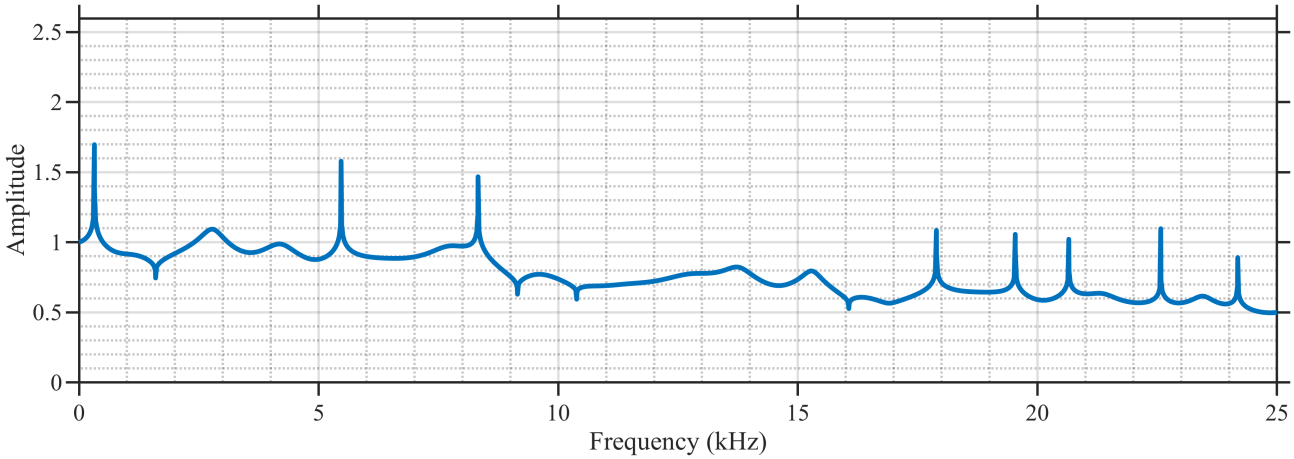


Figure 4.7: The Pseudo-FRF profile (Category 3) is generated with parameters N_p=20, peak_{RES}=2, $\text{peak}_{\text{ARES}}$=0.75, and ADR=0.5. This profile establishes the target spectral envelope for the structural noise synthesized via the Inverse Fast Fourier Transform (IFFT).

4.4.5 Results without Additive Noise

4.4.5.1 Case A — Results using BDD–T

To assess the impact of the proposed transformation on class separability, the feature distributions within the two-dimensional principal component (PC) space were initially examined, followed by an evaluation of the multi-class classification performance using k-nearest neighbor (KNN) and support vector machines (SVM). Five ranked feature sets—OWA, KW, LS, Var, and Mono—were considered using the first two PCs of both the original and transformed data for the visual and quantitative analyses.

Figure 4.8 presents 2-D scatter plots of multiple classes before (left column) and after (right column) the proposed transformation. In the original PC space, the class clusters were elongated and exhibited significant overlap, particularly for FIR025, FIR05, and FIR10. Post-transformation, the clusters became notably more compact and well-separated, with minimal

inter-class overlap. This visual evidence suggests that the transformation effectively homogenizes the feature distributions, enhancing intra-class cohesion and inter-class separation. Such spatial restructuring of the data is anticipated to improve the classification accuracy of both parametric and non-parametric models.

To quantify these improvements, Figures 4.9 and 4.10 present the normalized confusion matrices for the KNN and SVM classifiers, respectively. In each panel, the left column corresponds to the classification in the original PC space, and the right column shows the results after transformation. In both classifiers, a pronounced increase in diagonal dominance was observed post-transformation, indicating reduced misclassification and enhanced discriminative capabilities. This improvement is particularly notable for the LS, Var, and Mono feature sets, where the post-transformation classification approaches perfect accuracy. KNN, in particular, demonstrated a marked reduction in the off-diagonal elements across almost all feature sets.

Collectively, the qualitative scatter plot analysis and quantitative classification results confirm that the proposed transformation substantially enhanced class separability. The method yields a more structured feature space that is highly favorable for both visual interpretation and automated classification of the data.

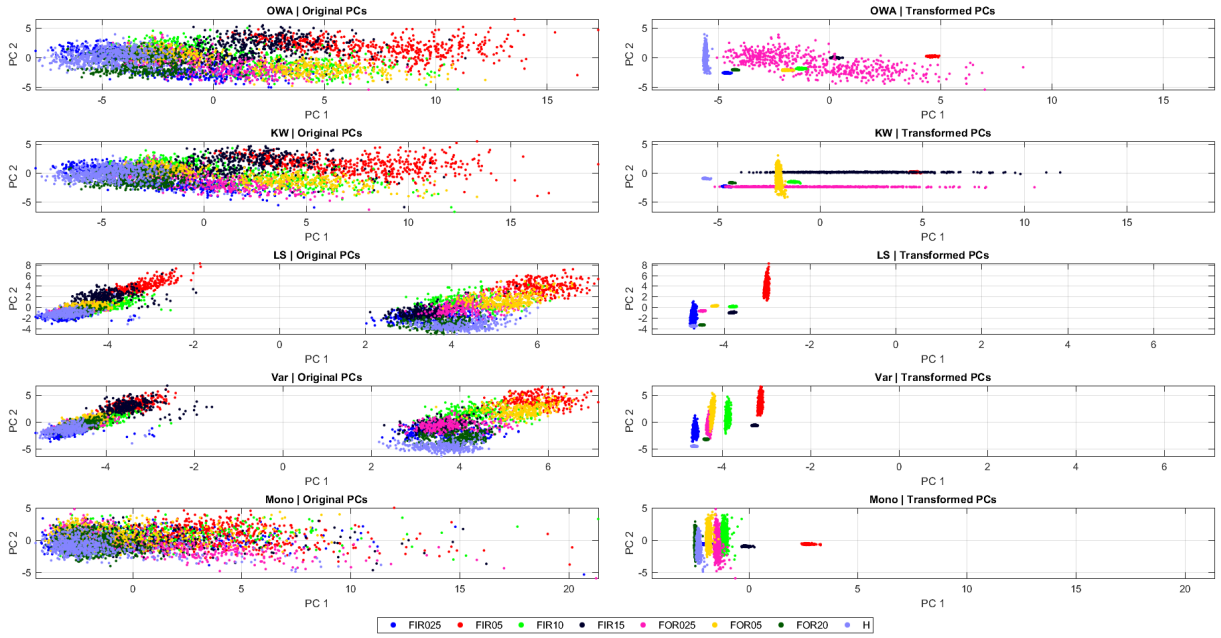


Figure 4.8: Scatter plots of the first two principal components (PC1 and PC2) are presented both prior to (left column) and following (right column) transformation for all five ranked feature sets (OWA, KW, LS, Var, Mono). Each color denotes a distinct fault or healthy state of the bearing. These plots demonstrate the impact of the transformation on inter-class separation and intra-class compactness.

4.4.5.2 Case B — Results using ADT

This section presents the diagnostic results obtained from both the single-cart and three-cart experimental setups, as documented in the bearing dataset (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a). In the single-cart configuration, a cart moved bidirectionally along a curved motor module at a nominal speed of 1000 mm/s (Figure 4.11). This reciprocal motion enables a distinct analysis of the vibration characteristics during forward and reverse traversal. Signal

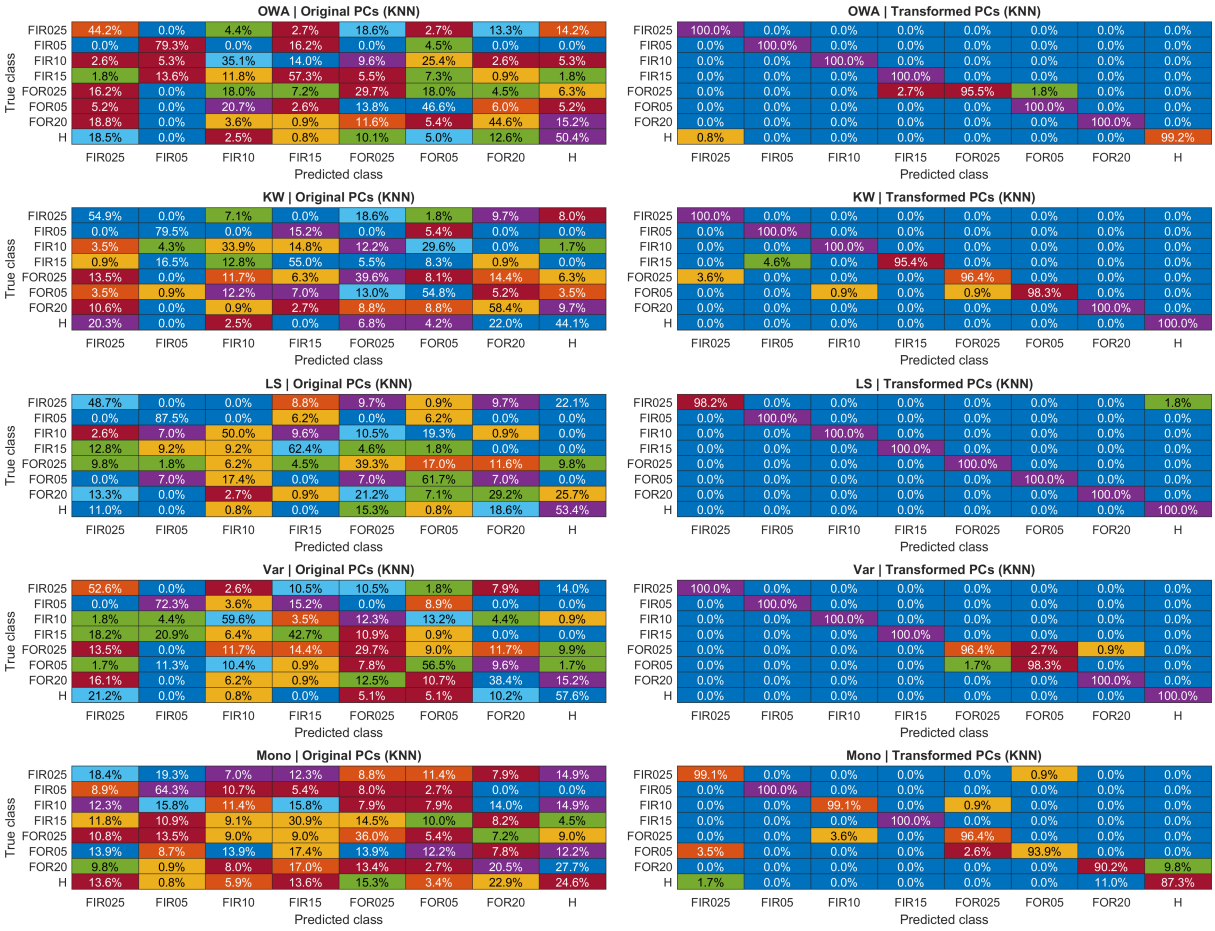


Figure 4.9: The confusion matrices derived from the K-Nearest Neighbors (KNN) algorithm are presented for both pre- and post-transformation scenarios across five ranked feature sets: OWA, KW, LS, Var, and Mono. The left column illustrates the original principal components (PCs), and the right column depicts the transformed PCs. The values represent row-normalized classification accuracies (in %), enabling a direct comparison of class-level prediction performance before and after the transformation.

windows were independently extracted for each direction, and the statistical features were computed for each windowed segment of the data. The corresponding speed profile is illustrated in Figure 1.5, highlighting the acceleration, deceleration, and steady-state phases during both the forward and reverse movements of the extraction windows.

In the single-cart experiments, the dataset comprised 14 channels, including nine vibration channels and five system variables relevant to cart motion: position, following error, actual velocity, velocity error, and set current. Vibration data were collected using two tri-axial and three mono-axial accelerometers strategically positioned along the track to capture the dynamic interactions between the cart bearings and guide rail (Figure 4.11). Statistical features (enumerated in Table 3.6) were extracted from all the sensor channels, except for the cart position and speed, resulting in high-dimensional feature space. To address feature redundancy and enhance class separability, five feature ranking techniques were employed, followed by Principal Component Analysis (PCA) for dimensionality reduction and t-distributed stochastic neighbor embedding (t-SNE).

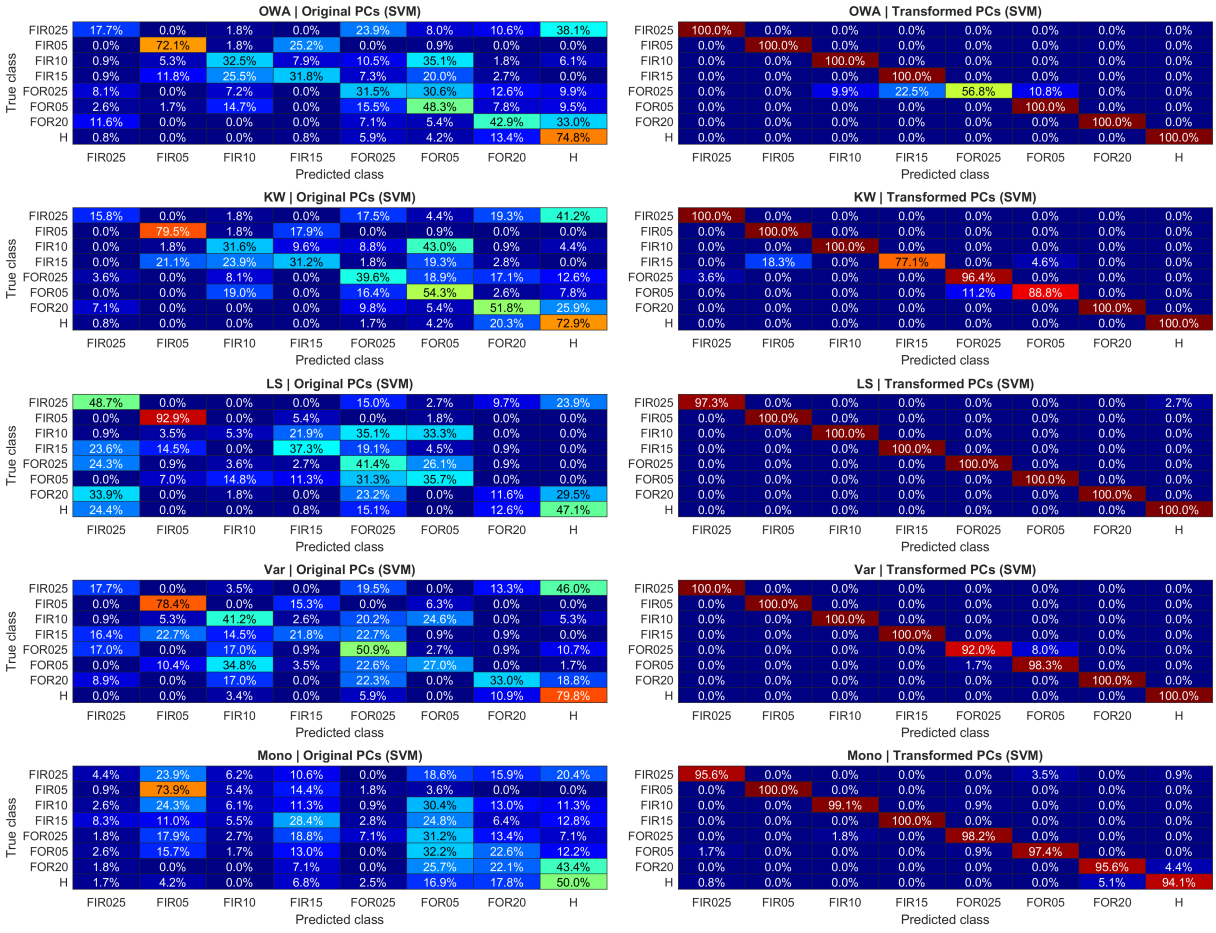


Figure 4.10: The confusion matrices derived from Support Vector Machine (SVM) analysis are presented for both pre- and post-transformation scenarios across five ranked feature sets: OWA, KW, LS, Var, and Mono. The left column displays the original principal components (PCs), and the right column presents the transformed PCs. The values represent row-normalized classification accuracies (in %), thereby illustrating the performance variations resulting from the transformation.

The results of the three-cart experiments are also presented in this section. In this configuration, three carts spaced 250 mm apart continuously moved around the full track in a closed-loop arrangement. The analysis focused on data windows extracted when the middle cart containing a faulty bearing traversed the right-side curved module between absolute positions of 250 mm and 750 mm, as shown in the shaded region of Figure 4.12. While the vibration channels remained unchanged (nine channels), the system variables included cart-specific measurements for each of the three carts, resulting in 15 system variable channels. This enriched context increases the complexity of the dataset, reflecting the diagnostic challenges associated with multi-cart industrial deployment. A comprehensive description of the experimental setup, configurations, and fault injection methodology is provided in (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a).

In this study, eight health conditions were analyzed, as detailed in Table 3.7. These conditions included one healthy class (H), four types of inner race faults (*FIR025*, *FIR05*, *FIR10*, and *FIR15*), and three types of outer race faults (*FOR025*, *FOR05*, and *FOR20*). Within this labeling framework, the prefix `F` indicates a faulty condition, whereas `IR` and `OR` specify the

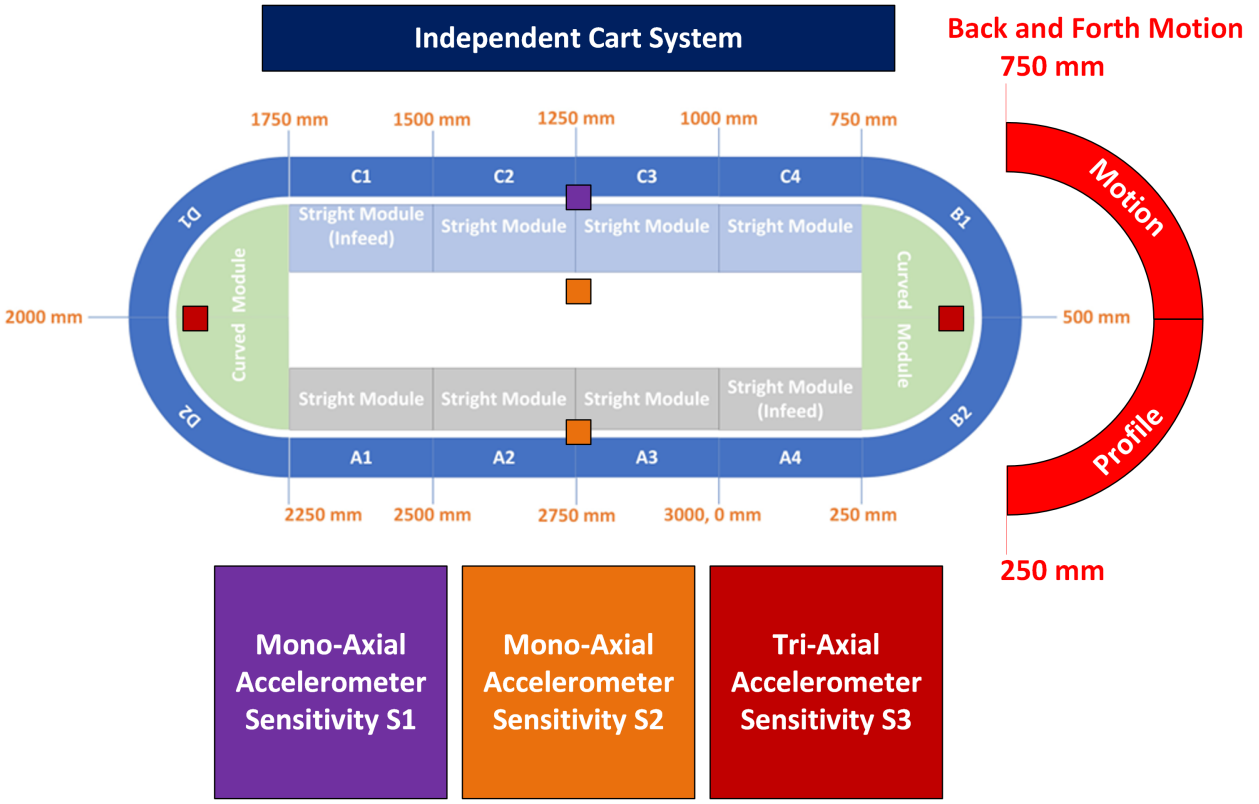


Figure 4.11: The topological schematic of the experimental Independent Cart System (ICS) is presented, featuring annotated dimensions, sensor locations, and the motion profile. The oval-shaped track comprises straight (C1–C4 and A1–A4) and curved (D1–D2 and B1–B2) modules that facilitate a closed loop. Colored squares indicate accelerometer placements: S1 and S2 represent the mono-axial sensor sensitivities on the straight sections, whereas S3 denotes the tri-axial sensor sensitivity on the curved module. The dimension labels specify the distances between the key module boundaries (mm). The red arc on the right delineates the 750 mm back-and-forth reversal zone employed during the testing to induce bidirectional motion. This configuration enables the analysis of sensor response under varying kinematics and track geometries.

defect location on the inner or outer race, respectively. The numerical suffix represents the notch width in millimeters (e.g., `025` = 0.25 mm, `05` = 0.5 mm). Owing to the constraints associated with the deep-groove bearing geometry, the fault depth was not recorded, and the severity was assessed solely based on the notch width.

Effect of the Transformation in PCA Space Figures 4.13 and 4.14 illustrate the class distributions prior to (top rows) and after (bottom rows) the proposed transformation across the five ranking strategies (OWA, KW, LS, Var and Mono). In the *pre-transform* maps, the clusters are elongated and exhibit significant overlap, particularly among the healthy (H) and small fault classes (FIR025 and FOR025). After transformation, the clusters became compact and well separated, indicating density regularization and enhanced inter-class margins. This trend persisted when transitioning from the single-cart to the more complex three-cart setup, demonstrating robustness against increased coupling effects.

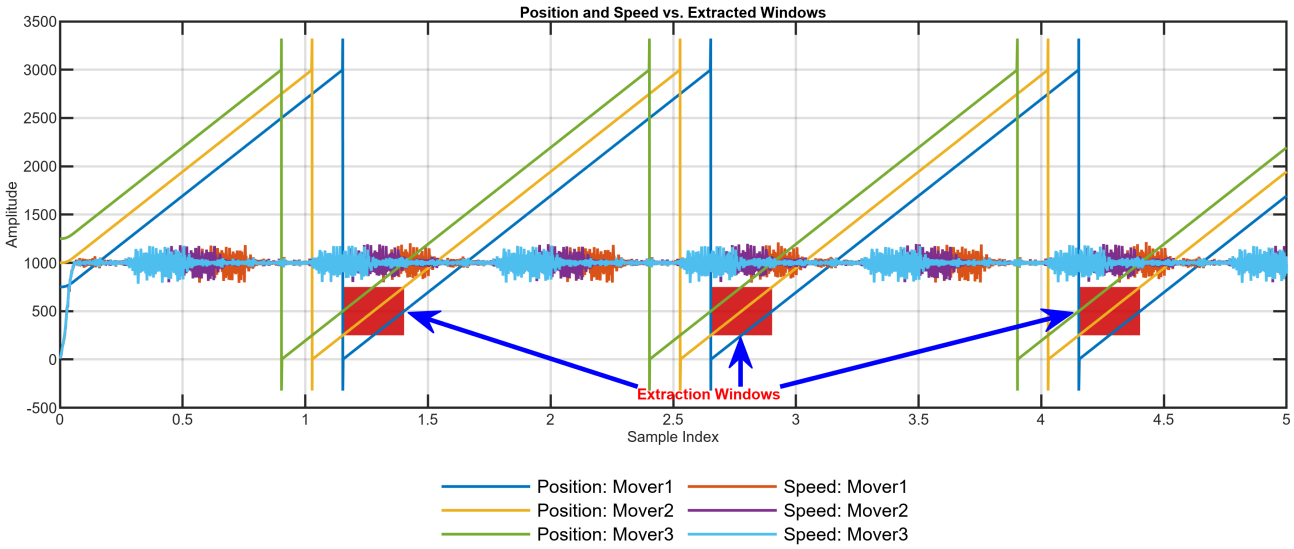


Figure 4.12: The trajectories of position and velocity for three carts moving continuously along the XTS track are analyzed. Extraction windows, marked in red, were designated for the middle cart when its absolute position ranged from 250–750 mm. The carts are evenly spaced at intervals of 250 mm and operate in a cyclic motion pattern, enabling consistent data segmentation across repetitive cycles.

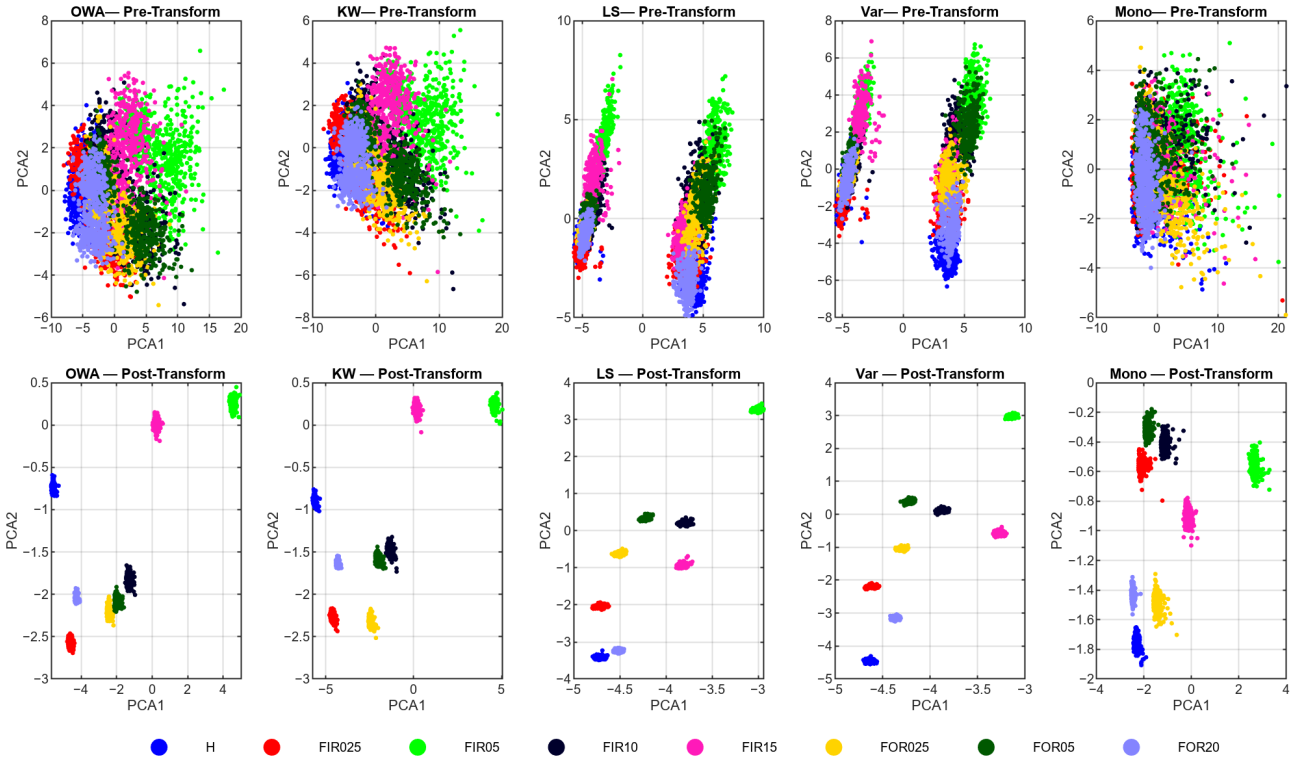


Figure 4.13: PCA embeddings (**single-cart**): The pre-transformation results (top row) display overlapping, elongated clusters, whereas the post-transformation results (bottom row) exhibit compact and distinct class clusters across all five ranking strategies.

Effect of the Transformation in t–SNE Space Figures 4.15 and 4.16 illustrate the t–SNE embeddings generated using four distance metrics: Mahalanobis, Cosine, Chebychev, and Euclidean. Prior to transformation, the feature points exhibited scattered and overlapping clusters

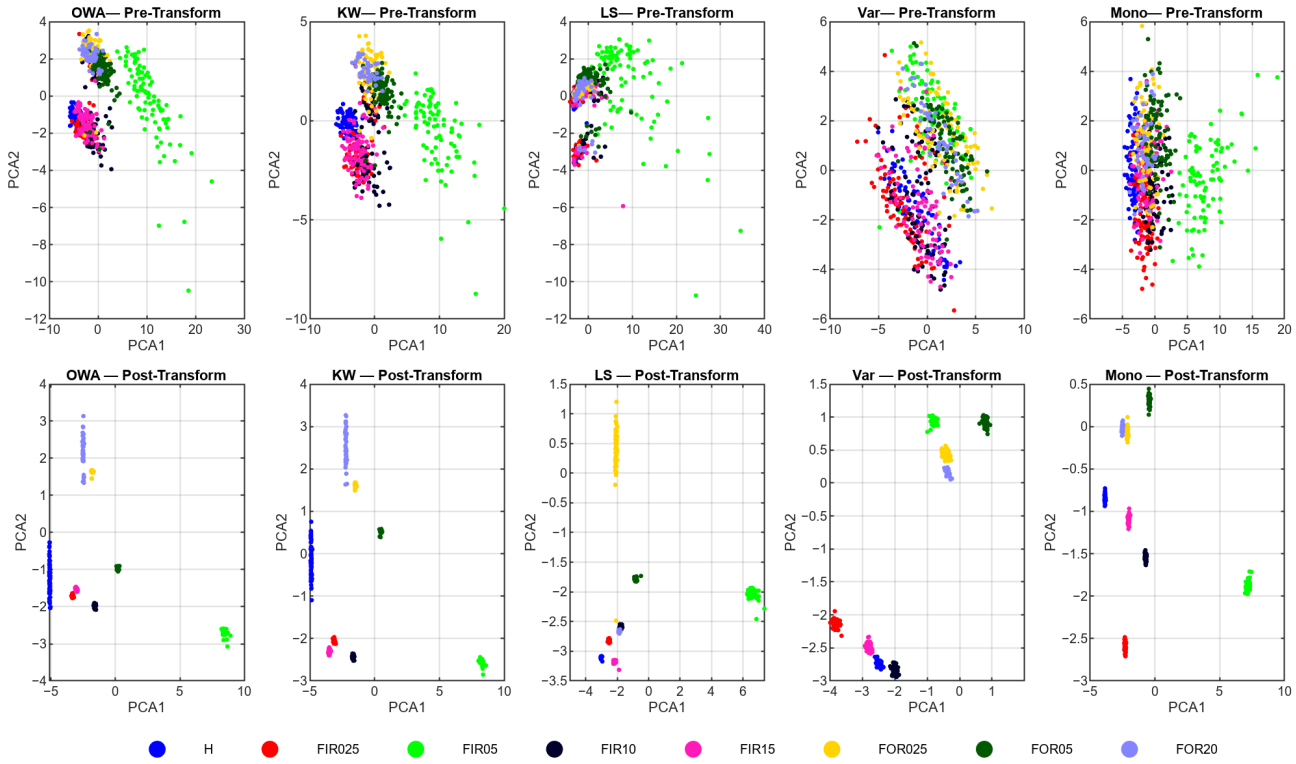


Figure 4.14: The proposed transformation of PCA embeddings (**three-cart**) results in compact, non-overlapping clusters, even with increased system complexity, thereby confirming its robustness.

with uneven densities. Small-fault cases are challenging to distinguish because their samples are intermingled with other classes. Following the application of the proposed transformation, each class forms a compact, distinctly separated cluster. The spacing between the classes becomes larger and more uniform across all metrics. This balanced distribution of points facilitates the reliable identification of faults by anomaly detection models, such as the OCSVM and Isolation Forest.

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

4.4.5.3 Comparative Analysis

Figure 4.17 presents a comparative analysis of various transformation methodologies for feature embedding within the single cart dataset. Traditional embedding techniques, such as PCA and t–SNE, display elongated overlapping clusters, where minor faults (FIR025 and FOR025) are interspersed with healthy samples. This pattern suggests uneven data density and inadequate discrimination between closely related operational states. Marginal transformations, including Box–Cox and Yeo–Johnson, address skewness at the feature level but do not significantly enhance global separability. The orientation and shape of the clusters remained largely unchanged, and inter-class mixing persisted. Gaussianization-based methods, such as hyperbolic power transform (HPT) and RBIG, transform data into Gaussian marginals. However, HPT excessively compresses one dimension, whereas RBIG results in an almost isotropic scatter that obscures underlying fault structures. Although these methods achieve statistical regularity,

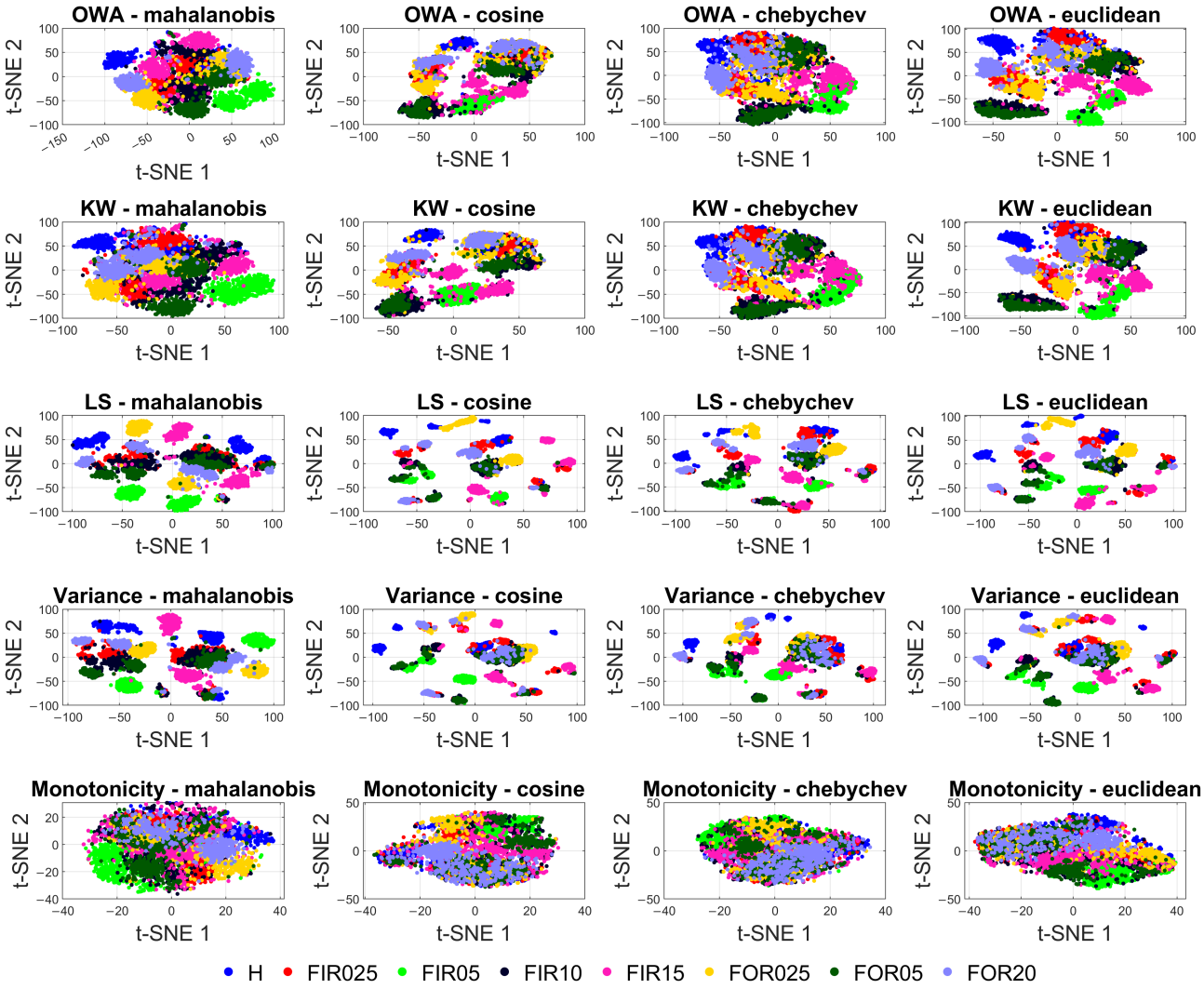


Figure 4.15: t–SNE (**single-cart**, pre-transform) exhibits scattered and overlapping clusters, with a noticeable imbalance in density across various metrics, including Mahalanobis, Cosine, Chebychev, and Euclidean.

they fail to preserve the class geometry that is essential for diagnostic interpretation. The density-homogenizing transform CDF–TS provides more uniform local densities and partially improves separation; however, overlaps remain, and the progression of fault severity is not clearly preserved. Figure 4.18 compares the two multimodality-aware transformations, DipExt and ICDMDip, projected into the PCA domain using the proposed method. DipExt→PCA produces vertically elongated structures with limited distinction between fault levels, whereas ICDMDip→PCA generates scattered, partially overlapping clouds that obscure class boundaries. In contrast, the proposed transformation yields distinct compact clusters aligned in an interpretable progression from healthy to severe-fault states. It effectively regularizes the density while preserving the geometric relationships among classes, thereby enabling improved cluster compactness and stability under varying conditions.

The proposed transformation results in compact, well-separated clusters with consistent inter-class spacing and monotonic ordering from mild to severe fault conditions. This balanced reshaping of the feature space reduces local density variations, enhances separability, and supports downstream anomaly detection and classification by providing clear and uniform decision-

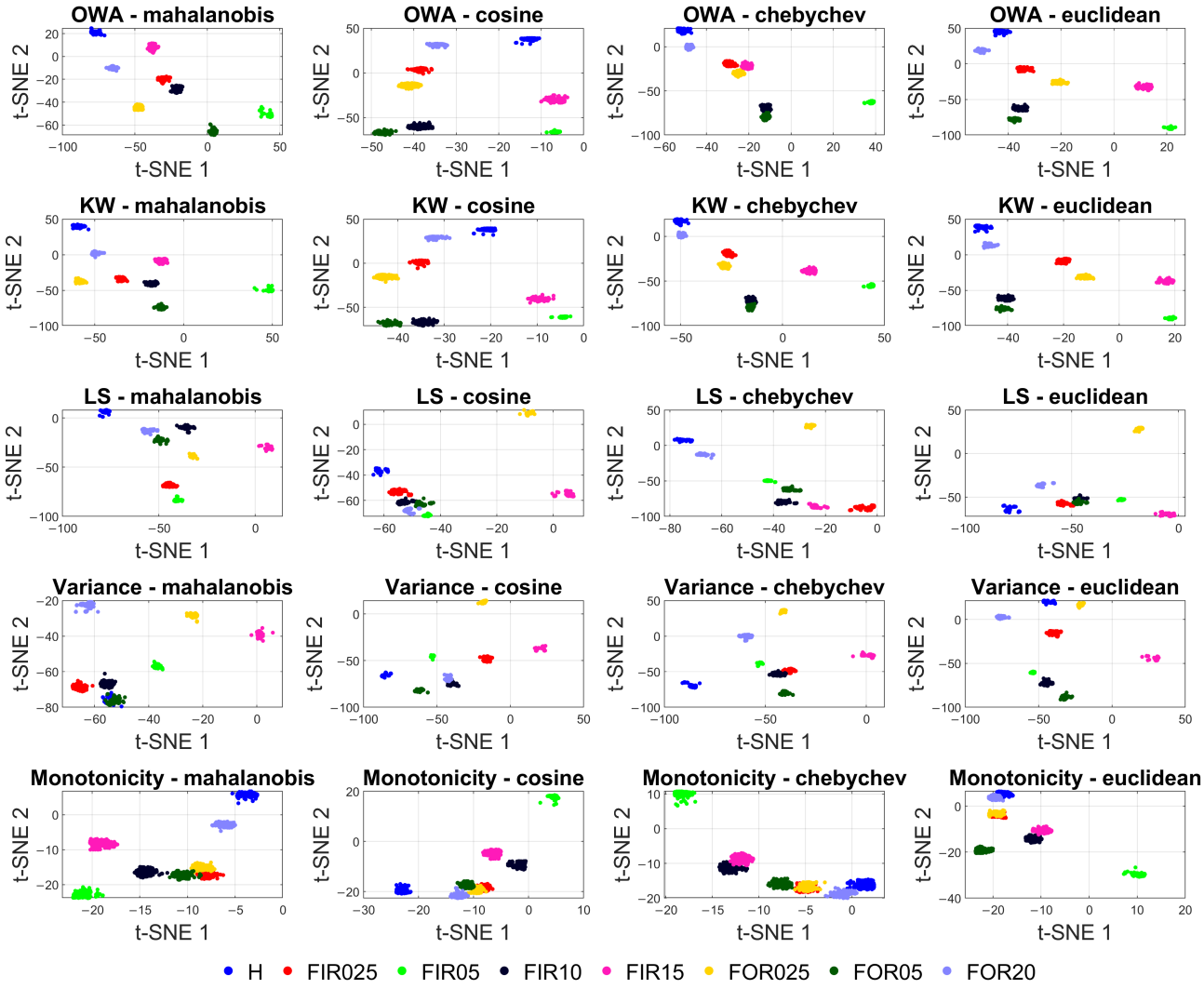


Figure 4.16: t–SNE (**single-cart**, post-transform): The method yields compact and well-separated clusters with consistent spacing across all distance metrics, demonstrating effective density regularization by the proposed approach.

making regions.

4.4.6 Robustness under noisy conditions

To assess the robustness against environmental and instrumentation noise, three representative noise families were exclusively introduced into the vibration channels: (i) additive white Gaussian noise combined with uniform background noise (AWGN + U), (ii) narrowband interference (NB), and (iii) sinusoidal structural noise (SN). The drive and command signals remained unaltered to isolate the effect of the vibration corruption.

In all instances, PCA projections of the seven classes—Healthy (H), FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20—are depicted before and after the transformation. In the *pre-transform* plots, the presence of noise results in significant overlap, shape distortion, and loss of separability among fault categories. The proposed Adaptive Distribution-Based Transformation (IWCCT) restores compact, linearly separable structures across all noise conditions, confirming its robustness to both stochastic and deterministic disturbances.

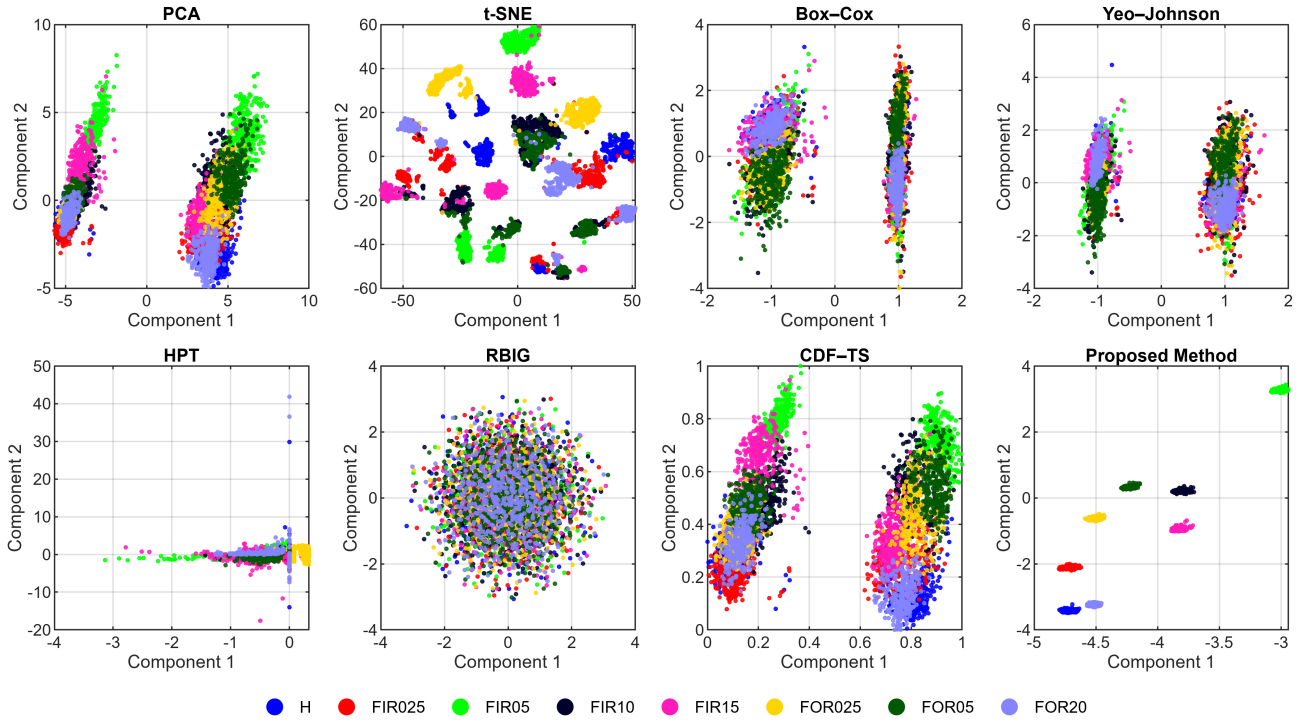


Figure 4.17: In the comparative analysis of two-dimensional embeddings across various methods, as demonstrated in the single-cart experiment, the top row illustrates the PCA and t-SNE baselines alongside the Box–Cox and Yeo–Johnson marginal transformations. The bottom row presents the HPT and RBIG density transformations, CDF–TS, and proposed method. Each color represents a distinct class, namely H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20.

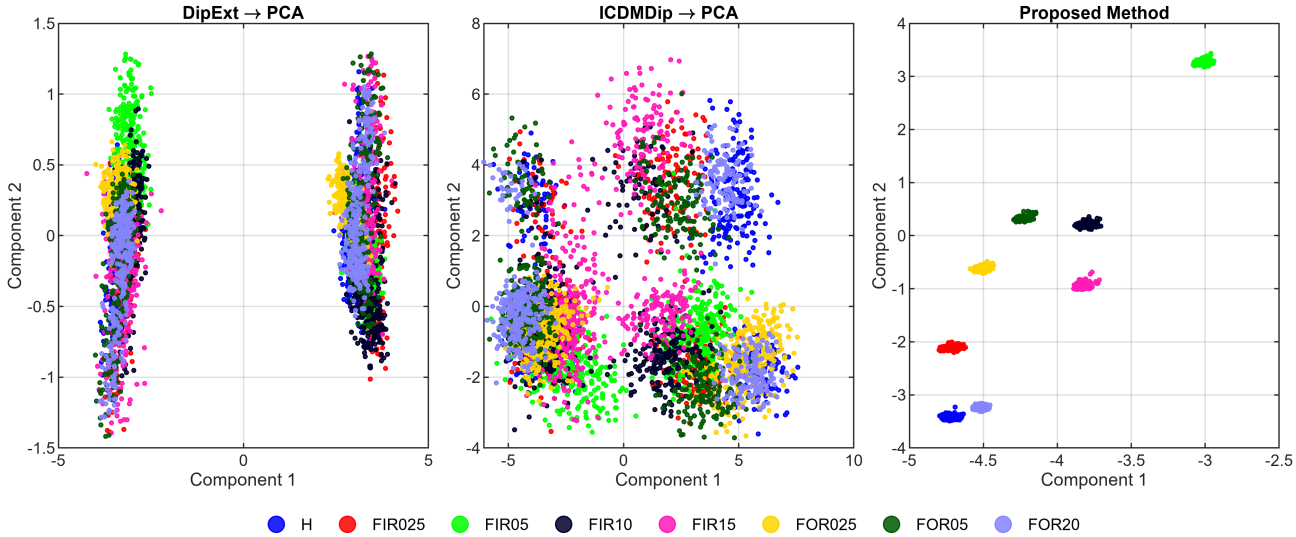


Figure 4.18: Embedding comparison (single-cart): (left) DipExt projected by PCA, (middle) ICDMDip projected by PCA, and (right) the proposed method (native 2D image). The proposed method produces compact and well-separated clusters for all classes.

Broadband noise (Figure 4.19). The combination of AWGN and uniform jitter increases both the angular and radial spreads, causing the healthy and mild fault classes to merge. Following the proposed transformation, all classes are consolidated into compact, well-separated

clusters with uniform spacing and reduced intra-class variance. The method effectively suppresses stochastic noise while maintaining the intrinsic geometry of the fault manifolds, demonstrating its resilience to the additive random perturbations that are common in industrial measurements.

Narrowband interference (Figure 4.20). The injected narrowband amplitude-modulated disturbance introduced anisotropic deformation and ring structures, particularly in LS- and Var-based features. These coherent distortions expand the intra-class variance and obscure the fault boundaries. After the transformation, the rings were eliminated, and each fault class reverted to a circular and compact form. A slight residual overlap appeared only in the transition regions between mild and moderate faults; however, the global separability was preserved. The transformation adaptively suppresses structured and frequency-localized noise, revealing stable manifolds in the principal component analysis (PCA) domain.

Sinusoidal structural noise (Figure 4.21). This represents the most challenging case, which is indicative of structural resonance-type contamination. The pseudo-FRF noise produces curved and entangled manifolds that obscure the class boundaries. After the proposed transformation, each fault class formed a compact cluster with a distinct centroid and a controlled variance. Although the cluster radii were slightly larger than those in the broadband case, the discriminative geometry was fully preserved, confirming the robustness of the method against deterministic-resonance effects.

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

4.4.7 Anomaly Detection in the Original Feature Space

Although PCA and t–SNE were utilized in this study to enable low-dimensional analysis and visualization, the proposed transformation (ADT) does not inherently rely on dimensionality reduction. To evaluate its efficacy within the original feature space, anomaly detection was assessed by applying the transformation directly to the ranked feature set without any intermediate projections. Figure 4.22 presents the anomaly detection results obtained using a one-class SVM in the original feature space for both classical statistical descriptors and the Hjorth parameters. Compared with the pre-transform scenario, the proposed distribution-based transformation results in a significant reduction in false negatives and false positives, thereby markedly enhancing the detection performance. These findings confirm that the proposed method effectively improves class separability and anomaly detectability directly in high-dimensional feature spaces, with dimensionality reduction serving solely as an auxiliary tool for visualization rather than as a prerequisite for performance improvement.

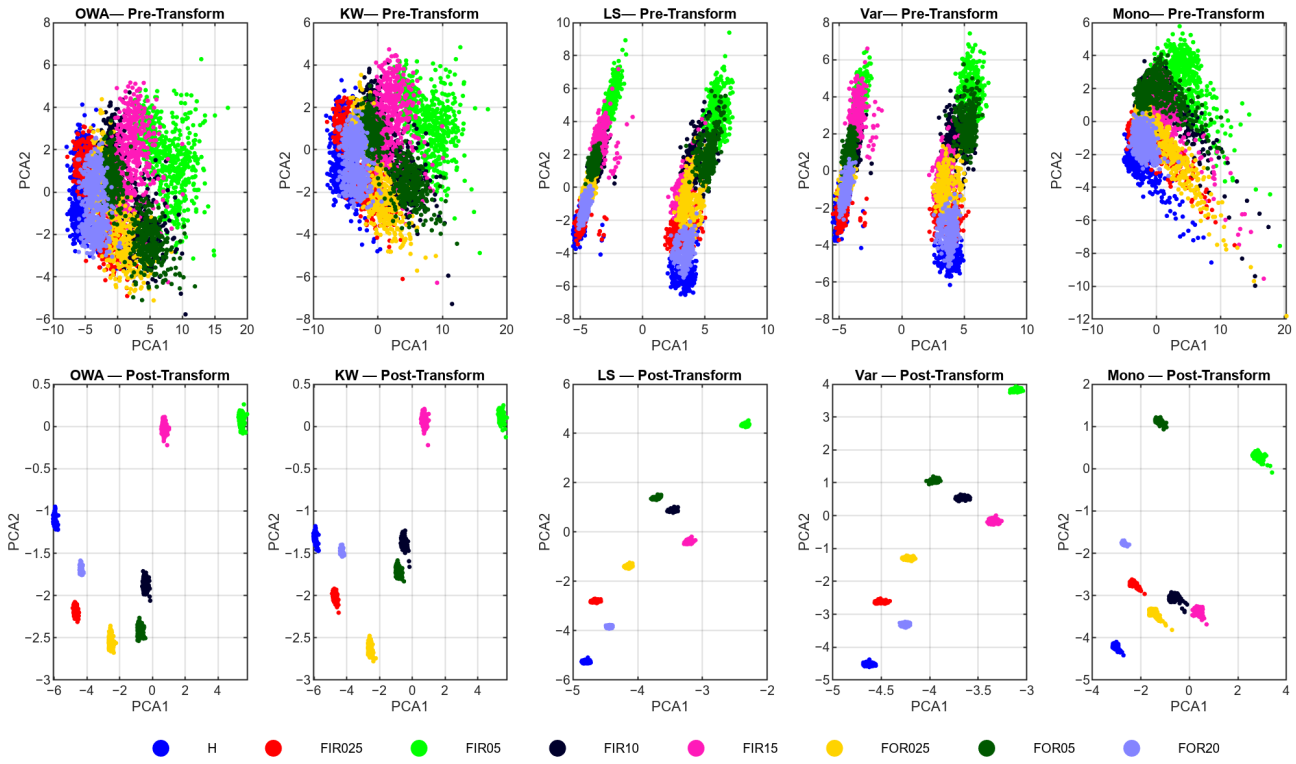


Figure 4.19: The impact of broadband noise, specifically AWGN at 10 dB combined with 2% uniform jitter, on polar diagnostics is illustrated before (top row) and after (bottom row) the application of IWCCT. The columns represent OWA, KW, LS, Variance, and Monotonicity. Prior to transformation, noise increases the angular and radial dispersion; however, the proposed transformation effectively consolidates each class into a compact cluster, thereby restoring class separability.

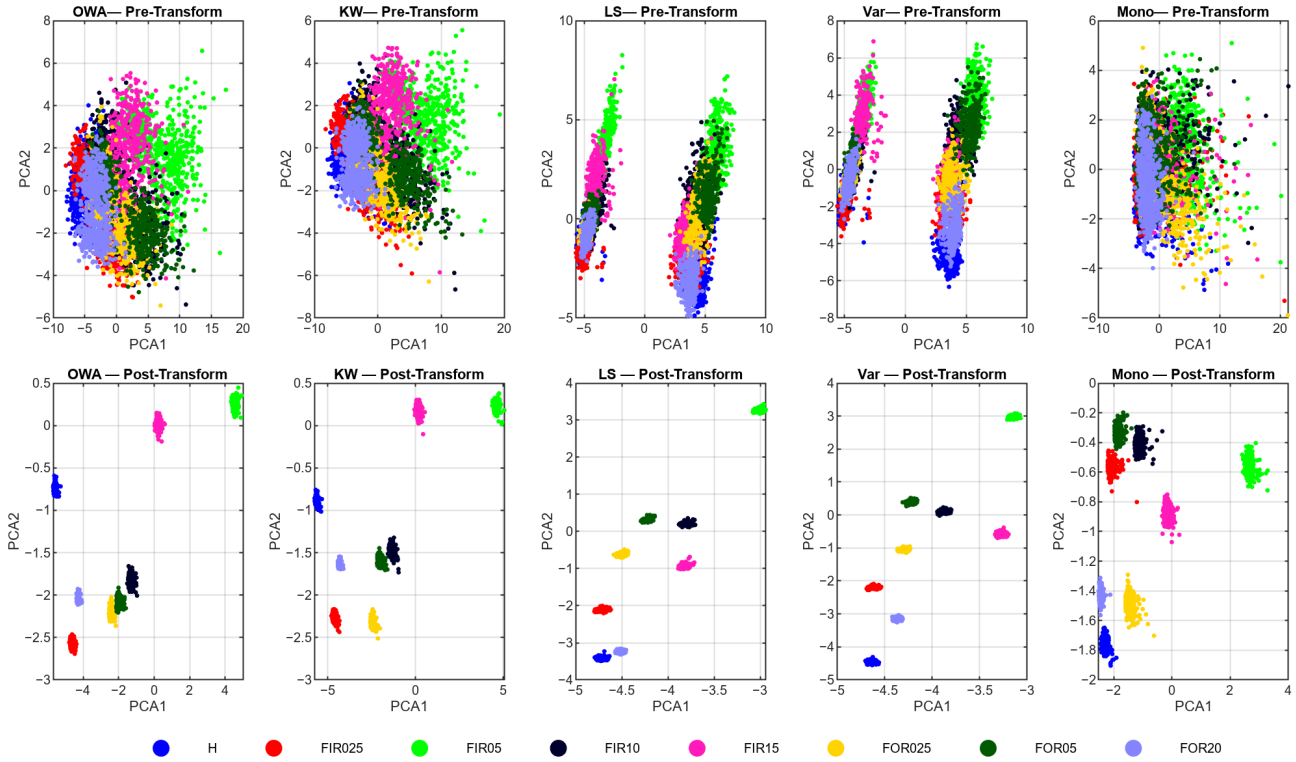


Figure 4.20: The effect of narrowband, amplitude-modulated interference (band-pass $[20, 40]$ Hz, pulse rate f_p=0.5 Hz) was examined. Prior to the application of the proposed transformation (top row), anisotropic smearing and ring-like structures were observed in the images. However, following the implementation of the proposed transformation (bottom row), the clusters became compact with minimal leakage, demonstrating robustness against coherent tonal noise.

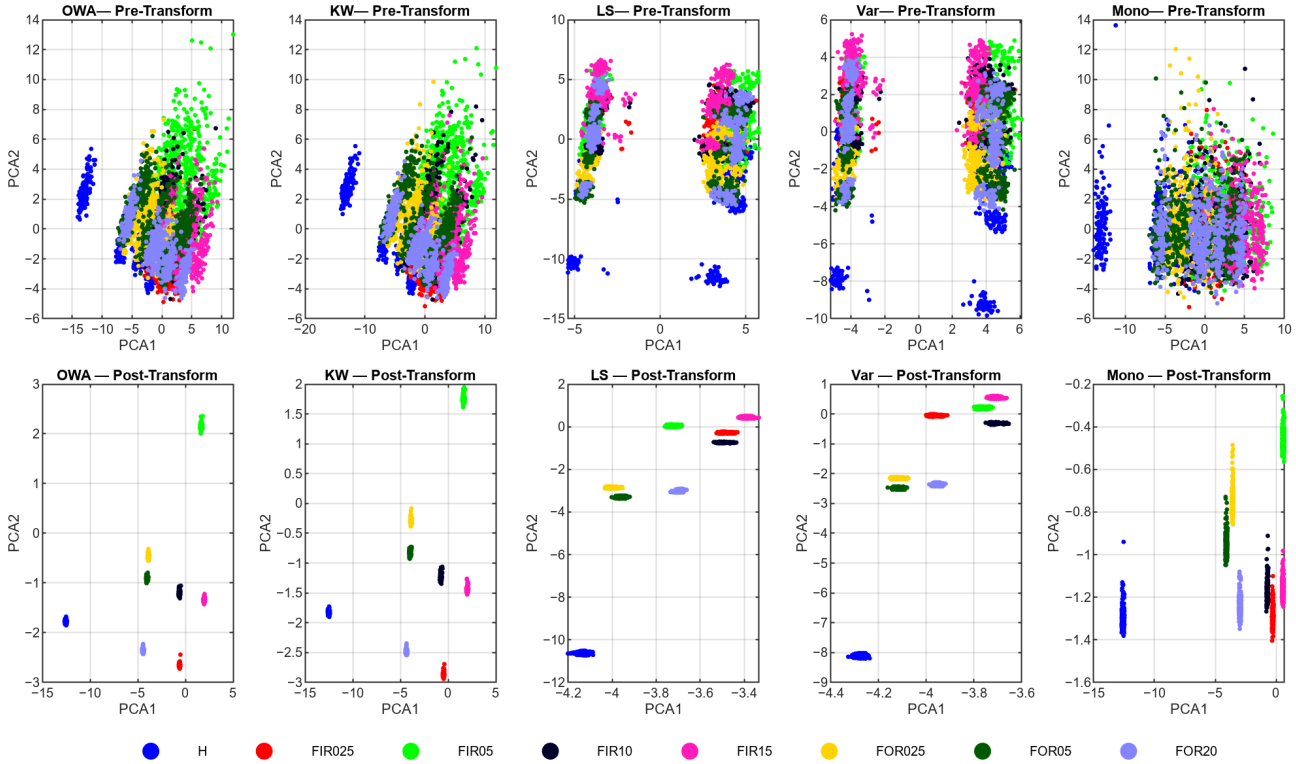


Figure 4.21: The impact of structural noise synthesized from a pseudo-FRF profile is examined. The most significant pre-transform distortion is observed in this context (top row); however, the proposed transformation continues to compact each class into distinct clusters(Bottom row).

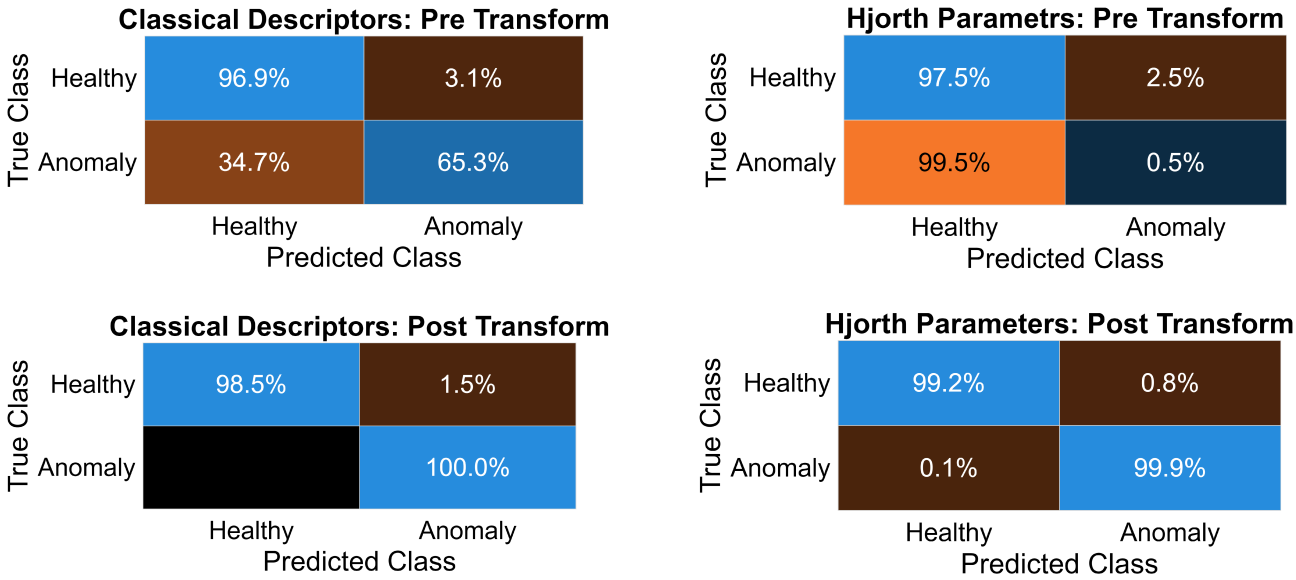


Figure 4.22: The performance of anomaly detection in the *original feature space* without the application of dimensionality reduction is evaluated. Confusion matrices are presented for both classical statistical descriptors and Hjorth parameters, which were assessed before and after implementing the proposed distribution-based transformation. The findings indicate a significant enhancement in anomaly detection accuracy post-transformation, thereby affirming the efficacy of the proposed method in operating directly within high-dimensional feature spaces.

Chapter 5

Iterative Weighted Compact Clustering Transformation

This chapter introduces a fully unsupervised and computationally efficient diagnostic framework for bearing fault detection based on a novel *Iterative Weighted Compact Clustering Transformation* (IWCCT). This method reorganizes dispersed or overlapping feature distributions into compact and more separable structures without requiring class labels or model training. By reshaping the marginal feature distributions prior to dimensionality reduction, the transformation enhances the performance of anomaly detection and clustering algorithms operating in reduced spaces. Previous research has investigated statistical transformations such as Box–Cox (Box & Cox, 1964), Yeo–Johnson (Yeo & Johnson, 2000), and Rotation-Based Iterative Gaussianization (RBIG) (Laparra et al., 2011), along with density-homogenizing approaches including CDF–TS (Zhu et al., 2021), dip scaling (Schelling & Plant, 2020), dip transformation (Schelling & Plant, 2018), and nonlinear cluster enhancement (Schelling et al., 2020) to regularize feature distributions. However, these methods often impose global marginal or joint normalization, which may suppress class-specific statistics. Additionally, density-homogenizing transformations can struggle with bimodal, skewed, and overlapping distribution characteristics that arise in the coupled translational–rotational dynamics of Independent Cart Systems (ICS). In contrast, the proposed transformation performs an *intraclass weighted compaction* that reduces feature dispersion while preserving the interclass structure within the original feature space. This targeted reshaping avoids the limitations of global Gaussianization, which frequently eliminates discriminative variance and does not rely on the uniform or unimodal density assumptions typically required by density-homogenizing methods.

The IWCCT employs a provably convergent, non-expansive update with a computational complexity of $\mathcal{O}(ND)$, enabling label-free and class-preserving reshaping. Experiments conducted on two publicly available datasets—*MOIRA-UNIMORE* (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a) and *Politecnicodataset* (Daga et al., 2019)—demonstrate consistent improvements in fault detectability, particularly under subtle fault conditions. Owing to its deterministic formulation and low computational overhead, the proposed approach is well-suited for real-time industrial environments in which labeled data are scarce. The originality of this study lies in the introduction of a distribution-reshaping mechanism specifically tailored to the highly variable operating conditions and nonstationary behaviors characteristic of ICS platforms.

5.1 Proposed Methodology

This section introduces the (*IWCCT* Algorithm 3), which is an unsupervised per-feature distribution-shaping transform specifically designed for non-stationary vibration diagnostics. The IWCCT is applied to data from a *single class* and functions independently of *labels*. When labels are available, they are utilized solely to *select* one class at a time; the transform itself remains label-independent and is executed separately for every class.

5.1.1 Problem Setup and Notation

Let $X \in \mathbb{R}^{n\times m}$ represent n samples and m features, $x_{(i,j)}$ denote the value of sample i in feature j, and x_j refer to column j. By default, we used a single unlabeled set $\mathcal{I} = \{1, \ldots, n\}$, which was assumed to contain only one class. If coarse labels are present, let $\{\mathcal{I}_c\}_{c=1}^{C}$ be the class index set and apply IWCCT *independently* to each $\mathcal{I}_c$.

5.1.2 Standardization

Each feature was standardized using z-scores to stabilize the scales as follows:

$$x_j \;\leftarrow\; \frac{x_j - \mu_j}{\sigma_j}, \qquad j \in \{1, \ldots, m\}, \tag{5.1}$$

where μ_j and σ_j represent the mean and standard deviation of x_j.

5.1.3 Iterative Compaction on a Single-Class Subset

For R rounds, and for each operating subset $\mathcal{I}$ (defaulting to a single class) and feature j, the scalar list $x_{\mathcal{I},j} = \{x_{(i,j)} : i \in \mathcal{I}\}$ is divided into two groups using 2-means:

$$\{z_i\}, \{\mu_z\} \;=\; \underset{\substack{z_i\in\{1,2\}\\ \mu_1,\mu_2\in\mathbb{R}}}{\arg\min} \sum_{i\in\mathcal{I}} \big(x_{(i,j)} - \mu_{z_i}\big)^2. \tag{5.2}$$

The centers are the means of the clusters.

$$\mu_z \;=\; \frac{1}{|\{i \in \mathcal{I} : z_i = z\}|} \sum_{i\in\mathcal{I}:\, z_i=z} x_{(i,j)}, \quad z \in \{1, 2\}. \tag{5.3}$$

5.1.4 Distance Normalization and Gaussian Weighting

For each $i \in \mathcal{I}$, compute the absolute deviation, as follows:

$$d_i \;=\; \big|x_{(i,j)} - \mu_{z_i}\big|, \tag{5.4}$$

and subsequently normalize it within the set $\mathcal{I}$:

$$\tilde{d}_i \;=\; \frac{d_i}{\max_{u\in\mathcal{I}} d_u}. \tag{5.5}$$

Thereafter, assign a Gaussian weight:

$$w_i \;=\; \exp\!\Big(-\frac{\tilde{d}_i^2}{2\sigma^2}\Big), \tag{5.6}$$

and construct the *weighted compact center* as follows:

$$x_j^{\text{comp}} \;=\; \frac{\sum_{i\in\mathcal{I}} w_i\, x_{(i,j)}}{\sum_{i\in\mathcal{I}} w_i}. \tag{5.7}$$

5.1.5 Adaptive Contraction

We define $v_j = \mathrm{Var}(\{x_{(i,j)} : i \in \mathcal{I}\})$. To mitigate the excessive contraction of features with low variance, we employ the following scaling factor:

$$s_j \;=\; \begin{cases} \dfrac{\alpha}{1+v_j}, & v_j > \tau, \\ \alpha, & \text{otherwise}, \end{cases} \qquad \alpha \in (0,1]. \tag{5.8}$$

Update the values towards x_j^{comp} using the rule:

$$x_{(i,j)}^{(r)} \;=\; x_{(i,j)}^{(r-1)} - s_j\Big(x_{(i,j)}^{(r-1)} - x_j^{\mathrm{comp}}\Big), \quad i \in \mathcal{I}. \tag{5.9}$$

This can be equivalently expressed as $x_{(i,j)}^{(r)} = (1-s_j)\, x_{(i,j)}^{(r-1)} + s_j\, x_j^{\mathrm{comp}}$, representing a convex combination.

5.1.6 Final Rescaling

Upon completion of R iterations, final z-score rescaling was applied as described in (5.1) to ensure comparability across features.

5.1.7 Operating Modes and Downstream Use

One-class (default). The IWCCT was trained on a single-class dataset, which eliminated the need for labels. In the transformed space, there is an increase in compactness and a reduction in heavy tails, which facilitates the application of straightforward decision rules (e.g., distance to x^{comp}, Mahalanobis distance to the compacted cluster, or one-class detectors such as one-class support vector machine (OCSVM)/isolation fores (IF)) to identify *non-healthy* data points. *Multiclass (optional).* When labels are available, the IWCCT is applied *independently* to each class subset, $\mathcal{I}_c$. The labels are utilized solely for the selection of $\mathcal{I}_c$ and do not influence the transformation process, thereby ensuring procedural consistency and preservation of class margins.

5.2 Theoretical Properties of IWCCT

5.2.1 Convergence Analysis

The iterative update rule of IWCCT is articulated as follows:

$$x_{(i,j)}^{(r)} = (1-s_j)\, x_{(i,j)}^{(r-1)} + s_j\, x_j^{\mathrm{comp}}, \tag{5.10}$$

where x_j^{comp} denotes the weighted compact center, and $s_j \in (0,1)$ is the adaptive scaling factor. Equation 5.10 represents the convex combination of the previous value and weighted compact center. This can be reformulated as:

$$x_{(i,j)}^{(r)} - x_j^{\mathrm{comp}} = (1-s_j)\big(x_{(i,j)}^{(r-1)} - x_j^{\mathrm{comp}}\big). \tag{5.11}$$

Consequently,

$$\big|x_{(i,j)}^{(r)} - x_j^{\mathrm{comp}}\big| = (1-s_j)^r \big|x_{(i,j)}^{(0)} - x_j^{\mathrm{comp}}\big|. \tag{5.12}$$

Algorithm 3 IWCCT: Iterative Weighted Compact Clustering Transform

Require: $X \in \mathbb{R}^{n \times m}$; R; σ; τ; α; $k \leftarrow 2$
Require: $c(i)$ (optional; used only to select a class subset)
Ensure: $\widetilde{X}$

Standardize
for $j = 1$ to m **do**
    $x_j \leftarrow \dfrac{x_j - \mu_j}{\sigma_j}$
end for
Select operating subset(s)
if $c(i)$ available **then**
    $\mathcal{S} \leftarrow \{\mathcal{I}_c\}_{c=1}^{C}$
else
    $\mathcal{S} \leftarrow \big\{\{1, \ldots, n\}\big\}$
end if
Iterative compaction
for $r = 1$ to R **do**
    for each $\mathcal{I} \in \mathcal{S}$ **do**
        for $j = 1$ to m **do**
            $x_{\mathcal{I},j} \leftarrow \{x_{(i,j)} : i \in \mathcal{I}\}$
            $(z, \{\mu_z\}_{z=1}^{k}) \leftarrow \mathrm{kmeans}(x_{\mathcal{I},j},\ k{=}2)$
            for $i \in \mathcal{I}$ **do**
                $d_i \leftarrow |x_{(i,j)} - \mu_{z_i}|$
            end for
            $\tilde{d}_i \leftarrow d_i \big/ \max_{u \in \mathcal{I}} d_u \quad$ for all $i \in \mathcal{I}$
            $w_i \leftarrow \exp\big(-\tilde{d}_i^2/(2\sigma^2)\big) \quad$ for all i
            $x_j^{\mathrm{comp}} \leftarrow \dfrac{\sum_{i \in \mathcal{I}} w_i\, x_{(i,j)}}{\sum_{i \in \mathcal{I}} w_i}$
            $v_j \leftarrow \mathrm{Var}\big(\{x_{(i,j)} : i \in \mathcal{I}\}\big)$
            $s_j \leftarrow \begin{cases} \alpha/(1+v_j), & v_j > \tau \\ \alpha, & \text{otherwise} \end{cases}$
            for $i \in \mathcal{I}$ **do**
                $x_{(i,j)} \leftarrow x_{(i,j)} - s_j\big(x_{(i,j)} - x_j^{\mathrm{comp}}\big)$
            end for
        end for
    end for
end for
Final rescale
for $j = 1$ to m **do**
    $x_j \leftarrow \dfrac{x_j - \mu_j}{\sigma_j}$
end for
return $\widetilde{X} \leftarrow X$

Given that $0 < s_j \leq 1$, it follows that

$$\lim_{r \to \infty} x_{(i,j)}^{(r)} = x_j^{\mathrm{comp}},$$

which implies that the iteration constitutes a *contractive mapping* that converges exponentially to a unique fixed point x_j^{comp}. The convergence rate is $(1 - s_j)^r$, indicating accelerated convergence for larger s_j values.

5.2.2 Stability

Consider two initial states differing by $\Delta^{(0)} = x_{(i,j)}^{(0)} - \tilde{x}_{(i,j)}^{(0)}$. From (5.10),

$$\Delta^{(r)} = (1 - s_j)\, \Delta^{(r-1)}, \tag{5.13}$$

which results in

$$|\Delta^{(r)}| = (1 - s_j)^r |\Delta^{(0)}|. \tag{5.14}$$

Therefore, minor perturbations in the input lead to geometrically decaying deviations in the output, confirming that the transformation is *Lipschitz-stable* with constant $L = (1 - s_j)$. For any two datasets X_1, X_2 with bounded entries, the final transformed outputs satisfy

$$\|\widetilde{X}_1 - \widetilde{X}_2\|_2 \leq (1 - s_{\min})\, \|X_1 - X_2\|_2, \tag{5.15}$$

ensuring bounded-input–bounded-output (BIBO) stability.

5.2.3 Role of Gaussian Weighting

The Gaussian weighting term

$$w_i = \exp\Big(-\frac{\tilde{d}_i^2}{2\sigma^2}\Big) \tag{5.16}$$

functions as a localized attention kernel, modulating the extent to which each sample contributes to the compact center. A smaller σ value assigns greater weights to proximate samples, thereby achieving sharper local compaction, whereas a larger σ value facilitates a smoother and more globally stable contraction. Consequently, σ effectively mediates the balance between *locality* and *stability*.

5.2.4 Hyperparameter Interpretation

The parameters $(\alpha, \sigma, \tau, R)$ affect convergence speed, compaction strength, and variance preservation. The theoretical roles and practical ranges are listed in Table 5.1.

5.2.5 Non-Expansive Operator Property

The IWCCT iteration can be represented as an operator $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\mathcal{T}(x) = (1 - s_j)\, x + s_j\, x_j^{\text{comp}}. \tag{5.17}$$

For any $x, y \in \mathbb{R}^n$,

$$\|\mathcal{T}(x) - \mathcal{T}(y)\|_2 = (1 - s_j)\|x - y\|_2. \tag{5.18}$$

Given that $0 < (1 - s_j) < 1$, the operator $\mathcal{T}$ is characterized as *non-expansive* (Banach contraction), which ensures that the IWCCT update possesses a unique fixed point and globally convergent dynamics. This property further implies that the iterative applications of $\mathcal{T}$ form a Cauchy sequence in Euclidean space, thereby guaranteeing both stability and convergence.

5.2.6 Hyperparameter Selection and Tuning Criteria

The IWCCT algorithm incorporates four hyperparameters $(\alpha, \sigma, \tau, R)$ that regulate contraction strength, locality, scaling, and iteration depth. The theoretical foundation and tuning rationale for these parameters are as follows:

5.2.6.1 Adaptive scaling factor α

The parameter α determines the contraction rate $s_j = \alpha/(\alpha + \tilde{v}_j)$, where $\tilde{v}_j$ represents the normalized feature variance. From the convergence condition $(1 - s_j)^r \to 0$, it is evident that $0 < \alpha < \infty$ ensures convergence. A smaller α results in stronger compaction (faster convergence but potential over-shrinkage), whereas a larger α retains more variance. A practical range is $\alpha \in [0.1, 0.95]$, selected such that $\mathrm{Var}(\widetilde{X})/\mathrm{Var}(X) \approx 0.6$–$0.9$.

5.2.6.2 Gaussian bandwidth σ

The bandwidth σ defines the effective support of the weighting kernel $w_i = \exp(-\tilde{d}_i^2/2\sigma^2)$ and determines locality. Theoretically, σ balances the bias and variance: a smaller σ increases the bias (strong local compaction), whereas a larger σ increases the variance (weaker local compaction). It is recommended to set σ proportional to the mean pairwise distance, $\sigma = \kappa\,\mathrm{mean}(\tilde{d}_i)$ with $\kappa \in [0.1, 1.0)$.

5.2.6.3 Threshold τ

The threshold τ is a critical parameter that dictates the cessation of the iterative update process, contingent upon the condition that the normalized mean change $\|\widetilde{X}^{(r)} - \widetilde{X}^{(r-1)}\|_2/\|X\|_2 < \tau$. According to the exponential decay described in (5.12), convergence within the bounds of numerical precision ϵ is assured when $\tau < \epsilon$. Typically, values of τ within the range $[0.01, 0.5]$ effectively balance the precision and computational runtime. In the current implementation, this parameter also functions as a variance-based stability gate, which moderates the scaling factor in scenarios of high local dispersion while theoretically embodying the convergence bound derived in (5.12).

5.2.6.4 Iteration depth R

Considering the exponential convergence rate $(1 - s_j)^r$, the requisite number of iterations is determined by the condition $R > \log(\epsilon/\|x^{(0)} - x^{\mathrm{comp}}\|)/\log(1 - s_j)$. Empirical evidence suggests that $R = 3$–5 iterations are sufficient to achieve a residual error of less than 1% for s_j. These criteria facilitate deterministic parameter initialization without the need for cross-validation or label tuning, thereby maintaining the unsupervised nature of the IWCCT.

Table 5.1: Role of IWCCT Hyperparameters and their Typical Ranges

Parameter	**Role**	**Typical Range / Effect**
α	Global contraction strength	$[0.1, 0.95]$; higher $\alpha \to$ faster convergence, less compaction
σ	Gaussian weighting spread	$\kappa \in [0.1, 1.0)$ of mean distance; controls locality
τ	threshold	$[0.01, 0.5]$
R	Iteration count	3–5; ensures residual $< 1\%$

5.2.7 Overall Guarantee

Given bounded X and fixed parameters (α, σ, τ), the IWCCT iteration establishes a monotone contraction mapping in $\mathbb{R}^{n\times m}$. Consequently, the transformation converges to a stable equilibrium that continuously depends on the input, thereby ensuring robustness and consistency across various initialization conditions.

5.3 Experimental Protocol and Evaluation Setup

5.3.1 Dataset and Signal Segmentation

Experiment Type2 was utilized from the MOIRA–UNIMORE ICS bearing dataset (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a). The test cart consistently traversed a 500 mm curved module at a nominal speed of 1000 mm/s (Figure 4.11). The vibration signals were sampled at a frequency of 50kHz. The position served as a gating variable to divide each traversal into *forward* and *reverse* windows, thereby ensuring condition-consistent segments for feature extraction (Figure 1.5).

5.3.1.1 Window length

A nominal traversal of 500 mm at 1000 mm/s requires 0.50 s . Accounting for acceleration and deceleration near direction changes, each gated segment spanned approximately 0.66s, which, at 50kHz, corresponds to $\approx$ 33,000 samples per window. Forward and reverse passes were processed independently.

5.3.1.2 Recording protocol and data volume

For each class under investigation (healthy and each fault condition), the experiment was conducted for five independent realizations, with each realization lasting 30–45 s (multiple back-and-forth traversals). Position-gated windowing results in approximately 550–600 windows *per class*, with each window containing $\approx$ 33,000 samples.

5.3.1.3 Partitioning strategy

All realizations for a given class were initially segmented as described above and then *pooled by class*. Subsequently, we performed a stratified split of the pooled windows into 80% training and 20% testing per class (the forward/reverse proportions were preserved). This split was applied at the window level (i.e., windows from the same realization could appear in both partitions). All feature extraction and model fitting utilized only the training portion, and a 20% hold-out was reserved exclusively for testing.

5.3.2 Feature Extraction

Features were derived from each segmented window by utilizing the vibration channels and system variables, with the exception of cart position and actual velocity. To ensure simplicity, interpretability, and reproducibility, two complementary feature families were employed: (1) statistical descriptors and (2) Hjorth parameters. These attributes were selected for their computational efficiency, minimal requirement for expert tuning or specialized pre-processing, and high interpretability of the physical signal behavior, such as energy distribution, impulsiveness,

and temporal variability. Compared with complex time–frequency or machine learning-based feature extraction techniques, these descriptors require significantly fewer computational resources and less domain expertise, rendering them particularly suitable for real-time industrial monitoring, where data volume and latency constraints are critical issues. Furthermore, their application ensures that the observed performance improvements originate from the proposed IWCCT transformation rather than alterations in the feature design or model complexity. Thus, these well-established and physically meaningful features serve as standardized and transparent baselines for evaluating distribution-shaping transformations in nonstationary vibration diagnostics. Table 3.6 and Table 4.2 summarize the definitions of the two sets of parameters extracted for each windowed segment of the data.

5.3.3 Feature Ranking Strategies

Five complementary ranking criteria were employed to identify informative features: one-way Analysis of Variance (ANOVA) (Fisher, 1925; Montgomery, 2017; Neter et al., 1996), the Kruskal–Wallis test (Kruskal & Wallis, 1952; Sheskin, 2003), Laplacian Score (He et al., 2005), variance ranking, and a monotonicity index (Coble & Hines, 2009). These collectively capture the linear separability (ANOVA), distributional shifts robust to non-Gaussian data (Kruskal–Wallis), locality/manifold structure (Laplacian Score), dispersion (variance), and progression trends (monotonicity).

5.3.4 Feature Transformation (IWCCT)

The IWCCT was applied independently to each channel feature set to reduce dispersion and overlap in the raw feature distributions. The operating assumption was a single-class dataset (labels were not required). When labels are available, they are used only to select one class subset at a time, whereas the transformation is the same. The IWCCT reshapes skewed distributions towards compact, unimodal forms, decreases intraclass variance, and enhances local clustering. Subsequent dimensionality reduction yielded clearer embeddings for visualizing cluster structures and anomalies. The hyperparameters are listed in Table 5.2.

5.3.5 Rationale for chosen hyperparameters

Our selection was guided by an application-driven criterion to *avoid false alarms*, while producing *compact and well-separated class clusters*. The choices are as follows: (i) *cluster count* k=2 (in 1-D) is sufficient to isolate heavy tails or bimodality while maintaining the step linear time, $k>2$ yields negligible gains with additional cost. (ii) *Round* R is kept small (1–3); beyond three rounds, marginal compaction saturates as the execution time increases. (iii) The *weight scale* σ is set as a fixed fraction of the normalized distance (distances are rescaled to $(0, 1)$ inside the algorithm), which makes the locality of the weights data-size invariant; the reported value consistently produced compact but non-collapsed marginals across classes. (iv) The *variance threshold* τ triggers adaptive shrinkage only for highly dispersed features. We used a unit variance reference after standardization so that features near the unit spread remained unchanged. (v) The *base step size* α is selected from a coarse grid as the smallest value that achieves stable monotonic compaction across classes without oscillations. (vi) For the auxiliary parameters, 1-D k-means `MaxIter` and `Replicates` were kept small to maintain throughput (replicates reduce sensitivity to initializations), and the added *noise factor* was set to a very small value to break numerical ties without altering the statistics.

Table 5.2: IWCCT hyperparameters used throughout. Values are held fixed across all experiments unless noted.

Parameter	N_{KMeans}	N_r	α	σ	τ	R	η
Value	500	5	0.8	0.01	0.01	3	0.01

Notes: N_{KMeans} and N_r are the implementation settings for the 2-means step for each feature; the core transform is otherwise label-free and a single class by default. α, σ, τ, R match the notation in Section 5.1. η is an optional robustness constant (e.g., the weight floor or numerical stability term) and is not required in the updated analytical equations.

5.3.6 Dimensionality Reduction

Principal component analysis (PCA) (Jolliffe & Cadima, 2016; Kitao, 2022; Pořízka et al., 2018) was applied to the top-ranked features to generate low-dimensional embeddings. For comparative purposes, t-distributed stochastic neighbor embedding (t-SNE) was used to visualize the structures that were not linearly separable. Given that the IWCCT compacts per-feature distributions, the leading principal components capture a larger portion of the signal's variance. t-SNE preserves more neighborhood structures. This resulted in embeddings with reduced overlap and enhanced separability, which benefited downstream clustering and anomaly detection.

5.3.7 Evaluation Setup

The performance was assessed across various combinations of embeddings, transformations, and detectors/classifiers. The results are presented as F1 scores in Tables 5.7 and 5.8.

5.3.7.1 Unsupervised anomaly detection

Table 5.7 compares the one-class SVM (OCSVM) and Isolation Forest (IF) operating on embeddings from PCA and t-SNE. For each embedding, the following transformations were contrasted: original (no transform), Box–Cox, Yeo–Johnson, hyperbolic power transformation (HPT), RBIG, CDF–TS, and the proposed IWCCT. Two feature families were evaluated: (i) classical descriptors (ranked by OWA, KW, LS, Variance, and Monotonicity) and (ii) Hjorth parameters.

5.3.7.2 Supervised classification

Table 5.8 extends the comparison to include Support Vector Machines (SVM) and k-Nearest neighbors (KNN), paralleling the unsupervised setup to facilitate direct cross-paradigm comparison.

5.3.7.3 Evaluation Design Overview

This study included four factors: (i) *embeddings* (PCA and t-SNE); (ii) *feature transforms* (original, Box–Cox, Yeo-Johnson, HPT, RBIG, CDF–TS, and IWCCT); (iii) *learning paradigms*, unsupervised anomaly detection (OCSVM and IF), and supervised classification (SVM and KNN); and (iv) *feature families* (classical descriptors and Hjorth parameters). This design isolates the incremental effect of the IWCCT while maintaining other factors constant. Figure 5.1 shows the complete pipeline of this study.

5.3.7.4 Computational complexity and execution time

The principal computational costs associated with all transformations are detailed in Table 5.3, where N denotes the number of samples, D the number of features, L the number of RBIG layers, ℓ the number of ICDMDip iterations, and T the number of CDF–TS iterations. To evaluate real-time applicability, we measured the execution time exclusively for the transformation steps, excluding file I/O, plotting, and other preprocessing steps, following a single warm-up across ten runs. Table 5.4 displays the mean time in seconds on a laptop equipped with an Intel® Core™ i9-12900HK (12th Gen, 2.50GHz), 32GB RAM, and 64-bit Windows 11. MATLAB R2024a was used for the IWCCT, HPT, RBIG, CDF–TS, and Box–Cox analyses. Yeo–Johnson was executed in Visual Studio Code with Python 3.13.5. DipExt was run in R, and ICDMDip was run using Java. For class-conditional methods (IWCCT, Box–Cox, Yeo–Johnson, and HPT), the transformation was applied independently to each of the eight classes and the combined dataset runtime was reported as the sum. Minor discrepancies between the empirical times in Table 5.4 and the large-O trends in Table 5.3 are expected owing to the constant factors and implementation details from different authors, libraries and languages.

Table 5.3: Computational complexity of distribution–shaping transforms compared in this study

Method	**Time complexity**
IWCCT	$O(ND)$
Box–Cox	$O(ND)$
Yeo–Johnson	$O(ND)$
HPT	$O(ND)$
RBIG	$O\big(L\,(D\,N\log N + N\,D^2) + L\,D^3\big)$
CDF–TS	$O(T\,N^2 D)$
DipExt	$O(N\log N) + O(D\log D)$
ICDMDip	$O(\ell\,D^2\,N\log N)$

Notes: N: number of samples; D: feature dimension; L: RBIG layers; T: CDF–TS iterations; ℓ: ICDMDip iterations.

Table 5.4: Execution time (s) for transforms on the combined dataset (N_{tot}, D). N_{tot} and D denote the number of samples and features, respectively.

Method	**Time (s)**
IWCCT (sum over classes)	0.327
Box–Cox (sum over classes)	0.048
Yeo–Johnson (sum over classes)	0.021
HPT (sum over classes)	0.199
RBIG (L=500, PCA rotations)	0.596
CDF–TS (T iters to threshold)	0.875
DipExt	4.779
ICDMDip	61.28

Notes: "Sum over classes" add per-class runtime to give the total for the combined dataset.

5.3.7.5 Computational efficiency and resource consumption

Under consistent hardware conditions, the approximate memory footprint was evaluated to assess the resource utilization. For the largest dataset (4500 samples×30 features), the transformation required less than 10MB of memory and completed each run in less than 0.4s, thereby

confirming the lightweight nature of IWCCT and its suitability for real-time industrial deployment. As the algorithm involves only feature-wise variance scaling, exponential weighting, and iterative averaging, its time complexity scales linearly with both the number of samples and features ($O(ND)$), without requiring matrix inversion or backpropagation. Extrapolating from these results to lower-power processors, we estimated the latency using proportional clock rate scaling and a small allowance for memory overhead. Table 5.5 summarizes the empirical and extrapolated runtimes across representative platforms, demonstrating that even entry-level embedded hardware (e.g., Raspberry Pi 4 or Jetson Nano) can execute a full IWCCT transformation within 1 s per feature batch. The deterministic, training-free formulation also eliminates stochastic initialization and parameter tuning overheads, ensuring consistent execution under constrained computational and memory budgets.

Table 5.5: Estimated execution latency of IWCCT across representative hardware platforms

Platform	**GHz**	**Scale**	**Time (s)**	**Rationale**
Intel i9-12900HK	2.5	1.00	0.33–0.40	Empirical measurement
Intel i7-12700H	2.3	$2.5/2.3 \approx 1.09$	0.36–0.44	∼9% slower, same core class
Raspberry Pi 4 (ARM A53)	1.5	$2.5/1.5 \approx 1.67$	0.55–0.67	Fewer cores; lower clock
Jetson Nano (ARM A57)	1.4	$2.5/1.4 \approx 1.79$	0.59–0.72	Similar to Pi 4; GPU idle
Embedded buffer (thermal/memory)	—	+20%	≈0.7–0.9	Conservative bound

Note: The embedded runtimes were extrapolated from the measured 0.33–0.40 s on an Intel i9–12900HK (2.5 GHz) by employing proportional clock-rate scaling and incorporating a minor memory-latency allowance. Given that the IWCCT scales linearly with both samples and features, these projections offer realistic upper bounds for scenarios involving low-power deployment.

5.3.8 Results without additive Noise

Figure 5.2 and Figure 5.3 depict the geometric impact of the IWCCT on PCA embeddings. For each ranking criterion (columns: OWA, KW, LS, Variance, Monotonicity), features were initially ranked, and the selected set was projected into two dimensions (PC1 and PC2) using PCA. The top row illustrates the PCA embedding prior to the IWCCT, whereas the bottom row shows the PCA embedding post-IWCCT, employing the same ranking criterion and preprocessing. The points represent individual windows and are color-coded according to class (H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20). For clarity, the $(\mathrm{PC1}, \mathrm{PC2})$ plane is represented in polar coordinates (r, θ), where θ denotes the direction and r signifies the magnitude (Euclidean norm) after centering and scaling. Polar mapping was performed solely for visualization.

Before transformation, the embeddings formed elongated clouds with significant angular overlap and occasional bimodality, particularly between nearby severities. Following the IWCCT, within-class dispersion contracts into compact clusters, and the between-class margins expand, resulting in distinct locations in (r, θ) for each class. This effect was consistent across the ranking criteria and for both feature families (classical descriptors and Hjorth parameters), suggesting that the improvements stemmed from reshaping the distribution rather than from a specific feature selection. These visual trends correspond to the quantitative enhancements reported in Tables 5.7 and 5.8.

In both embeddings, classical distribution-matching transforms such as RBIG, Box-Cox, and Yeo-Johnson yielded the lowest F1 scores. Under PCA (Table 5.7), RBIG scores approximated chance levels ($\approx$ 0.09–0.12), and Box-Cox/Yeo-Johnson rarely exceeded 0.6, indicating weaker

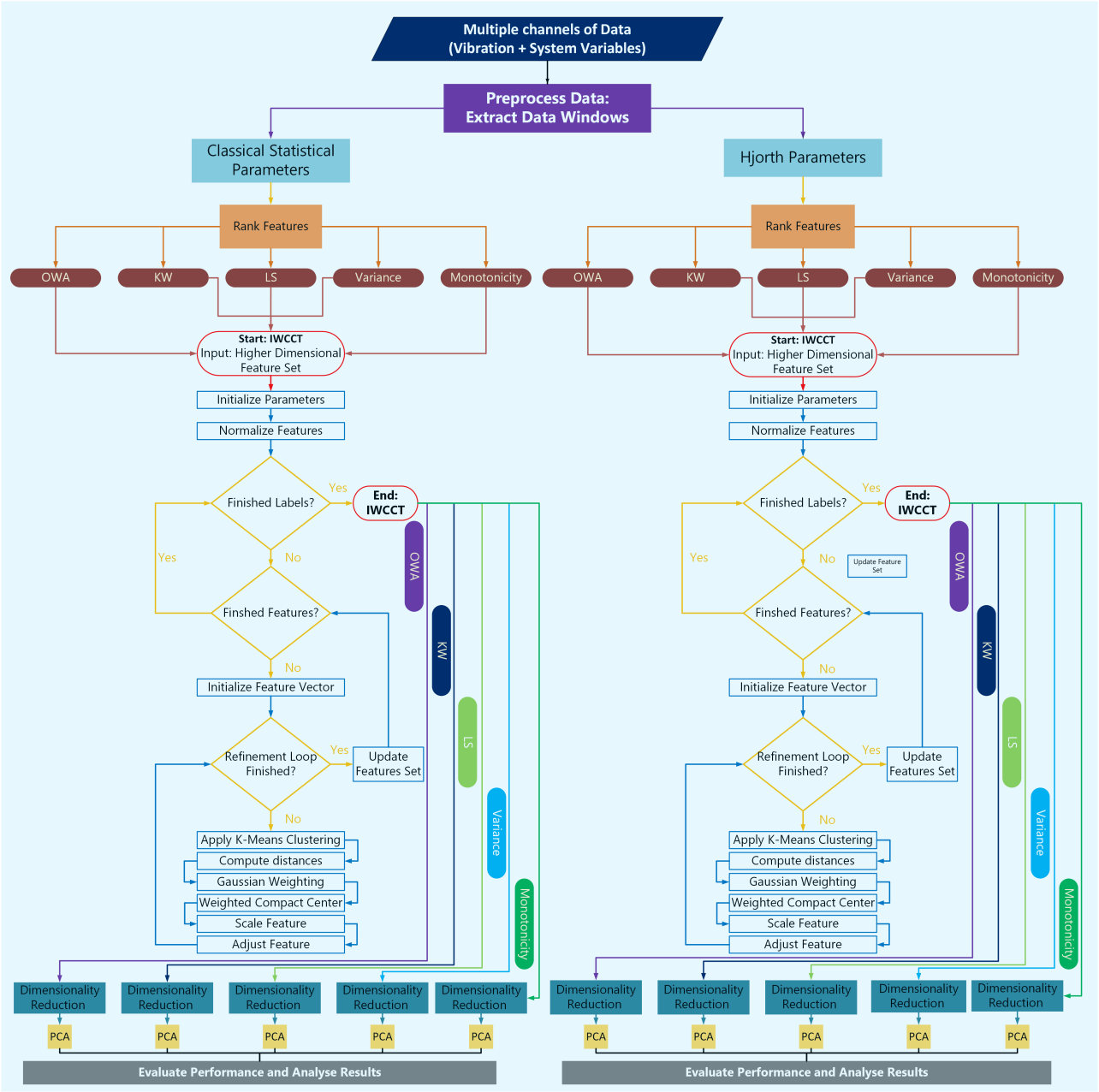


Figure 5.1: End-to-end workflow for IWCCT-based framework. Multiple synchronous channels (vibration and system variables) were windowed, and two feature families were formed in parallel: classical statistical descriptors and Hjorth parameters. Each family was rank-ordered using five criteria (one-way analysis of variance (ANOVA), Kruskal–Wallis, least-squares trend, variance, and monotonicity). The IWCCT then refines the features via a loop that normalizes the inputs, applies 2-means per feature, computes distances with Gaussian weighting, forms a weighted compact center, and rescales the features. Finally, low-dimensional embeddings are obtained (PCA = principal component analysis) for performance evaluation and analysis.

performance on Hjorth features. Similar patterns were observed for t-SNE, where the RBIG scores ranged from ≈ 0.10 to 0.15.

The original and CDF-TS transformations exhibited intermediate performances. On PCA, OCSVM/IF in the raw space achieved scores between 0.73 and 0.83 for classical descriptors

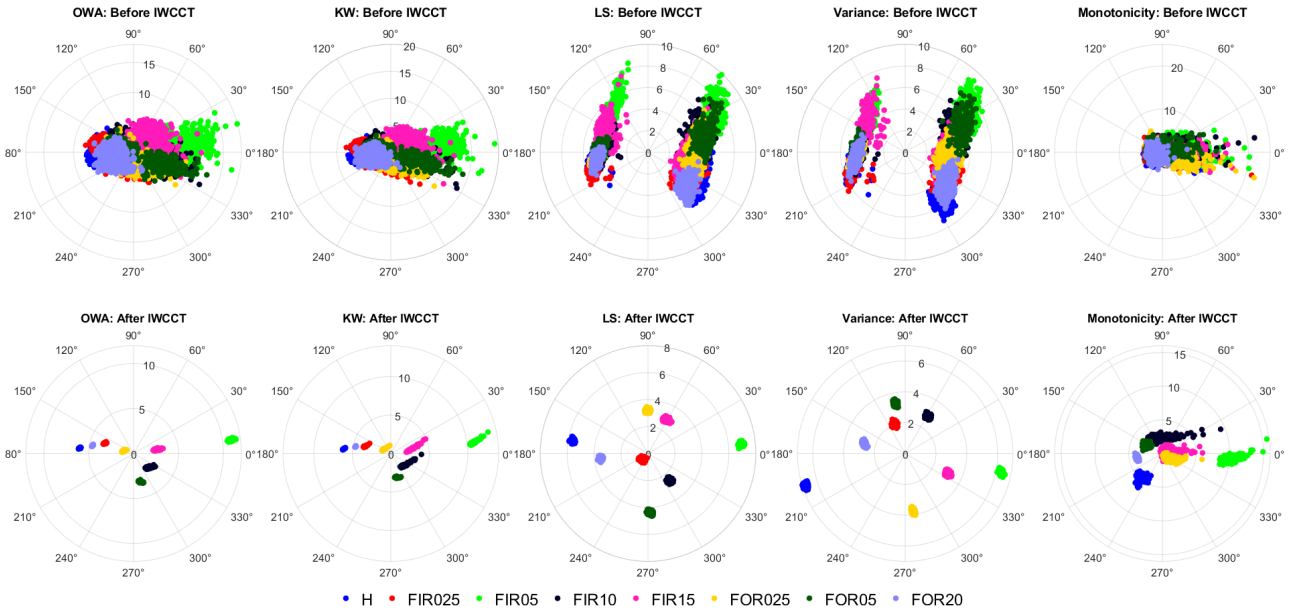


Figure 5.2: PCA embeddings in polar coordinates using classical statistical descriptors. The columns correspond to the ranking criteria (OWA, KW, LS, variance and monotonicity). Top row: PCA before IWCCT; bottom row: PCA after IWCCT. The samples are colored according to class (H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20). The $(\mathrm{PC1}, \mathrm{PC2})$ plane is rendered in polar coordinates (r, θ) for readability; IWCCT compacts elongated clusters and improves separability.

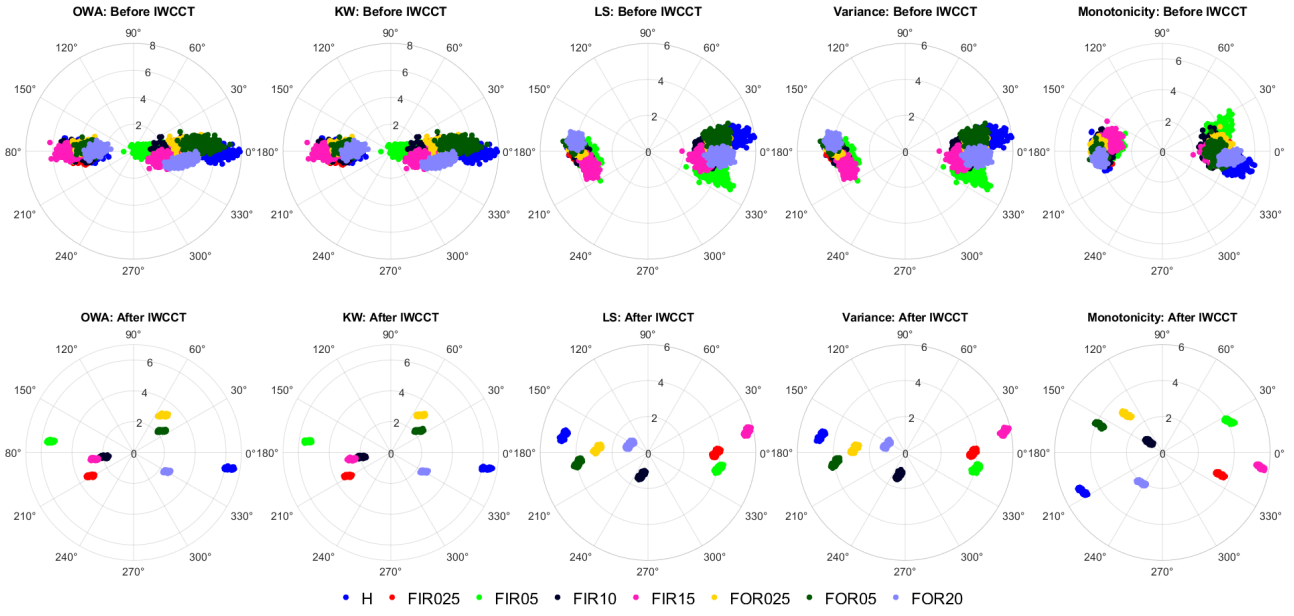


Figure 5.3: PCA embeddings in polar coordinates using Hjorth parameters. The columns correspond to the ranking criteria (OWA, KW, LS, variance, and monotonicity). Top row: PCA before IWCCT; bottom row: PCA after IWCCT. The samples are colored by class (H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20). The $(\mathrm{PC1}, \mathrm{PC2})$ plane is rendered in polar coordinates (r, θ); IWCCT yields more compact, better-separated clusters.

but could drop below 0.25 for Hjorth features. CDF-TS enhances some settings (e.g., PCA/IF ~ 0.79), yet variability persists in the results. Specialized transforms, such as ICDMDip and DipExt, outperform Box-Cox/RBIG in certain scenarios (e.g., t-SNE/IF ~ 0.81) but remain significantly below IWCCT, particularly for Hjorth features, where the scores can be as low as 0.04.

The IWCCT consistently achieved the highest F1 score. Under PCA (Table 5.7), IWCCT+IF attained scores of ≈ 0.97 across both feature families compared to ≈ 0.78 for the next best baseline (PCA/CDF-TS+IF). With t-SNE, IWCCT with OCSVM/IF maintained scores between 0.96 and 0.98, surpassing strong alternatives (e.g., t-SNE/original/IF ~ 0.98 in isolated cases). Crucially, the IWCCT preserved the robustness across feature families. While PCA/original/OCSVM can reach ~ 0.81 on classical descriptors, it falls below 0.24 on Hjorth features, and IWCCT remains above 0.94 throughout the experiment.

A similar pattern was observed for the supervised classifiers (Table 5.8). Box-Cox, Yeo-Johnson, RBIG, and HPT frequently remain below 0.25–0.35. The original and CDF-TS values often lie between 0.45 and 0.54. In contrast, IWCCT achieves scores of ≥ 0.99 with both SVM and KNN across embeddings and feature sets, indicating that the gains stem from the transformation rather than the choice of the classifier.

A comparative visualization of the transformed feature distributions is shown in Figures 5.4 and 5.5. Figure 5.4 contrasts the PCA, t-SNE, Box-Cox, Yeo-Johnson, HPT, RBIG, CDF-TS, and the proposed IWCCT. Both linear and nonlinear embeddings (PCA and t-SNE, respectively) resulted in only partial separation with significant overlap across the fault categories. Distribution-oriented transforms (Box-Cox, Yeo-Johnson, and HPT) offer slight improvements in compactness, whereas RBIG produces a more dispersed representation of the data. CDF-TS partially aligned the data but failed to form compact clusters. In contrast, the IWCCT generates compact, non-overlapping clusters, creating a representation that is well suited for downstream anomaly detection and classification.

To benchmark the distribution-shaping methods, Figure 5.5 compares the DipExt→PCA, ICDMDip→PCA, and IWCCT. DipExt→PCA and ICDMDip→PCA enhance compactness compared with plain PCA, yet noticeable overlap persists between the healthy and faulty categories. The IWCCT achieved clear interclass separation with dense, well-defined clusters across all fault types, indicating a superior enhancement in fault separability under non-stationary operating conditions.

5.3.9 Sensitivity Analysis of IWCCT Hyperparameters

A sensitivity analysis was conducted to assess the impact of each IWCCT hyperparameter on the diagnostic performance using the MOIRA–UNIMORE bearing dataset (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a). The analysis utilized the *Monotonicity*-ranked feature subset and the binary label pair $\{H, \text{FIR025}\}$, representing healthy and small inner–race fault conditions, respectively. The OCSVM detector was trained on healthy samples and tested on both classes, with the resulting F1-score reported for each parameter configuration. Initially, all parameters were jointly tuned to determine their optimal configuration, followed by a one-at-a-time perturbation. Each parameter was varied across its full admissible range, whereas the remaining parameters were held constant at the tuned values. The parameters σ, η, τ, and α were varied from 0.01 to 0.9 in uniform steps, whereas the iteration depth R was varied from

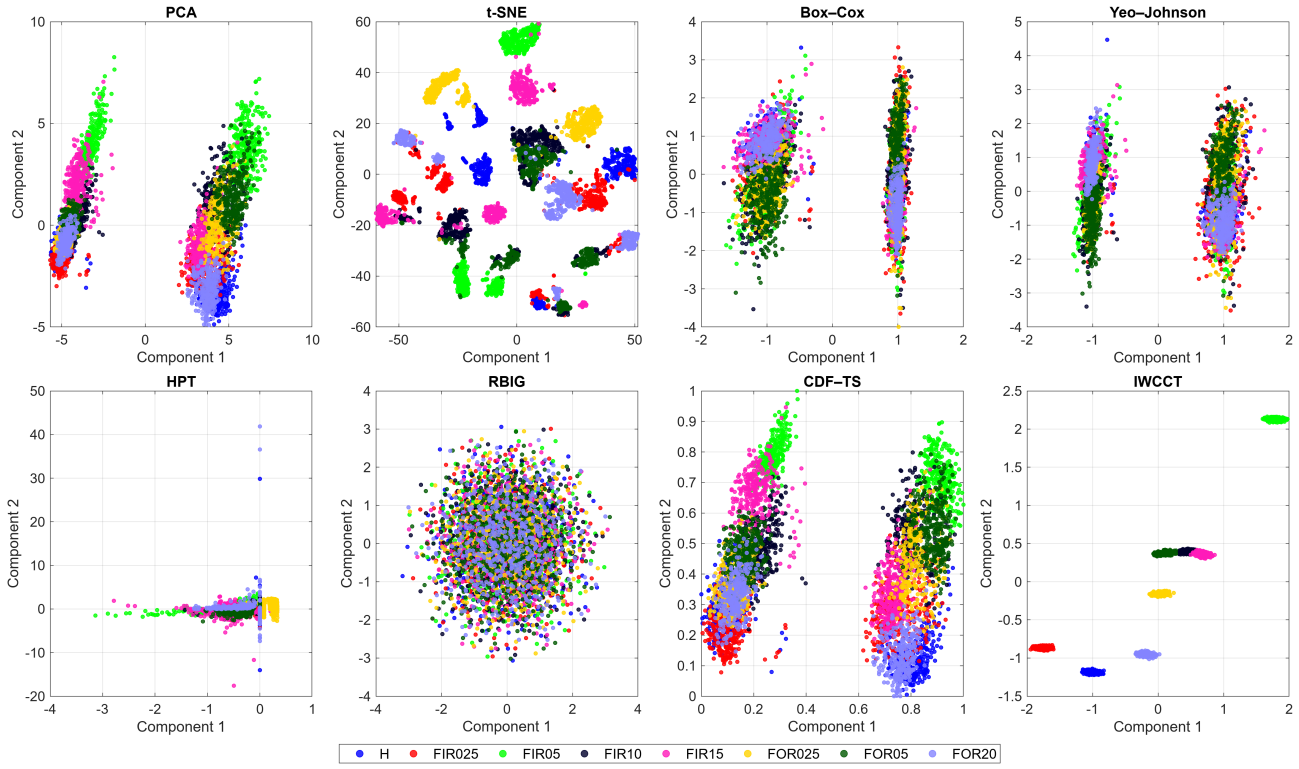


Figure 5.4: Comparative 2-D embeddings across baseline and transform methods. Scatter plots of component 1 vs. component 2 for PCA, t-SNE, Box–Cox, Yeo–Johnson, HPT, RBIG, CDF–TS, and IWCCT applied to the same feature set. The points are colored according to class (H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20). Visually, IWCCT yields the most compact, well-separated clusters among methods.

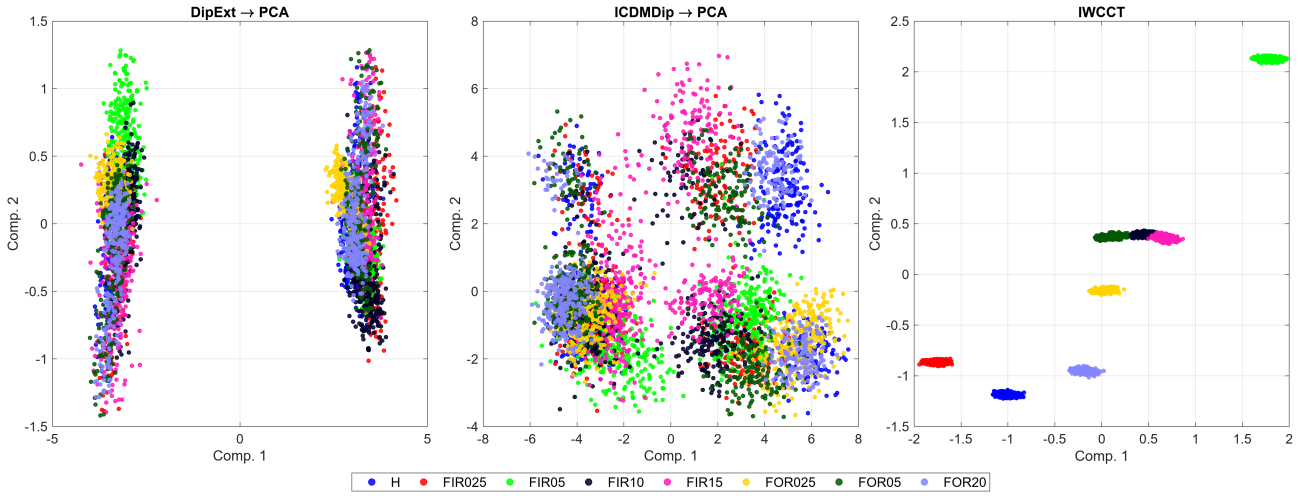


Figure 5.5: Embeddings for DipExt→PCA, ICDMDip→PCA, and IWCCT. Each panel shows component1 vs.component2 from the same dataset: DipExt followed by PCA (left), ICDMDip followed by PCA (middle), and IWCCT (right). The points are colored by class (H, FIR025, FIR05, FIR10, FIR15, FOR025, FOR05, and FOR20). DipExt→PCA and ICDMDip→PCA yield only partial compaction with residual overlap, whereas IWCCT produces dense, well-separated clusters across all categories.

1 to 14. Figure 5.6 illustrates the corresponding F1-score trends. The results indicate that

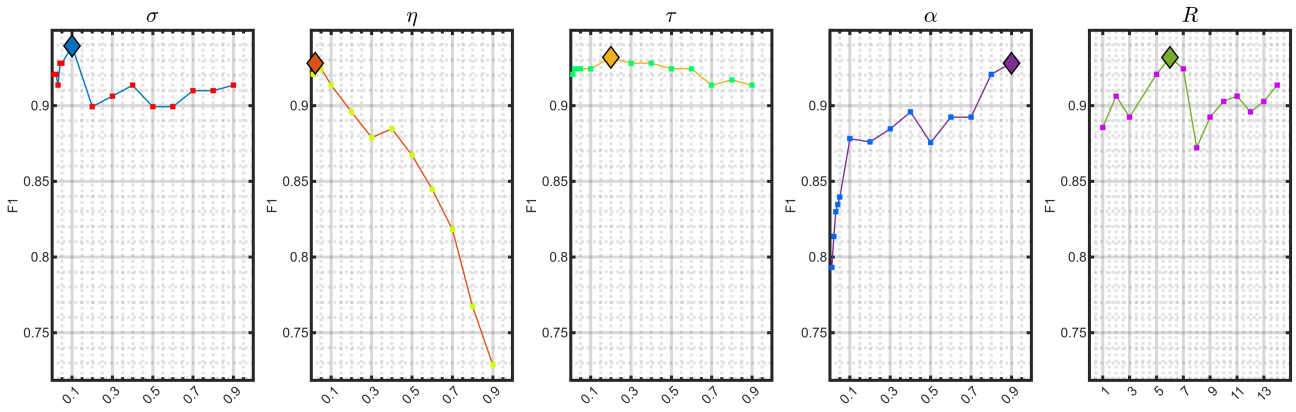


Figure 5.6: Sensitivity analysis of IWCCT hyperparameters on the *Monotonicity*-ranked subset of the MOIRA–UNIMORE bearing dataset. The One-Class SVM (OCSVM) was trained on healthy samples and tested between the labels {H, FIR025}. Each curve shows the variation in the F1-score when a single hyperparameter is perturbed, whereas all others remain fixed at their tuned values. All parameters were varied within the range 0.01–0.9, except for the iteration depth R, which was varied between 1 and 14.

(i) moderate σ values provide the best balance between local compactness and global stability, and (ii) increasing η beyond its optimum rapidly degrades the compactness and reduces separability. (iii) τ exerts limited influence once compaction stabilizes, consistent with its role as a stability gate, rather than a dominant control parameter. (iv) α regulates the contraction intensity, with mid-range values yielding the most consistent improvements. (v) Only a few refinement rounds (R=3–5) are sufficient for convergence, beyond which the F1-score saturates. These observations confirm that the IWCCT maintains a robust diagnostic performance across a wide parameter range. As expected for data-driven transformations, the precise optimal settings may vary slightly across datasets because of differences in scale, noise level, and interclass dispersion; however, the overall sensitivity trends remain consistent.

5.3.10 Validation on the Politecnico di Torino Dataset

The proposed framework was further validated using the *Politecnico di Torino* bearing dataset (Daga et al., 2019). This publicly accessible dataset was derived from a specialized bearing test bench equipped with two triaxial IEPE accelerometers, resulting in six channels positioned at bearings B1 and B3. The signals were synchronously acquired using an OROS OR38 data acquisition system that employed 24-bit delta–sigma converters at a sampling frequency of 51.2 kHz. Each record consisted of 10 s of vibration data, equating to 512,000 samples per channel, with no multiplexing applied to maintain channel coherence. The dataset comprised one healthy bearing and six faulted bearings, which were initially labeled 0A–6A. Each defect was introduced via a conical indentation using a Rockwell indenter, applied either on the inner race or on a single roller. The fault diameters were 150, 250, and 450 μm, resulting in the classes summarized in Table 5.6. For consistency with the MOIRA–UNIMORE nomenclature, the faults were renamed as H, FIR450, FIR250, FIR150, FR450, FR250, and FR150, where "F" denotes a fault, "IR" an inner-race defect, and "R" 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 the same procedure, commencing at 0–N and 100 Hz, with the load and speed incrementally increased until steady state conditions were achieved. A total of 119 data files were recorded (17 per class across the load and speed combinations). Each

file contained six vibration channels corresponding to the two triaxial sensors. For the present analysis, the data were grouped according to the load level, yielding seven independent data subsets for the analysis. The proposed IWCCT framework was applied to each subset, and the results were compared before and after transformation.

Table 5.6: Correlation of the Politecnico di Torino fault nomenclature with the labels and defect specifications utilized in the current 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 PCA-based embeddings shown in Figure 5.7–5.9 show the influence of the proposed IWCCT transformation across three representative load levels: 0, 1000, and 1400 N. In each scenario, a comparative analysis was conducted between the original PCA representation and the IWCCT-transformed feature space using classical statistical descriptors. The analysis identified the following key trends.

- **0 N Load (Figure 5.7):** Under zero-load conditions, the vibration response was minimal, rendering fault-related signatures less discernible in the original PCA space. Upon the application of the IWCCT, the features formed compact, non-overlapping clusters, significantly enhancing the separability of early stage faults (FIR150, FIR250) from the healthy class.

- **1000 N Load (Figure 5.8):** At moderate load levels, there is an increase in contact stiffness and nonlinear vibration interactions. The transformation stabilizes the class boundaries by reducing the dispersion and emphasizing the local density structure, ensuring that the fault clusters remain distinct even as the vibration energy increases.

- **1400 N Load (Figure 5.9):** At higher load levels, the signal becomes highly complex. Despite this complexity, the transformed PCA space maintains distinct fault clusters with minimal overlap between roller and inner-race faults, confirming the robustness and adaptability of IWCCT under varying mechanical stress.

Overall, the consistent improvement across all load cases demonstrates that the IWCCT generalizes effectively beyond the MOIRA–UNIMORE dataset, preserving separability and compactness across different experimental setups, sensor placements, and acquisition systems.

5.3.11 Confusion Matrix and Error Analysis

A confusion matrix-based error analysis was conducted to evaluate the diagnostic performance of the proposed IWCCT and visualize class-wise fault separability both before and after transformation. This analysis employed the *monotonicity* -ranked feature subset and a multiclass linear SVM classifier with an 80/20% train/test split. The evaluation was performed on two independent datasets, MOIRA–UNIMORE and *Politecnico di Torino*, to assess the generalization

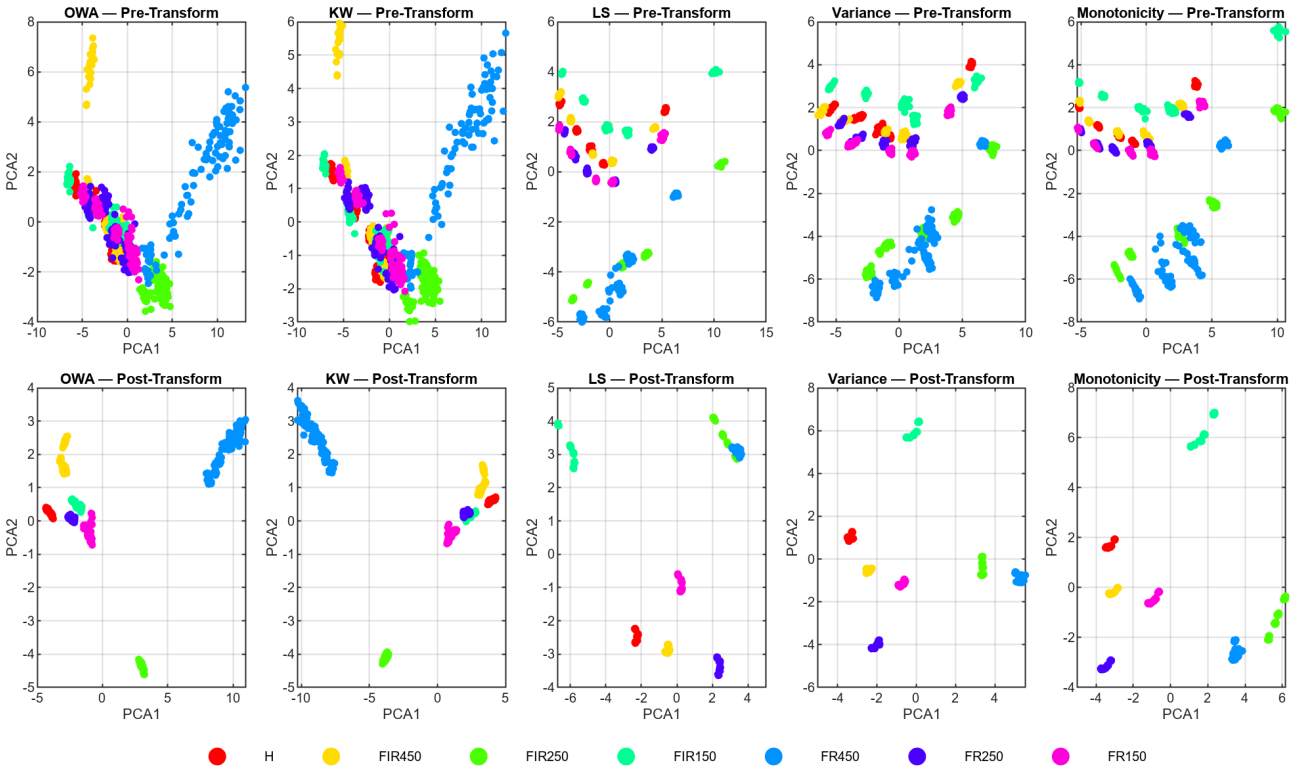


Figure 5.7: The principal component analysis (PCA) embedding of classical statistical features derived from the Politecnico di Torino dataset at a nominal static load of 0 N is presented. The plots illustrate a comparison of the pre- and post-IWCCT feature distributions across various fault types (H, FIR450, FIR250, FIR150, FR450, FR250, and FR150). The transformation results in visibly tighter, non-overlapping clusters.

capability of the IWCCT across diverse acquisition conditions.

- **MOIRA–UNIMORE dataset (Figure 5.10):** Prior to the transformation, there was a notable overlap among several fault categories, particularly between small inner-race faults (FIR025 and FIR05) and outer-race faults (FOR025 and FOR05), resulting in scattered off-diagonal responses. However, following the application of the IWCCT, the confusion matrix demonstrated strong diagonal dominance for nearly all classes, indicating compact and well-separated feature clusters. The sole exception was the FIR10 category, which achieved a lower recognition accuracy of approximately 8.8%, suggesting a partial overlap with adjacent inner-race fault distributions. Overall, the transformation significantly enhanced class separability and reduced cross-fault confusion, confirming that the IWCCT effectively improved the discriminative structure of the feature space without supervision.

- **Politecnico di Torino dataset (Figure 5.11):** Similar patterns were identified in this independent dataset, which was obtained using distinct mechanical and sensor configurations. Prior to transformation, several fault types, particularly those between the inner and roller-fault categories (e.g., FIR150 and FR150), were misclassified owing to overlapping feature distributions. Following the application of the IWCCT, the confusion matrix predominantly exhibited a diagonal structure, indicating a high classification accuracy across most classes. Only two categories demonstrated residual misclassification: FR250 with 16.7% correct recognition and FIR150 with 75%, reflecting the inherent

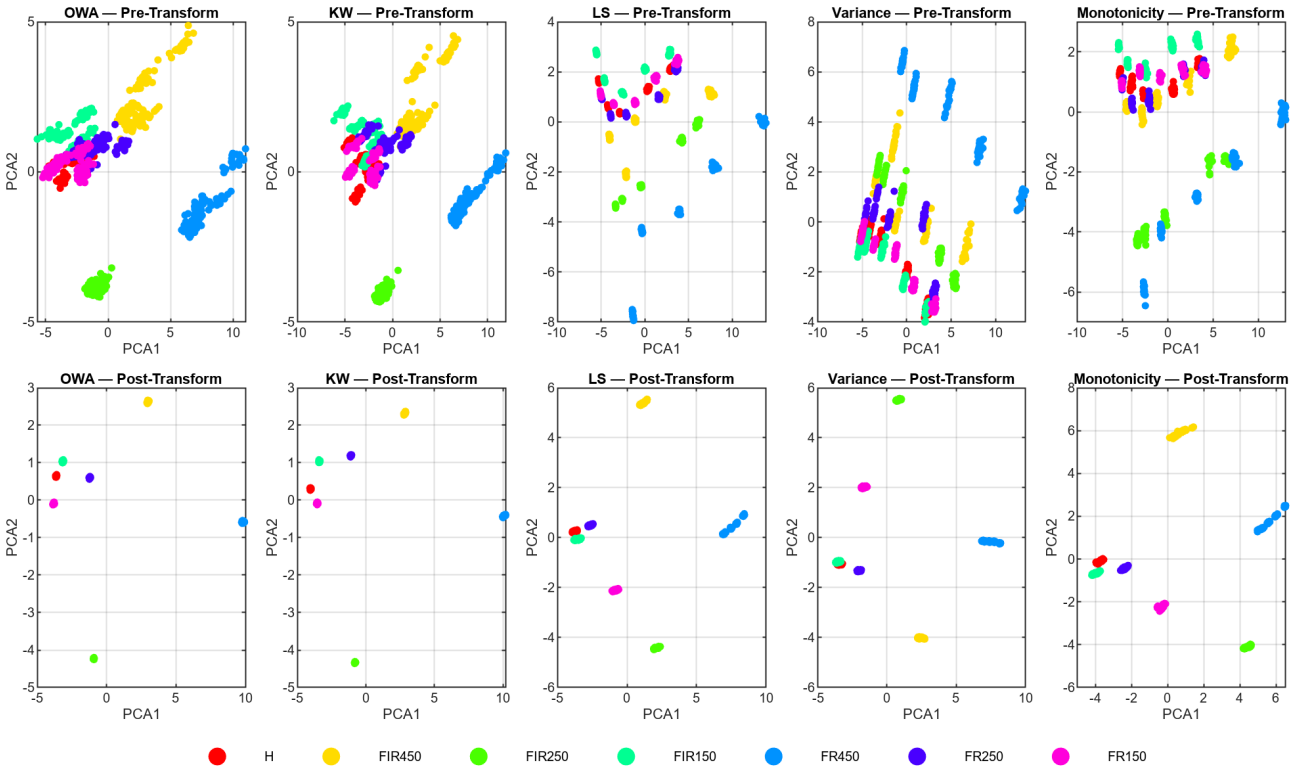


Figure 5.8: PCA embedding of classical statistical features from the *Politecnico di Torino* bearing dataset at a nominal static load of 1000 N. Despite the increased vibration coupling under moderate loading conditions, the IWCCT-transformed feature space maintains compact and separable fault clusters that can be distinguished from one another. This transformation outperforms the original PCA representation and confirms consistent fault separability across varying load conditions.

similarity in their vibration responses under a 900 N nominal static load. Overall, the post-transformation results indicate that the IWCCT effectively maintains strong fault separability and compactness, even under variable loads and operating conditions.

These observations substantiate that the IWCCT consistently enhances interclass separability and compactness across various datasets while maintaining robustness against acquisition variability and mechanical stress.

5.3.12 Robustness to Additive Noise

In practical applications, vibration measurements are inevitably influenced by environmental and instrumental noise. To assess robustness, three representative noise families were introduced exclusively into the vibration channels while maintaining the drive/command signals unchanged, and subsequently reran the entire pipeline (transform → reduction → clustering) in accordance with standard practices in condition monitoring (Jablonski, 2021). A detailed description of the specifications is provided in subsection 4.4.4.

5.3.12.1 Results under noise

This subsection delineates the results obtained from the *statistical descriptor–based features* (time/frequency-domain descriptors extracted per window). Unless specified otherwise, the

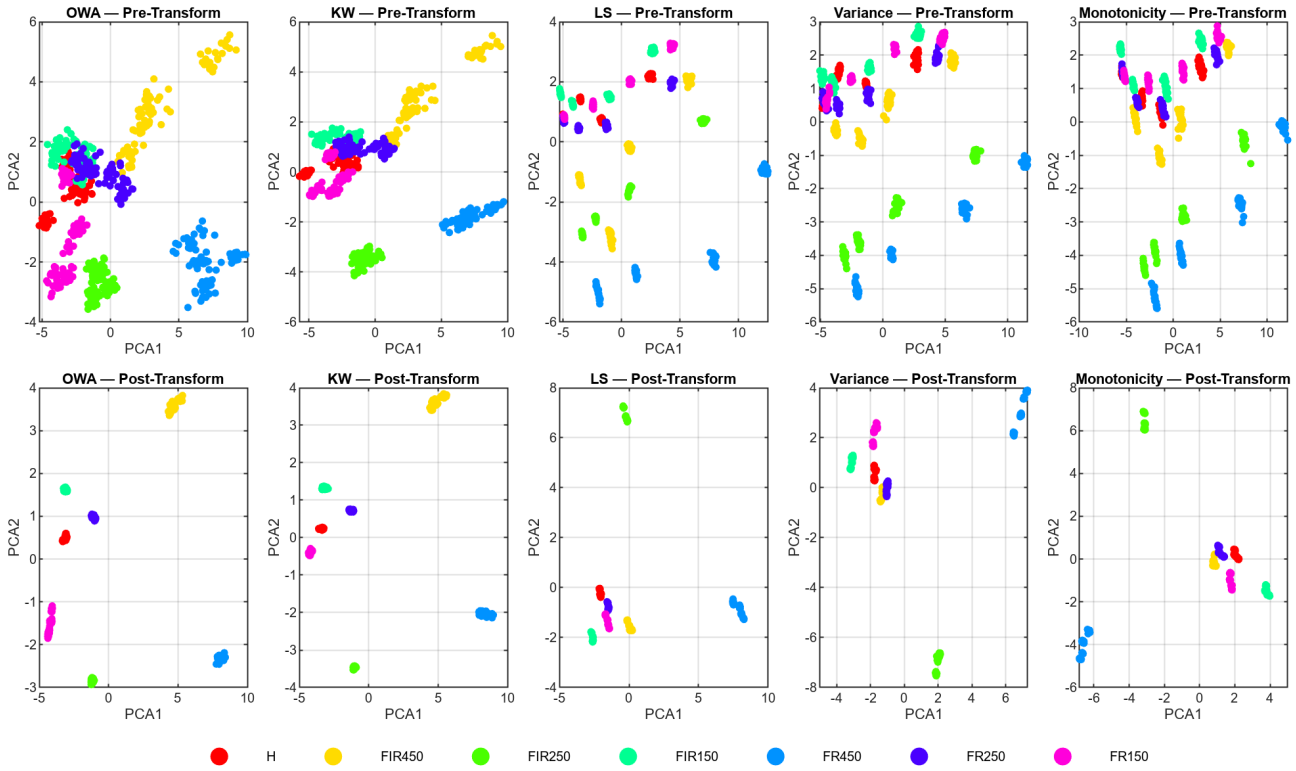


Figure 5.9: The PCA embedding of classical statistical features derived from the *Politecnico di Torino* bearing dataset at a nominal static load of 1400 N demonstrates that, under increased mechanical stress, the IWCCT-transformed feature space maintains compact, stable, and distinctly separated fault clusters. This observation affirms the robustness of the transformation across varying load conditions.

IWCCT hyperparameters correspond to those inTable 5.2, except for the number of outer rounds, which was increased from R=3 to R=5 to mitigate noise-induced dispersion while maintaining the linear $O(ND)$ profile.

Across all three noise families (Figures 5.12–5.14), the "Before IWCCT" rows exhibit anticipated degradations—greater angular spread in KW, inflated radii in LS/Variance, and loss of radial ordering in Monotonicity—whereas the "After IWCCT" rows consolidate each class into a small, well-separated locus.

AWGN + Uniform: As illustrated in Figure 5.12, broadband perturbation disperses the pre-transform clouds and thickens the radial bands in the LS/variance. Post-IWCCT, each class contracts to a tight point, the angular spread in KW approaches zero, and the variance/monotonicity returns to a compact and class-specific range. Visual separation was effectively restored, indicating strong robustness to moderate broadband noise.

Narrowband AM: In Figure 5.13, band-limited interference introduces directionality (anisotropic smearing) and ring-like structures before the IWCCT. The transform suppresses these anisotropies, and the post-transform clusters are compact with a small residual angular bias, demonstrating that local compaction effectively manages coherent tonal leakage.

Structural Noise: Figure 5.14 shows the largest pre-transform distortion (broad rings and class overlap). After the IWCCT, the clusters remained separated, albeit with slightly larger

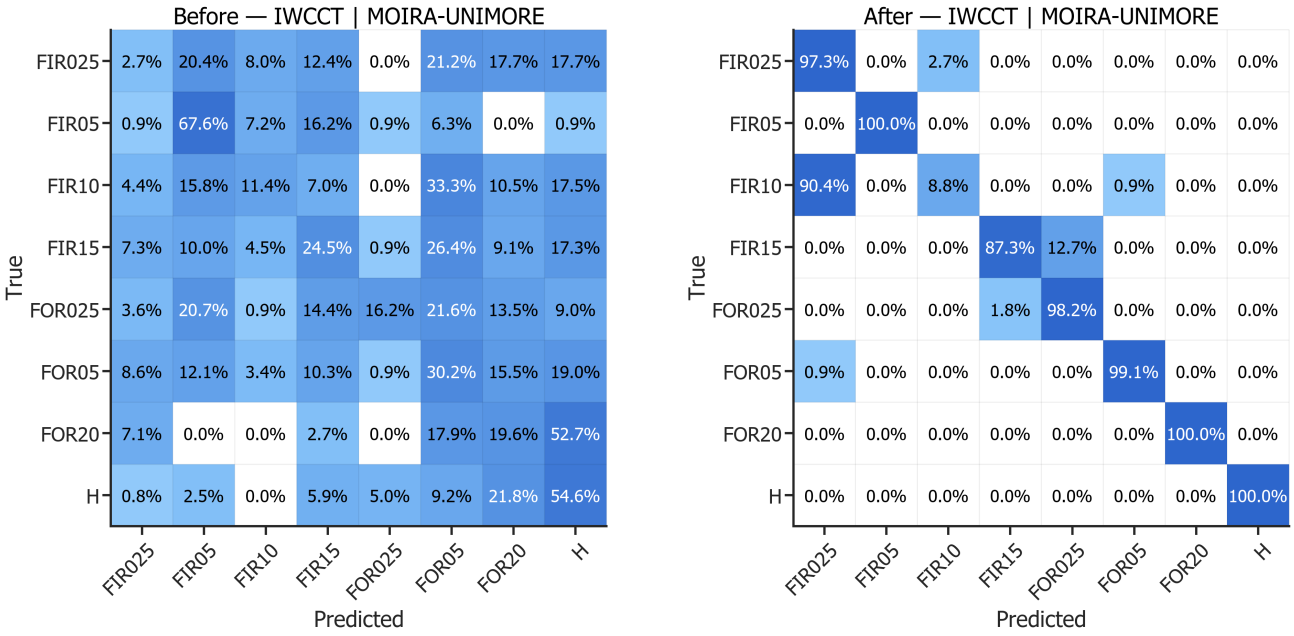


Figure 5.10: Normalized confusion matrices, both before and after the IWCCT transformation, are presented for the *Monotonicity*-ranked feature set derived from the MOIRA–UNIMORE bearing dataset. The classifier utilized was a multiclass linear SVM, employing an 80/20% train/test split. The matrix following the transformation exhibits pronounced diagonal dominance, indicative of compact and separable class structures.

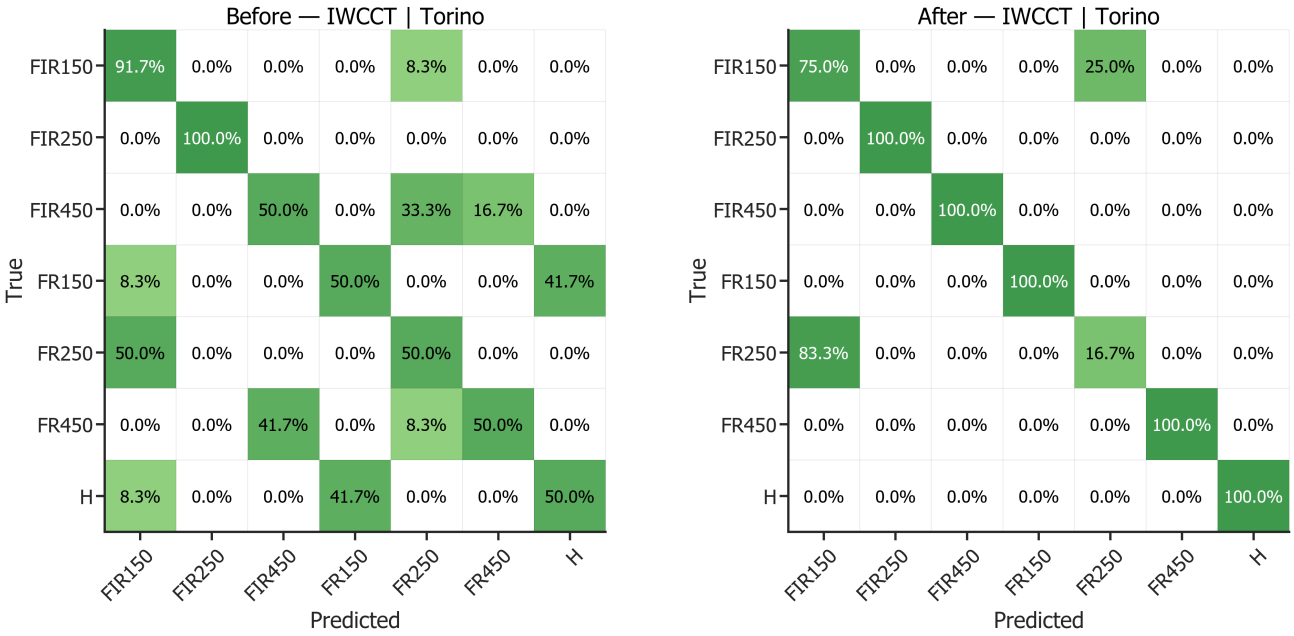


Figure 5.11: Normalized confusion matrices, both before and after the IWCCT transformation, are presented for the *Monotonicity*-ranked feature set derived from the *Politecnico di Torino* bearing dataset under a nominal static load of 900 N. The results following the transformation exhibit minimal cross-class confusion, thereby confirming the robustness of the IWCCT method.

radii than those in the AWGN case, which is consistent with the stronger low-frequency energy from the FRF shape.

Collectively, Figures 5.12–5.14 demonstrate that the IWCCT maintains its compacting behavior

under additive noise, recovering clear class separation in all five polar diagnostics (OWA, KW, LS, Variance, Monotonicity) with minimal visual degradation compared to the clean case.

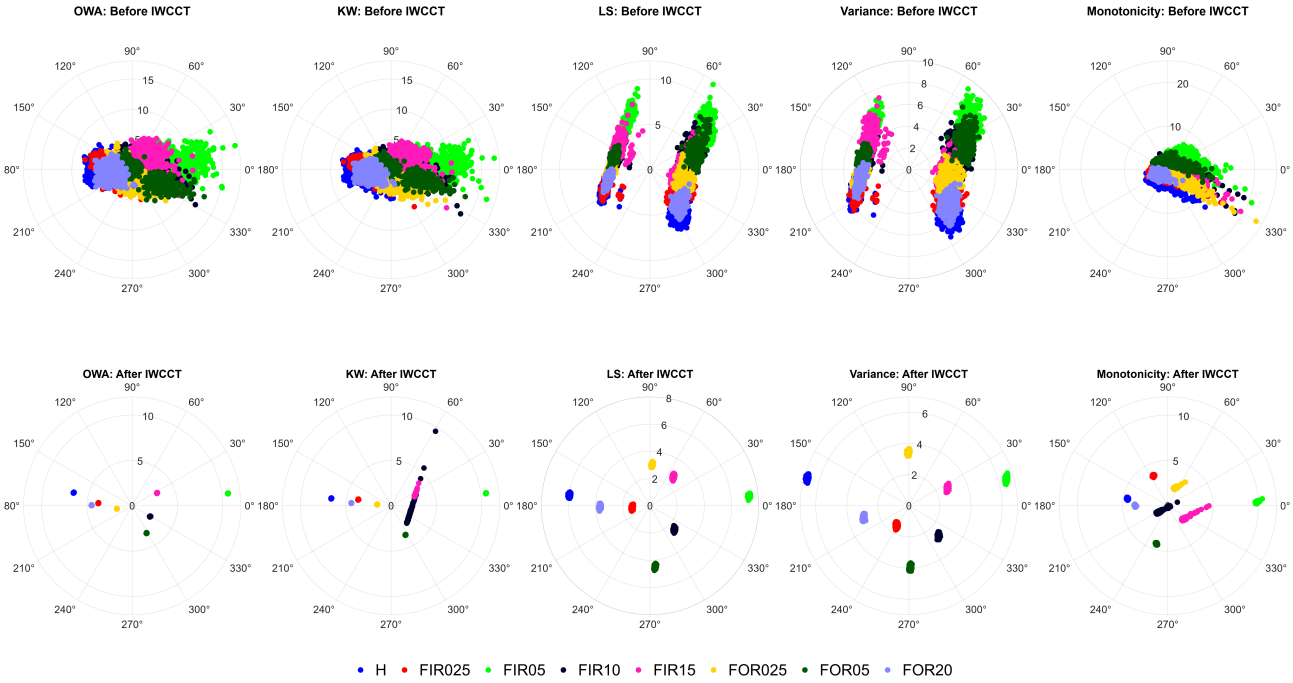


Figure 5.12: Effect of broadband noise (AWGN at 10 dB plus 2% uniform jitter) on polar diagnostics before (top row) and after (bottom row) the IWCCT. The columns show the OWA, KW, LS, Variance, and Monotonicity. Noise inflates angular/radial spread pre-transform; IWCCT reduces each class to a compact locus with distinct angle/radius, restoring visual separability.

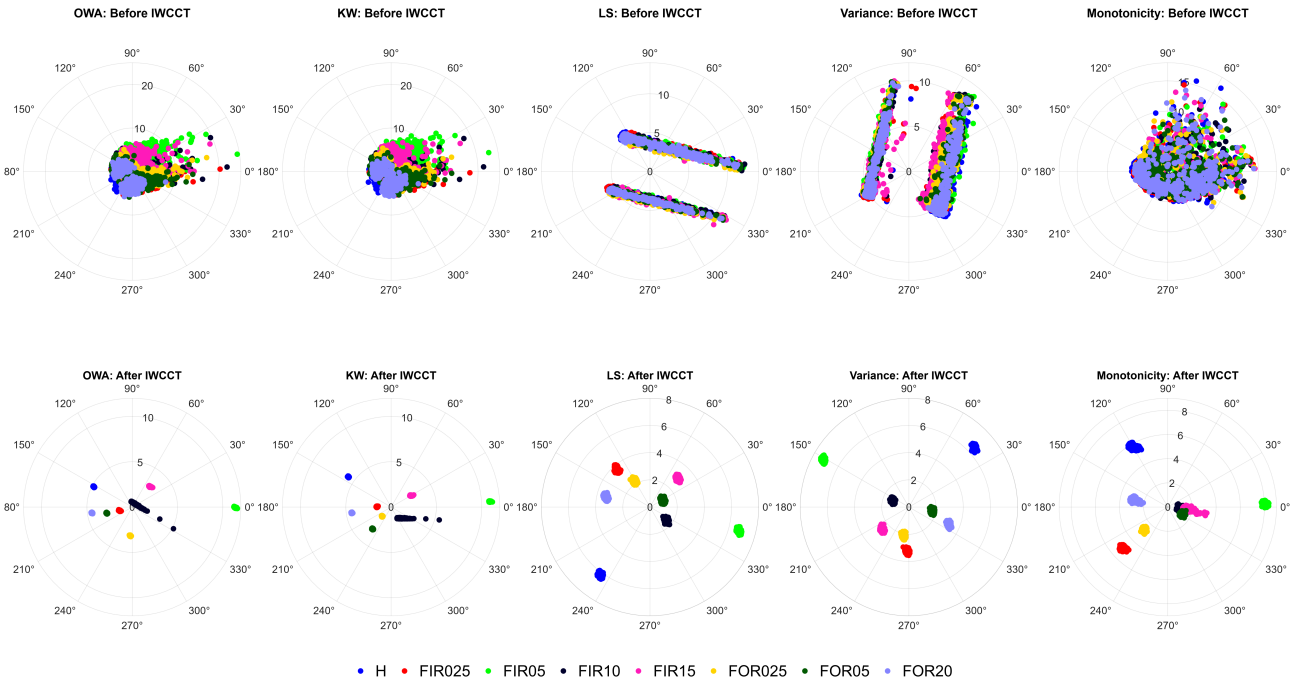


Figure 5.13: Effect of narrowband, amplitude-modulated interference (band-pass $[20, 40]$ Hz, pulse rate f_p=0.5 Hz). Before IWCCT, anisotropic smearing and ring structures appear; after IWCCT, clusters are again compact with small residual angular bias, indicating robustness to coherent tonal leakage.

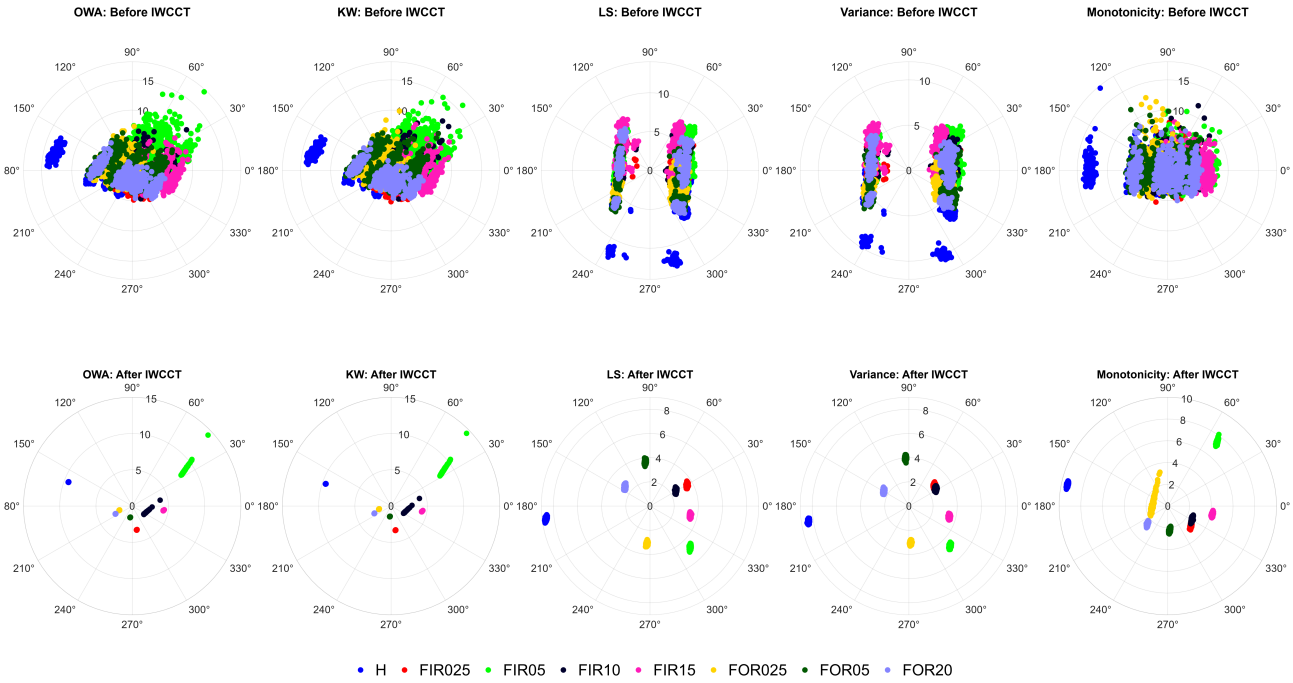


Figure 5.14: Effect of structural noise synthesized from a pseudo-FRF profile. The strongest pre-transform distortion (broad rings/overlap) is observed here; IWCCT still compacts each class to a separated locus, with modestly larger radii than in the broadband case.

Table 5.7: F1 Scores across embedding, transformation, and detector configurations. For each feature family (left block: classical descriptors; right block: Hjorth parameters), the columns correspond to the feature ranking criteria (OWA, KW, LS, variance, and monotonicity). Rows sweep embeddings (PCA, t-SNE) and transforms (Original, Box–Cox, Yeo–Johnson, HPT, RBIG, CDF–TS, IWCCT;. For DipExt and ICDMDip, the detectors are applied directly to the transformed features in the original space—no dimensionality reduction (no PCA or t-SNE) is performed.

Embedding	Transformation	Detector	Classical descriptors					Hjorth parameters				
			OWA	KW	LS	Var	Mono	OWA	KW	LS	Var	Mono
PCA	Original	OCSVM	0.8247	0.7745	0.7349	0.8122	0.5000	0.2363	0.2134	0.5551	0.6374	0.4728
		IF	0.8102	0.7711	0.7840	0.8338	0.5268	0.6963	0.7189	0.7880	0.7558	0.7587
	Box–Cox	OCSVM	0.0776	0.0731	0.4648	0.6193	0.0817	0.1189	0.0865	0.2393	0.4769	0.3514
		IF	0.1072	0.0960	0.5494	0.6843	0.13318	0.2461	0.2439	0.5414	0.5042	0.4729
	Yeo–Johnson	OCSVM	0.0730	0.0869	0.4828	0.6078	0.1020	0.1166	0.0820	0.2165	0.5080	0.3242
		IF	0.1145	0.1027	0.5757	0.7084	0.1526	0.3040	0.3125	0.5346	0.5300	0.4796
	HPT	OCSVM	0.1452	0.1187	0.5642	0.2878	0.1291	0.4660	0.3825	0.2011	0.2590	0.3340
		IF	0.1556	0.1430	0.6385	0.4794	0.2342	0.4409	0.4440	0.5391	0.4427	0.4166
	RBIG	OCSVM	0.0940	0.1099	0.1239	0.0910	0.0842	0.1007	0.0844	0.0998	0.1050	0.0935
		IF	0.1120	0.0916	0.0982	0.1003	0.1269	0.1033	0.0751	0.1140	0.1200	0.1266
	CDF–TS	OCSVM	0.8333	0.8021	0.7608	0.8229	0.5849	0.2647	0.1870	0.5490	0.6286	0.5406
		IF	0.8228	0.7907	0.7927	0.8482	0.6029	0.6314	0.7513	0.7824	0.7636	0.7814
	IWCCT	OCSVM	0.9752	0.9843	0.9782	0.9758	0.9734	0.9806	0.9734	0.9794	0.9800	0.9758
		IF	0.9722	0.9704	0.9758	0.9782	0.9758	0.9776	0.9734	0.9752	0.9764	0.9740
t-SNE	Original	OCSVM	0.9021	0.8835	0.5851	0.8406	0.6915	0.6512	0.5811	0.5339	0.3270	0.4071
		IF	0.9044	0.8727	0.9917	0.9807	0.7163	0.6925	0.7271	0.8190	0.8045	0.7699
	Box–Cox	OCSVM	0.3389	0.4591	0.3385	0.5854	0.3138	0.1480	0.1597	0.3049	0.1664	0.2464
		IF	0.4160	0.6234	0.8713	0.9014	0.3399	0.4734	0.4659	0.5801	0.7353	0.7638
	Yeo–Johnson	OCSVM	0.2748	0.4080	0.4621	0.6129	0.2823	0.2371	0.1747	0.2663	0.1579	0.2375
		IF	0.2726	0.5652	0.8530	0.9168	0.2915	0.5727	0.5361	0.5163	0.6880	0.8056
	HPT	OCSVM	0.3811	0.0372	0.1445	0.5605	0.0703	0.3493	0.2989	0.2033	0.5037	0.1947
		IF	0.4745	0.1120	0.2232	0.9240	0.0960	0.5316	0.2378	0.3513	0.7055	0.4311
	RBIG	OCSVM	0.1096	0.1028	0.0955	0.1068	0.0865	0.1006	0.0682	0.0891	0.0847	0.1093
		IF	0.1163	0.1134	0.0658	0.0731	0.1050	0.1275	0.1458	0.1096	0.1289	0.1481
	CDF–TS	OCSVM	0.8708	0.7695	0.8701	0.7739	0.7178	0.5701	0.5269	0.5382	0.2766	0.3738
		IF	0.8987	0.8680	0.9945	0.9962	0.7325	0.7651	0.7460	0.8377	0.8059	0.7765
	IWCCT	OCSVM	0.9709	0.9764	0.9807	0.9704	0.9788	0.9734	0.9764	0.9722	0.9782	0.9758
		IF	0.9650	0.9758	0.9704	0.9674	0.9682	0.9656	0.9722	0.9716	0.9704	0.9752
	ICDMDip	OCSVM	0.8978	0.8832	0.5417	0.6789	0.6244	0.2535	0.2177	0.6315	0.6863	0.4891
		IF	0.8972	0.8642	0.7822	0.8145	0.5241	0.7478	0.7363	0.7920	0.7686	0.7660
	DipExt	OCSVM	0.8695	0.8788	0.3825	0.3396	0.2384	0.6093	0.5250	0.0658	0.0419	0.0393
		IF	0.8691	0.8398	0.7763	0.7722	0.2040	0.6655	0.6588	0.6315	0.6532	0.6277

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

Table 5.8: F1 Scores across embedding, transformation, and classifier configurations. For each feature family (left block: classical descriptors; right block: Hjorth parameters), the columns correspond to the feature ranking criteria (OWA, KW, LS, variance, and monotonicity). Rows sweep embeddings (PCA, t-SNE) and transforms (Original, Box–Cox, Yeo–Johnson, HPT, RBIG, CDF–TS, IWCCT). *DipExt and ICDMDip are evaluated in the original feature space and passed directly to the classifier (no PCA or t-SNE).*

Embedding	Transformation	Classifier	Classical descriptors					Hjorth parameters				
			OWA	KW	LS	Var	Mono	OWA	KW	LS	Var	Mono
PCA	Original	SVM	0.4787	0.4641	0.5361	0.5035	0.2683	0.5324	0.5364	0.5253	0.5048	0.4487
		KNN	0.4717	0.4433	0.5245	0.5053	0.2413	0.5084	0.5648	0.5018	0.4850	0.4398
	Box–Cox	SVM	0.1887	0.1615	0.2571	0.2494	0.1401	0.2305	0.2107	0.1969	0.2072	0.1814
		KNN	0.2049	0.1652	0.2525	0.2439	0.1401	0.2278	0.2009	0.2168	0.2072	0.1729
	Yeo–Johnson	SVM	0.2031	0.1744	0.2666	0.2550	0.1325	0.2319	0.2454	0.2015	0.2124	0.1775
		KNN	0.2058	0.1623	0.2574	0.2396	0.1402	0.2103	0.2179	0.2161	0.2009	0.1675
	HPT	SVM	0.2034	0.1861	0.3234	0.3324	0.1543	0.4372	0.4564	0.2712	0.2607	0.2500
		KNN	0.2101	0.1834	0.3322	0.3354	0.1618	0.3941	0.3808	0.3265	0.2841	0.2583
	RBIG	SVM	0.1006	0.0927	0.0966	0.0816	0.0651	0.0799	0.1053	0.0877	0.0953	0.0906
		KNN	0.0863	0.0788	0.0912	0.0828	0.0582	0.0833	0.0909	0.0925	0.0972	0.0838
	CDF–TS	SVM	0.4706	0.4699	0.5407	0.5057	0.2811	0.5262	0.5319	0.5249	0.5056	0.4443
		KNN	0.4758	0.4503	0.5292	0.4859	0.2472	0.5043	0.5538	0.5005	0.4960	0.4400
	IWCCT	SVM	1.0000	1.0000	0.9955	0.9677	0.9784	1.0000	1.0000	1.0000	1.0000	0.9933
		KNN	1.0000	1.0000	0.9955	0.9677	0.9774	1.0000	1.0000	1.0000	1.0000	0.9889
t-SNE	Original	SVM	0.8539	0.8697	0.9364	0.9033	0.5886	0.5369	0.5460	0.5256	0.5055	0.4443
		KNN	0.8586	0.8779	0.9431	0.9087	0.5579	0.5017	0.5311	0.5084	0.4918	0.4420
	Box–Cox	SVM	0.4691	0.5874	0.8601	0.8141	0.4298	0.5702	0.5940	0.5612	0.5175	0.5573
		KNN	0.4633	0.5383	0.8563	0.7969	0.4004	0.5563	0.5682	0.5403	0.5257	0.5597
	Yeo–Johnson	SVM	0.5022	0.5887	0.8345	0.8036	0.4227	0.5726	0.6377	0.5595	0.5516	0.5853
		KNN	0.4733	0.5411	0.8132	0.8062	0.3567	0.5594	0.6054	0.5463	0.5431	0.5402
	HPT	SVM	0.5257	0.6217	0.7651	0.7994	0.4533	0.5751	0.6771	0.5663	0.6060	0.6063
		KNN	0.5027	0.5611	0.7567	0.7480	0.4358	0.5492	0.6606	0.5172	0.5595	0.5928
	RBIG	SVM	0.1304	0.1342	0.1421	0.1283	0.1531	0.1323	0.1622	0.1324	0.1565	0.1507
		KNN	0.1170	0.1292	0.1226	0.1102	0.1315	0.1394	0.1534	0.1201	0.1443	0.1392
	CDF–TS	SVM	0.8560	0.8758	0.9173	0.9440	0.6376	0.5359	0.5680	0.5316	0.5044	0.4453
		KNN	0.8619	0.8645	0.9226	0.9379	0.6249	0.5006	0.5362	0.5083	0.4892	0.4331
	IWCCT	SVM	0.9684	1.0000	1.0000	1.0000	0.9819	1.0000	1.0000	1.0000	1.0000	1.0000
		KNN	0.9717	1.0000	1.0000	1.0000	0.9819	1.0000	1.0000	1.0000	1.0000	1.0000
	ICDMDip	SVM	0.8439	0.9212	0.8447	0.8504	0.5481	0.5402	0.5563	0.5462	0.5062	0.4935
		KNN	0.8187	0.8820	0.7771	0.7639	0.4220	0.5019	0.5349	0.5032	0.4884	0.4365
	DipExt	SVM	0.9296	0.9387	0.9479	0.9398	0.7297	0.5306	0.5521	0.5328	0.5040	0.4751
		KNN	0.8829	0.8989	0.8896	0.8745	0.5811	0.5203	0.5345	0.4947	0.4861	0.4309

Notes: OWA: one-way ANOVA; KW: Kruskal–Wallis; LS: Laplacian Score; Var: variance; Mono: monotonicity Embeddings: PCA = Principal Component Analysis; t-SNE = t-distributed Stochastic Neighbor Embedding. Classifiers: SVM = Support Vector Machine; KNN = k-nearest neighbors. The values are the F1 scores (0–1).

Chapter 6

Intelligent Time-Frequency Feature Embedding and Reshaping

This chapter introduces a comprehensive unsupervised diagnostic framework that incorporates two key components: (i) Theil-Aware Intelligent Windowing, an innovative strategy based on the Theil index, which autonomously selects spectrogram window sizes to optimize the time–frequency resolution of nonstationary vibration signals, and (ii) Distribution Reshaping via Bimodal Transformation, a novel feature-level transformation that converts irregular or overlapping manifolds into compact, uniformly dense, and cluster-friendly representations. Extensive experiments conducted on the MOIRA–UNIMORE bearing dataset demonstrated that the proposed framework significantly enhanced class compactness, reduced the nonlinear manifold curvature, and restored separability under variable speed and transient regimes. Comparative evaluations against established distribution reshaping techniques, including Box–Cox, Yeo–Johnson, HPT, RBIG, PIT-GMM, PIT-KDE, and CDF-TS, indicate that the proposed transformation achieves markedly superior cluster compaction and class separability. Robustness tests revealed that the framework remained stable under broadband sensor noise, narrowband amplitude-modulated interference, and structurally shaped pseudo-FRF (pseudoresponse function) noise. Cross-dataset validation on the Case Western Reserve University and Politecnico di Torino bearing datasets further confirmed the generalizability of the approach, with substantial improvements in the F1-scores for both t-SNE and UMAP-reduced features. The results demonstrate that combining adaptive time–frequency representation with principled distribution reshaping offers a powerful and computationally efficient solution for unsupervised bearing fault diagnosis under nonstationary conditions.

6.1 Methodology

This section delineates the proposed vibration-based fault diagnosis framework for independent cart systems (ICS). The workflow comprises three stages (Figure 6.1): (i) Theil-Aware Intelligent Windowing (TAIW) is employed to select 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) (Gove et al., 2022; Jung et al., 2024; van der Maaten & Hinton, 2008) or uniform manifold approximation and projection (UMAP) (Healy & McInnes,

2024; McInnes et al., 2018; McInnes et al., 2020); and (iii) Distribution Reshaping via Bimodal Transformation (DRBT) is utilized to enhance cluster compactness and separation, thereby facilitating downstream detection and classification (see Algorithm 4 for the end-to-end implementation).

Algorithm 4 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 $\mathcal{D}$ (e.g., OCSVM/IF), train/test split rule

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

1: **Stage I (TAIW):** Compute $(w^\star, \text{ov}^\star) \leftarrow \arg\max_{(w,\text{ov})} Q_{tf}(w, \text{ov})$ using Algorithm 5
2: **Stage II (TF features):** Compute STFT with $(w^\star, \text{ov}^\star)$ and obtain power spectrogram $P_{w^\star}$
3: Form frame-level feature matrix $Z \in \mathbb{R}^{N\times d}$ from $P_{w^\star}$ (each time frame is one sample)
4: Compute 2-D embedding $X \leftarrow \Phi(Z) \in \mathbb{R}^{N\times 2}$
5: **Stage III (DRBT):** Compute reshaped embedding $X_{\text{transformed}} \leftarrow \text{DRBT}(X; \Theta_{\text{DRBT}})$ using Algorithm 6
6: Split $X_{\text{transformed}}$ into training/testing according to the experimental protocol
7: Train detector $\mathcal{D}$ using training data (healthy-only for one-class settings)
8: Compute anomaly scores / predicted labels on test data using $\mathcal{D}$
9: Output decisions and compute evaluation metrics (Accuracy/Precision/Recall/F1, etc.)

6.1.1 Implementation Summary

6.1.1.1 Stage I — TAIW.

Given $(\mathcal{W}, \mathcal{O})$, the algorithm systematically examines candidate pairs (w, ov), computes a spectrogram for each pair, and assesses the *Theil-based concentration scores* over time and frequency, denoted as $Q_t(w)$ and $Q_f(w)$, respectively. These scores are then integrated as $Q_{tf}(w) = Q_t(w)\, Q_f(w)$, and the algorithm identifies

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

6.1.1.2 Stage II — Spectrogram and Feature Compression.

The spectrogram is recalculated using the parameters $(w^\star, \text{ov}^\star)$, with each frame considered a sample within the feature space. Subsequently, a 2-D t-SNE or UMAP embedding is computed, which encapsulates the latent manifold of spectrotemporal patterns in a condensed form that is suitable for density analysis.

6.1.1.3 Stage III — DRBT.

A succinct deterministic sequence of operations is executed, including centering/scaling, whitening, radial compaction, and adaptive cluster shift along a bimodal axis. These operations serve to consolidate the clusters, standardize the feature densities, and expand the inter-cluster margins. The entire transformation is characterized by its non-iterative, data-driven, and computationally efficient characteristics.

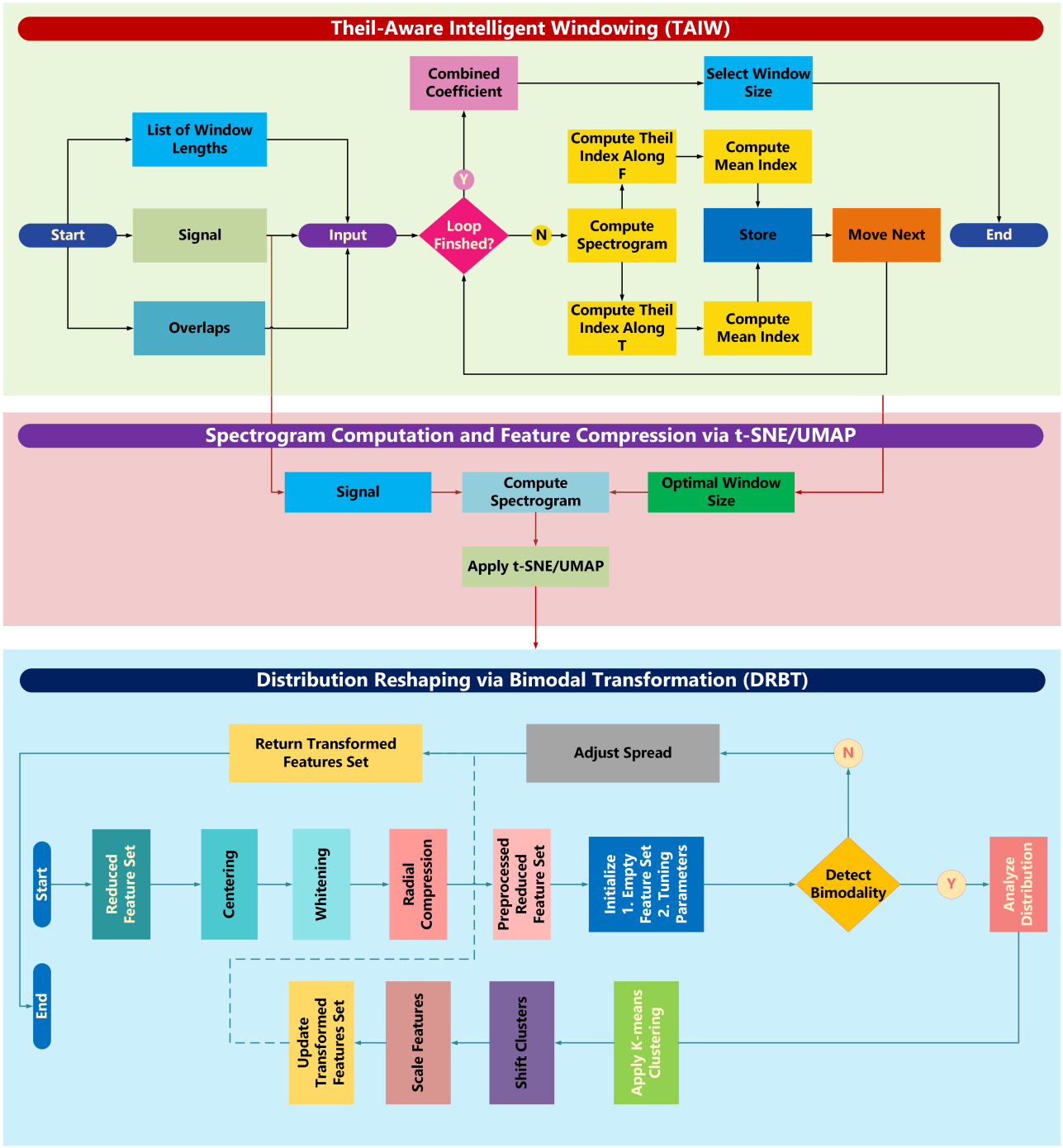


Figure 6.1: The proposed diagnostic framework comprises three stages. Stage 1 involves Theil-Aware Intelligent Windowing (TAIW) for the automatic selection of the window size. Stage 2 encompasses the computation of spectrograms and feature compression. Stage 3 entails Distribution Reshaping via Bimodal Transformation (DRBT) to derive compact, cluster-friendly features, thereby facilitating reliable fault diagnosis in Industrial Control Systems (ICS).

6.1.2 Notation and Conventions

Table 6.1 The text enumerates the symbols and terms employed throughout the document, without the inclusion of equations. Units are specified, where applicable.

Table 6.1: Notation used in §6.1.3 (TAIW) and §6.1.5 (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, $W = V\,\mathrm{diag}(d^{-1/2})$
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

6.1.3 Theil-Aware Intelligent Windowing (TAIW)

The selection of 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 employed to quantify inequality, was used to assess the degree of energy localization within the spectrogram. Higher Theil values indicate a stronger temporal and spectral concentration of fault-related events than lower values do.

6.1.3.1 Industrial Motivation

In industrial environments, condition monitoring systems continuously collect data on vibrations, currents, and system variables from several machines. Although each machine is monitored for only a brief period each day, the cumulative effect of monitoring numerous machines repeatedly daily, weekly, and over several years results in a substantial volume of 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 offering a fully automatic and data-driven criterion for selecting the optimal window parameters, thereby facilitating the efficient large-scale monitoring of industrial fleets.

6.1.3.2 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

$$\mathcal{T}(y) = \frac{1}{M}\sum_{i=1}^{M} \frac{y_i}{\bar{y}} \log\Big(\frac{y_i}{\bar{y}}\Big), \tag{6.1}$$

where the convention $0 \log 0 = 0$ is applied, which is numerically clamped by ε. By applying (6.1) to the columns (time slices) and rows (frequency tracks) of the spectrogram $P_w[k,n] = |S_w[k,n]|^2 + \varepsilon$, we obtain

$$T_w^{(t)}(n) = \mathcal{T}\big(P_w[:,n]\big), \qquad n = 0,\ldots,N_t-1, \tag{6.2}$$

$$T_w^{(f)}(k) = \mathcal{T}\big(P_w[k,:]\big), \qquad k = 0,\ldots,K-1. \tag{6.3}$$

The aggregate per-window scores are calculated as

$$Q_t(w) = \frac{1}{N_t}\sum_{n=0}^{N_t-1} T_w^{(t)}(n), \qquad Q_f(w) = \frac{1}{K}\sum_{k=0}^{K-1} T_w^{(f)}(k), \tag{6.4}$$

and the joint criterion for window selection is defined by

$$Q_{tf}(w) = Q_t(w)\,Q_f(w), \qquad w^\star \in \arg\max_{w\in\mathcal{W}} Q_{tf}(w). \tag{6.5}$$

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

Algorithm 5 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: **end for**
10: Select best window: $w^\star \leftarrow \arg\max_w Q_{tf}(w)$

6.1.3.3 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. 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 (6.5) 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 the axes.

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

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

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

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, while $m_{\max}$ is chosen such that $w_{m_{\max}} \le \lfloor L/4 \rfloor$, with L denoting 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 $\mathcal{W} = \{2^m \mid m_{\min} \le m \le m_{\max}\}$ 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.

6.1.4 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 the local vibration pattern. To achieve dimensionality

reduction and reveal latent structures, a nonlinear embedding technique such as t-SNE or UMAP was employed.

$$X = \Phi(P_{w^\star}) \in \mathbb{R}^{N\times 2}, \tag{6.6}$$

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

6.1.4.1 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 excitation. Because these embeddings preserve the local geometry but not the absolute scale, the resulting two-dimensional manifold frequently displays a non-uniform density or elongated clusters. This observation necessitates subsequent Distribution Reshaping via Bimodal Transformation (DRBT).

6.1.5 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 instead of irregular or elongated clusters.

6.1.5.1 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 by creating uneven cluster shapes and overlapping boundaries. From a diagnostic perspective, this results in ambiguous fault regions and unreliable clustering. DRBT addresses this limitation by implementing an adaptive data-driven reshaping procedure that balances global and local separability.

6.1.5.2 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.

6.1.5.3 Principle and Overview

The transformation process was performed in two distinct 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 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.

6.1.5.4 Step 1: Compactification

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

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad X_c = X - \mathbf{1}\mu^\top. \tag{6.7}$$

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

$$\Sigma = \frac{1}{N-1} X_c^\top X_c, \tag{6.8}$$

and is subsequently decomposed as

$$\Sigma = V\,\mathrm{diag}(d)\,V^\top, \tag{6.9}$$

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

$$W = V\,\mathrm{diag}(d^{-1/2}), \qquad X_w = X_c W, \tag{6.10}$$

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:

$$r_i = \|x_i^{(w)}\|_2, \tag{6.11}$$

$$s_i = \frac{\tanh(r_i)}{r_i + \varepsilon}, \tag{6.12}$$

$$x_i^{(\text{comp})} = s_i\, x_i^{(w)}. \tag{6.13}$$

Consequently, the compactified dataset is represented as

$$X_{\text{compact}} = [x_1^{(\text{comp})};\, x_2^{(\text{comp})};\, \ldots;\, x_N^{(\text{comp})}]. \tag{6.14}$$

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.

6.1.5.5 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

$$\hat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - f_i}{h}\right), \tag{6.15}$$

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

$$\text{prominence}\big(\hat{p}, x^*\big) > p_{\min}.$$

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

Bimodal Case (Reshape-worthy). Upon the detection of bimodality, 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 $\Delta = c_{\text{high}} - c_{\text{low}} > 0$.

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

$$\alpha = \min\left(\gamma,\, \gamma\frac{\Delta}{\sigma}\right), \qquad f_i \leftarrow f_i - \alpha\Delta, \quad \forall i \in C_{\text{high}}, \tag{6.16}$$

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

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

$$\beta = \max\left(\delta,\, \epsilon - \eta\frac{\Delta}{\sigma}\right), \qquad f \leftarrow m + \beta(f - m). \tag{6.17}$$

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:

$$f \leftarrow m + \beta_{\text{nonbimodal}}\frac{(f - m)}{\sigma}. \tag{6.18}$$

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 allow for the maintenance of variance.

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

$$X_{\text{transformed}} = [f^{(1)}_{\text{reshaped}},\, f^{(2)}_{\text{reshaped}}]. \tag{6.19}$$

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

$$\text{Var}(f_{\text{reshaped}}) \leq \beta^2 \,\text{Var}(f), \tag{6.20}$$

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 6 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: $X_{\text{transformed}}$ 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**
    $r_i \leftarrow \|X_w(i,:)\|_2$
    $X_{\text{compact}}(i,:) \leftarrow \tanh(r_i)/(r_i + \varepsilon)\ \cdot\ X_w(i,:)$
end for
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 $\#\{\text{peaks}\} > 1$ **then**
        Partition $f \to C_{\text{low}}, C_{\text{high}}$ (k-means, $k = 2$)
        $\Delta \leftarrow c_{\text{high}} - c_{\text{low}}$
        $\alpha \leftarrow \min\big(\gamma,\ \gamma\Delta/\sigma\big)$
        Shift: $f_i \leftarrow f_i - \alpha\Delta, \quad \forall i \in C_{\text{high}}$
        $\beta \leftarrow \max\big(\delta,\ \epsilon - \eta\Delta/\sigma\big)$
        Contract: $f \leftarrow m + \beta(f - m), \quad m = \text{mean}(f)$
    else
        $f \leftarrow m + \beta_{\text{nonbimodal}}(f - m)/\sigma$
    end if
end for
$X_{\text{transformed}} \leftarrow [f^{(1)}_{\text{reshaped}},\ f^{(2)}_{\text{reshaped}}]$

6.1.6 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.

6.1.6.1 Bounded Shift Fraction (γ)

The cluster-shift operation $f_i \leftarrow f_i - \alpha\Delta$ with

$$\alpha = \gamma \cdot \min\big(1,\, \Delta/\sigma\big), \qquad 0 < \alpha \leq \gamma \tag{6.21}$$

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 the separation.

6.1.6.2 Contraction Gate (δ, ϵ, η)

The subsequent contraction step,

$$\beta = \max\Big(\delta,\, \epsilon - \eta\frac{\Delta}{\sigma}\Big), \qquad f \leftarrow m + \beta(f - m) \tag{6.22}$$

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

$$\eta \leq \epsilon - \delta \tag{6.23}$$

6.1.6.3 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 approach mitigates the occurrence of spurious branching caused by noise or weak oscillations, thereby enhancing robustness under conditions of a low signal-to-noise ratio (SNR).

6.1.6.4 Non-bimodal Contraction

For unimodal features with $\sigma > \sigma_{\text{th}}$, the transformation

$$f \leftarrow m + \beta_{\text{nonbimodal}}\frac{(f - m)}{\sigma} \tag{6.24}$$

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 variance.

6.1.6.5 Invariances

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

Table 6.2: Role of DRBT Hyperparameters and Typical Practical Ranges

Parameter	**Role**	**Typical Range / Effect**
ε	Radial near-origin damping	$[0.01, 0.95]$; larger $\uparrow$ smooths small r
γ	Max shift fraction	$[0.05, 0.95]$; larger $\uparrow$ merges modes faster
δ	Min post-shift contraction	$[0.2, 0.7]$; prevents collapse ($\beta \geq \delta$)
ϵ	Base contraction cap	$(\delta, 1]$; larger $\uparrow$ 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 $\downarrow$ scatter more

6.1.6.6 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.

6.1.6.7 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.

6.1.6.8 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 the operating conditions.

6.1.7 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 bounded geometric operation within the algorithm. Unlike optimization-based approaches, the DRBT does not minimize an explicit objective function; instead, its hyperparameters are selected according to *the structural constraints imposed by the update rules themselves*, rather than empirical tuning toward 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 was 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 to the mean. 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 6.2) ensures numerical stability while maintaining a meaningful scatter correction.

The inter-mode shift parameter γ governs the maximum fraction of inter-cluster separation that can be translated during the 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 collapse. 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 the numerical collapse. The update rule $s = \max\big(\delta,\ \epsilon - \eta \frac{\Delta}{\sigma}\big)$ implies that δ must remain strictly positive, while ϵ must exceed δ to allow adaptive behavior. Selecting δ that is 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 the loss of structural information. Finally, $\beta_{\text{nonbimodal}}$ regulates the unimodal high-scatter regime, where updates driven by bimodality are inapplicable to the data. This parameter implements variance-normalized scaling around the feature mean, maintaining the overall distribution shape while preventing uncontrolled dispersion.

6.2 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) (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a), (ii) the Politecnico di Torino bearing dataset (Daga et al., 2019), and (iii) Case Western Reserve University (CWRU) dataset (CWRU, 2005). Collectively, these datasets provide complementary validation domains encompassing translational–rotational linear drive systems, laboratory test benches, and conventional motor–bearing assemblies.

6.2.1 Data Diversity

Although the MOIRA–UNIMORE dataset includes additional system variables, such as cart currents and position feedback, 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 the nature of 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 (Cavalaglio Camargo Molano et al., 2020; Cocconcelli et al., 2019; Jabbar, D'Elia, & Cocconcelli, 2024; Jabbar et al., 2023; Molano et al., 2020; Molano et al., 2017). This hybrid kinematic behavior of the ICS platform introduces time-varying excitation patterns, making it particularly suitable for evaluating the robustness of the proposed method in nonstationary operating conditions.

6.2.2 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 capability of the proposed framework. All analyses were conducted in MATLAB R2024b on a 12th Gen Intel® Core™ i9–12900HK (2.50 GHz) workstation with 32GB RAM. The performance evaluation was based on cluster compactness, separation, and anomaly detection accuracy using one-class SVM and Isolation Forest detectors.

6.3 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) 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 both stationary and non-stationary operating conditions.

6.3.1 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 (Figures. 6.2 and 6.3). 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 Hz + noise):** six $\sim 80\,\text{ms}$ packets of a 700 Hz tone embedded in low-level noise.
- **Signal 3 (step frequency + noise):** a tone transitioning from 450 Hz to 800 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.

Figure 6.2 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 ($\approx 0.205\,\mathrm{s}$) with ov $= 960$ (15/16 of w), maintaining temporal detail in the slowly evolving ridge and the chirp crossing area. For *Signals 2* and *3*, Q_{tf} prefers a longer window $w = 4096$ ($\approx 0.819\,\mathrm{s}$, 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 Figure 6.3 corroborate 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 any manual adjustment.

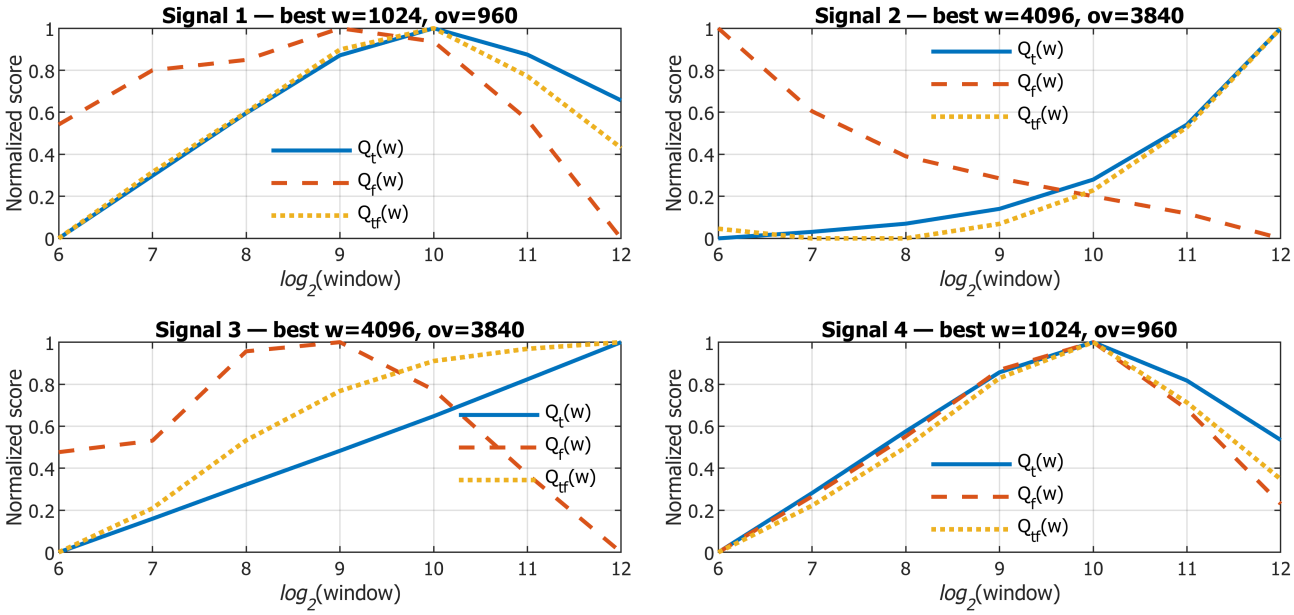


Figure 6.2: TAIW coefficient curves for the four synthetic signals. Normalized $Q_t(w)$ (solid), $Q_f(w)$ (dashed), and $Q_{tf}(w)$ (dotted) versus $\log_2$ window length. Titles indicate the selected $w^\star$ and overlap ov*.

6.3.2 Application of the Proposed Pipeline on Public Datasets

The proposed TAIW–DRBT pipeline was applied to the three datasets introduced in Section 6.2.

6.3.2.1 MOIRA–UNIMORE Dataset (Translational–Rotational System)

Experimental Conditions. The data analyzed correspond to *Experiment Type 2* of the MOIRA–UNIMORE dataset (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a), 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 (Figure 4.11). This section of the rail, highlighted in red, induces coupled translational–rotational motion as the cart navigates the curve, resulting in pronounced nonstationarity 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 Figure 6.4. The cart speed varied between 0 mm/s and approximately 1000 mm/s. Signals were recorded under no fault, inner-race fault,

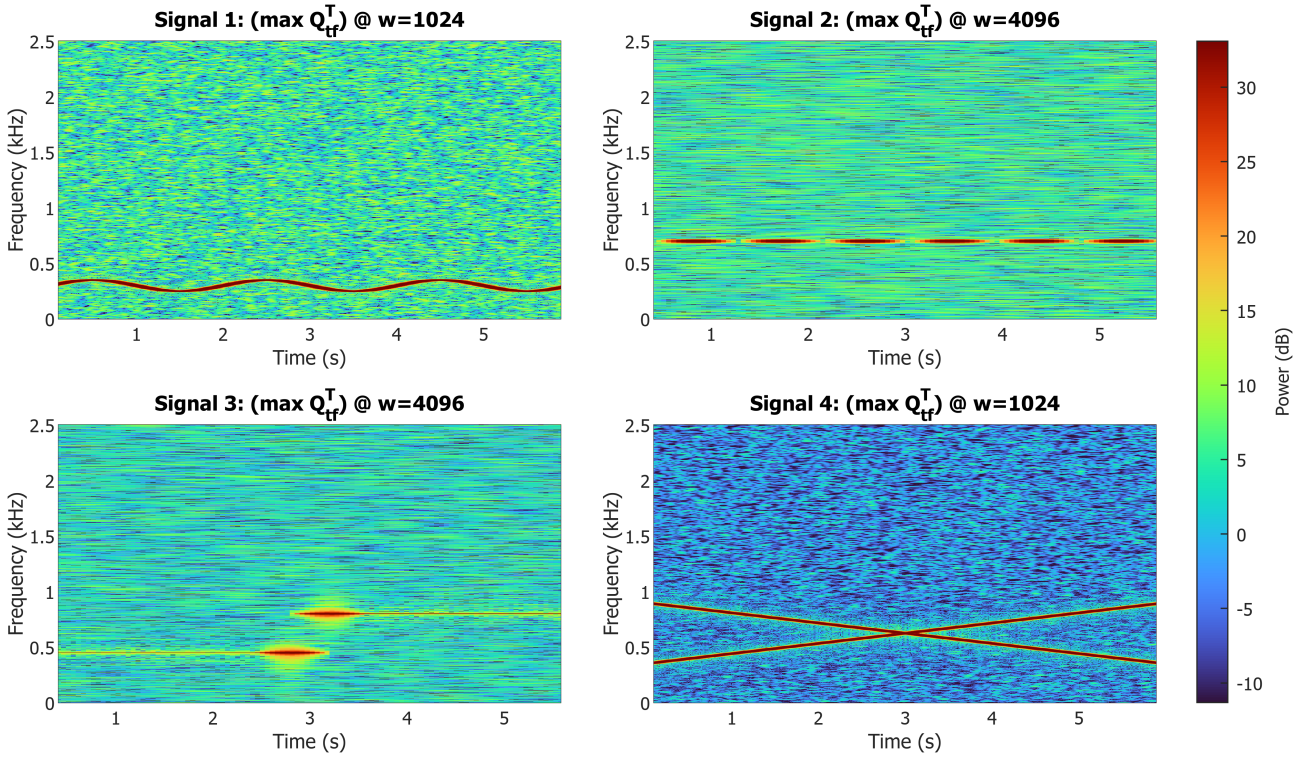


Figure 6.3: Spectrograms computed using the TAIW-selected window and overlap for each synthetic signal.

and outer race fault conditions, with fault severities as described in Table 3.7. Each experiment was repeated for multiple cycles and realizations to ensure the statistical consistency.

Vibration Channels The vibration channels of the MOIRA–UNIMORE bearing dataset correspond to the accelerometer placements shown in Figure 4.11. 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 (Jabbar, Cocconcelli, D'Elia, Borghi, et al., 2025a).

Embedding Analysis and Quantitative Evaluation The t-SNE and UMAP embeddings were calculated for all monoaxial vibration channels and the triaxial sensor located on the right side of the rail, both prior to 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. Figures 6.5–6.7 depict the comparative low-dimensional representations across channels.

In all channels, the pre-transformation embeddings (upper panels in each figure) demonstrated

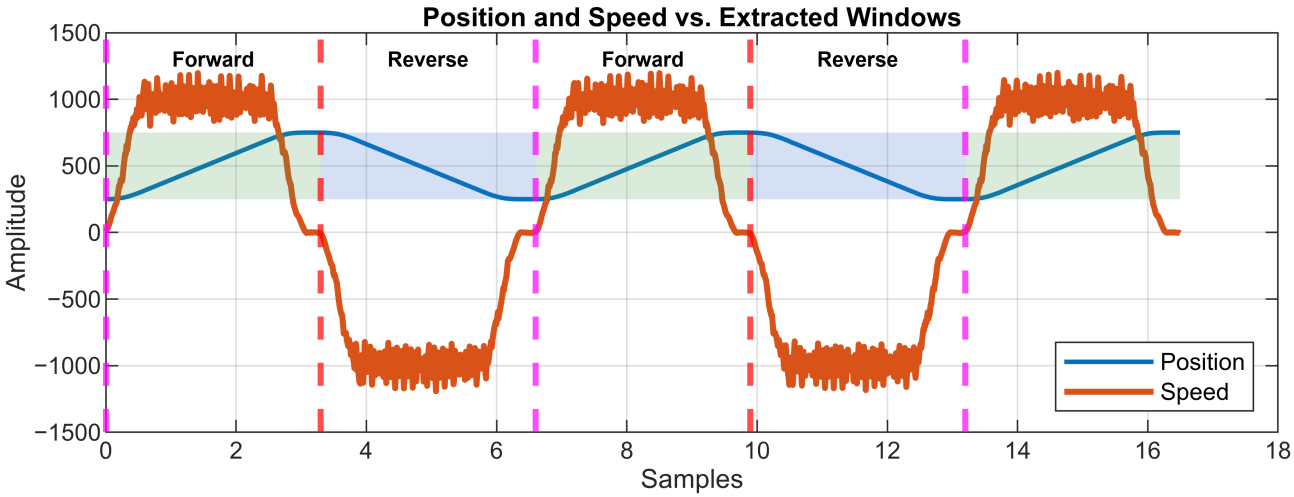


Figure 6.4: The position and velocity profile of the cart during Experiment Type 2 is illustrated, with the forward and reverse phases indicated by shaded regions and dashed lines.

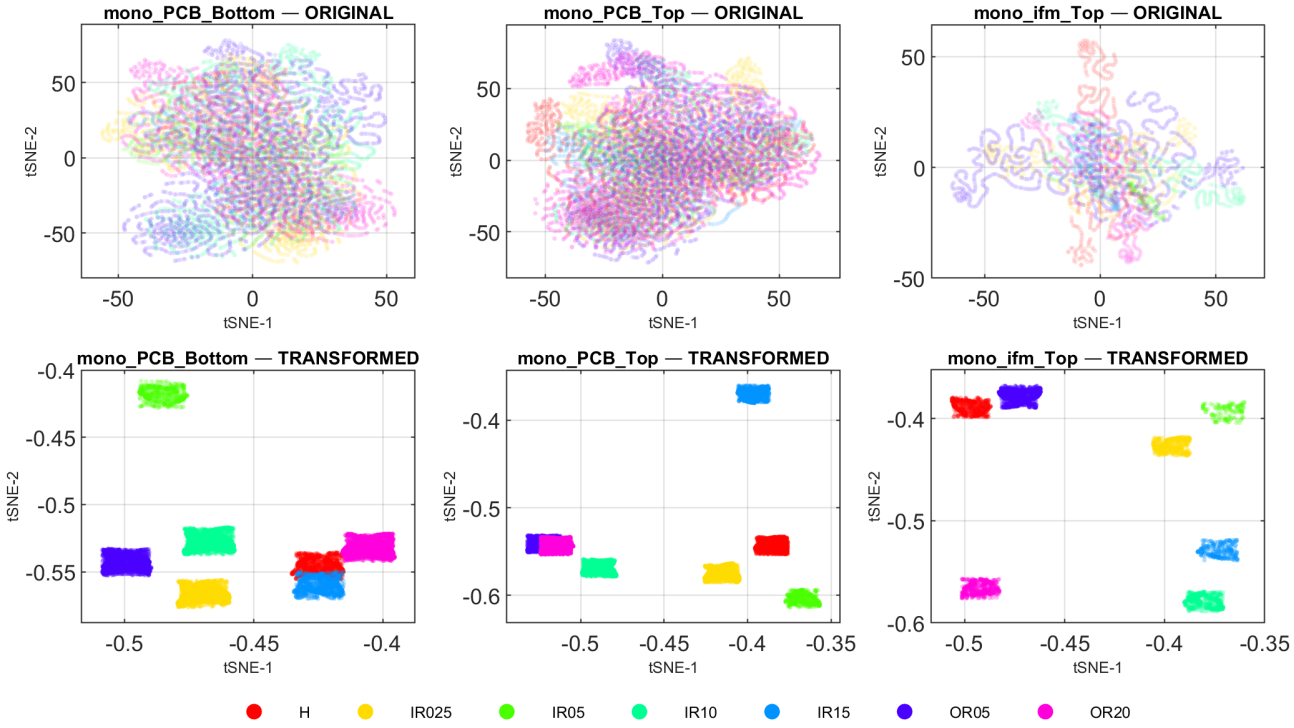


Figure 6.5: The t-SNE embeddings, both prior to and following the DRBT transformation, for mono-axial accelerometer channels (`mono_PCB_Bottom`, `mono_PCB_Top`, and `mono_ifm_Top`) demonstrate significant improvements. Post-transformation, the clusters exhibit enhanced compactness and improved fault separability across all bearing conditions.

a significant overlap between the 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, respectively. 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.

Quantitative validation using one-class SVM (OCSVM) and Isolation Forest (IF) is presented

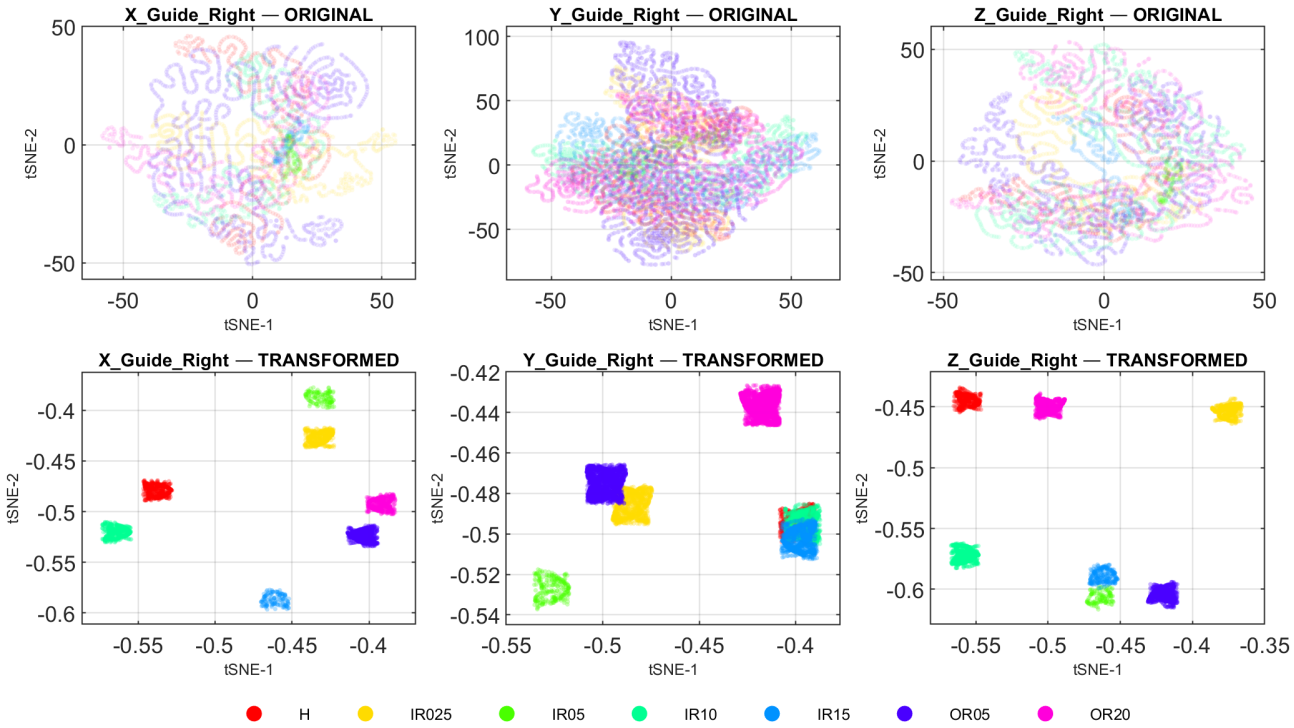


Figure 6.6: The t-SNE embeddings for the right-side tri-axial accelerometer channels (`X_Guide_Right`, `Y_Guide_Right`, and `Z_Guide_Right`) were analyzed both prior to and following the DRBT transformation. The embeddings post-transformation exhibit compact and distinctly separable clusters, effectively replacing the overlapping distributions that were evident in the original feature space.

in Table 6.3 for t-SNE. The post-transformation F1-scores exceeded 0.98 for 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 the speed variation.

Table 6.3: 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

To further evaluate the consistency of the DRBT transformation across various embedding techniques, Table 6.4 presents the results for UMAP-reduced features. Consistent with the t-SNE findings, all channels demonstrated significant performance enhancement 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.

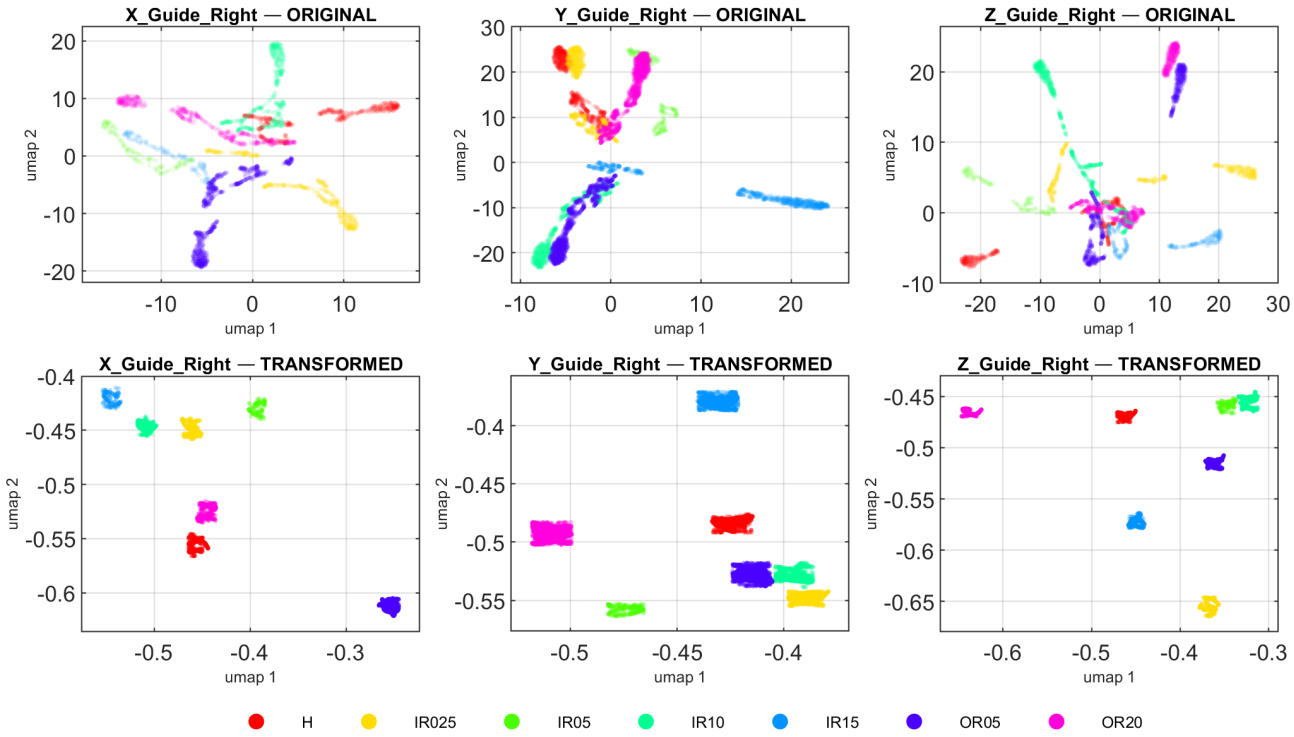


Figure 6.7: The UMAP embeddings for the right-side tri-axial accelerometer channels, both prior to and following the DRBT transformation, demonstrate that the transformation consistently improves inter-class separation and intra-class compactness. This finding confirms the robustness of the transformation across nonlinear embedding techniques.

Table 6.4: 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 **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

6.3.2.2 Politecnico di Torino Dataset (Purely Rotational System)

Experimental Context The proposed framework was further validated using the *Politecnico di Torino* bearing dataset (Daga et al., 2019), a publicly available benchmark designed to evaluate 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 (Daga et al., 2019). 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.

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 either on the inner race or on a single roller. The defect diameters were approximately 150, 250, and 450 µm, as listed in Table 5.6. 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 according to the load level, resulting in seven independent data subsets that were processed through 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. Figures 6.8–6.11 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 corresponding to 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 rotational speed. Notably, the improvement was more pronounced for the roller fault family, where the pre-transformation manifolds exhibited strong nonlinear distortions owing to the contact dynamics.

To further evaluate 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. Figures 6.12–6.15 depict 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) exhibit broad and entangled manifolds, where the fault classes overlap significantly, 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, facilitating the classification.

However, some residual overlap persisted under specific operating conditions. For the `FIR` family, the `FIR150` class partially overlaps with the healthy cluster at 300 Hz and 1400N (Figure 6.13), 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 (Figures 6.14–6.15). 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.

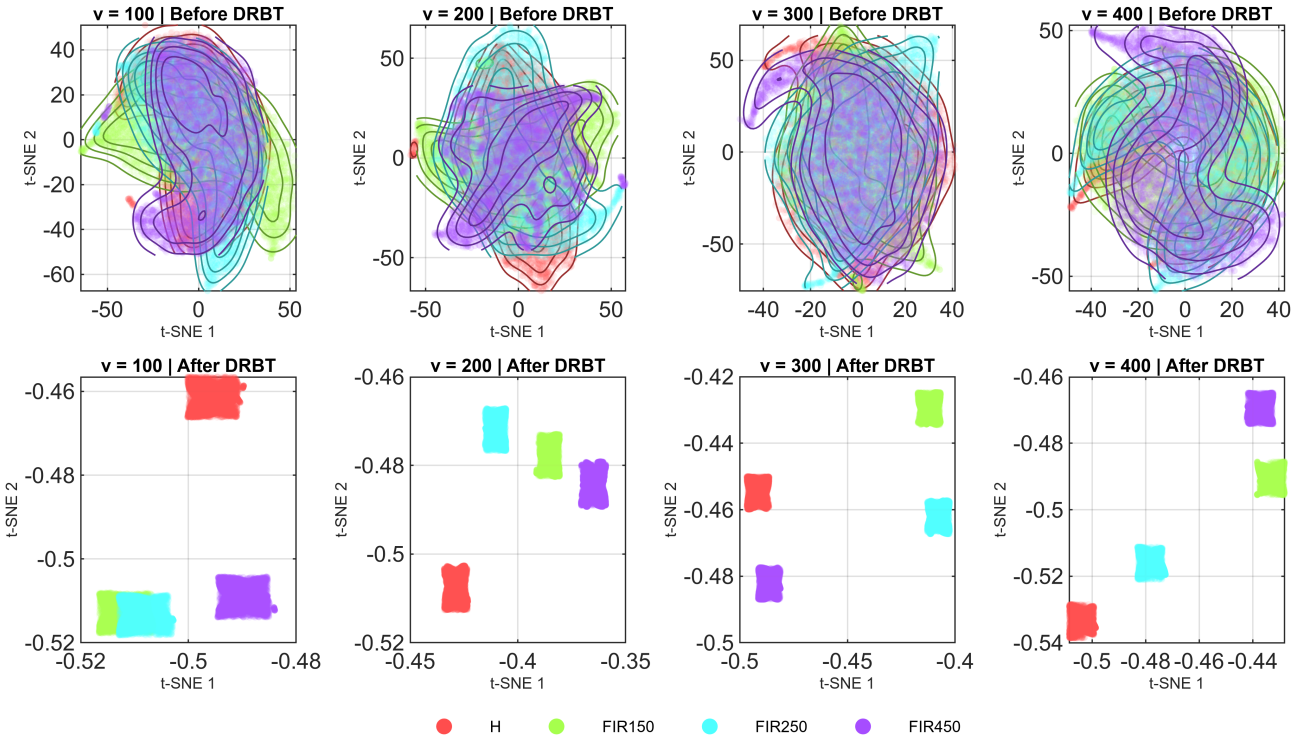


Figure 6.8: The t-SNE embeddings are depicted before (top row) and after (bottom row) the DRBT transformation for the inner-race fault family (`FIR`) at a load level of **1000 N**. The transformation results in compact, linearly separable clusters across increasing shaft speeds (100–400 Hz).

Quantitative Evaluation The results presented in Tables 6.5 and Table 6.6 indicate a consistent and significant enhancement in anomaly detection performance following the application of 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 6.5), the baseline F1-scores prior to 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 0N). This demonstrates that the 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 6.6). Prior to DRBT, there was moderate to high variability among speeds and loads, with OCSVM values ranging from 0.30 to 0.83 and IF values ranging from 0.40 to 0.83. Following 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 manifold visualizations.

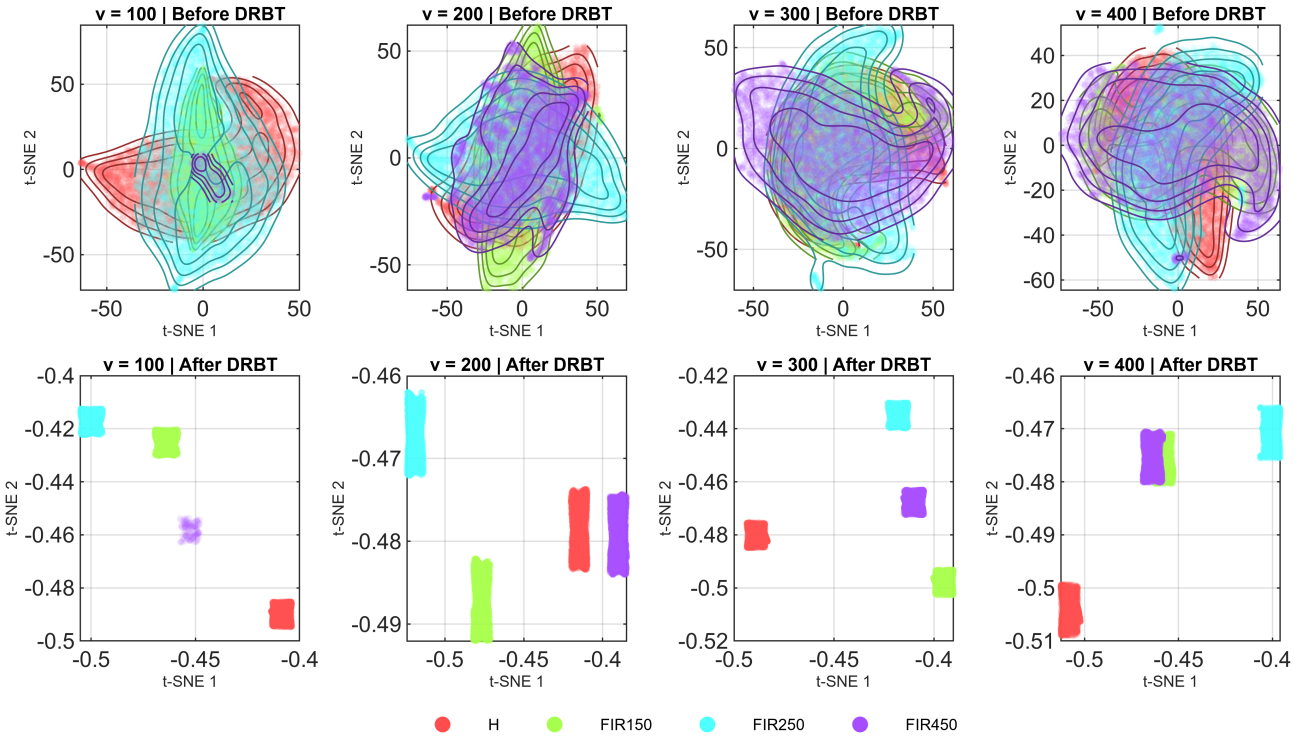


Figure 6.9: The t-SNE embeddings, both prior to and following the DRBT transformation, for the inner-race fault family (`FIR`) at a load level of **1400 N**, demonstrate that the DRBT effectively reduces intra-class spread and enhances the separation among fault severities (`FIR150-FIR450`).

Table 6.5: The F1 scores related to speed and load for OCSVM and IF on **t-SNE**-reduced features are presented both prior to Before and following After 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

6.3.2.3 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 (Smith & Randall, 2015), who systematically evaluated the entire CWRU dataset and classified each case according to its diagnoses. In this benchmark, the dataset was divided into six categories—Y1, Y2, P1, P2, N1, and N2—based on the clarity of the fault manifestation in the time and frequency domains. Specifically, Y1 and Y2 correspond to fully diagnosable

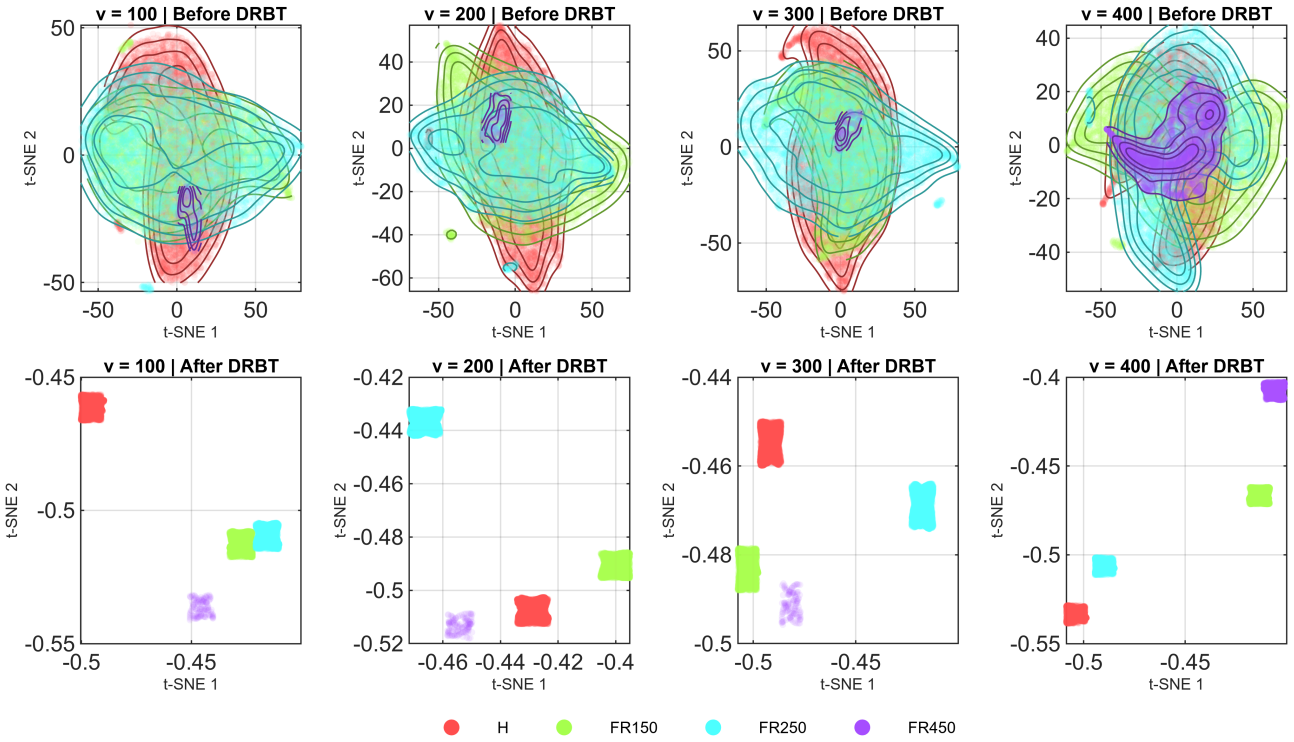


Figure 6.10: The t-SNE embeddings are depicted before (top) and after (bottom) the DRBT transformation for the roller fault family (FR) at a load level of **1000 N**. This transformation effectively compacts the fault clusters and rectifies local manifold distortions observed in the pre-transformation space.

Table 6.6: The F1 scores related to speed and load for OCSVM and IF on **UMAP**-reduced features are presented both prior to Before and following After 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

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 diagnosable owing to confounding mechanical effects or are indistinguishable from noise, containing only weak harmonic traces in the spectra envelope. Importantly, the authors of the tutorial (Smith & Randall, 2015) 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 (Smith & Randall, 2015), this study focuses exclusively on the N1 and N2 subsets, which correspond to the most noise-dominated and diagnostically

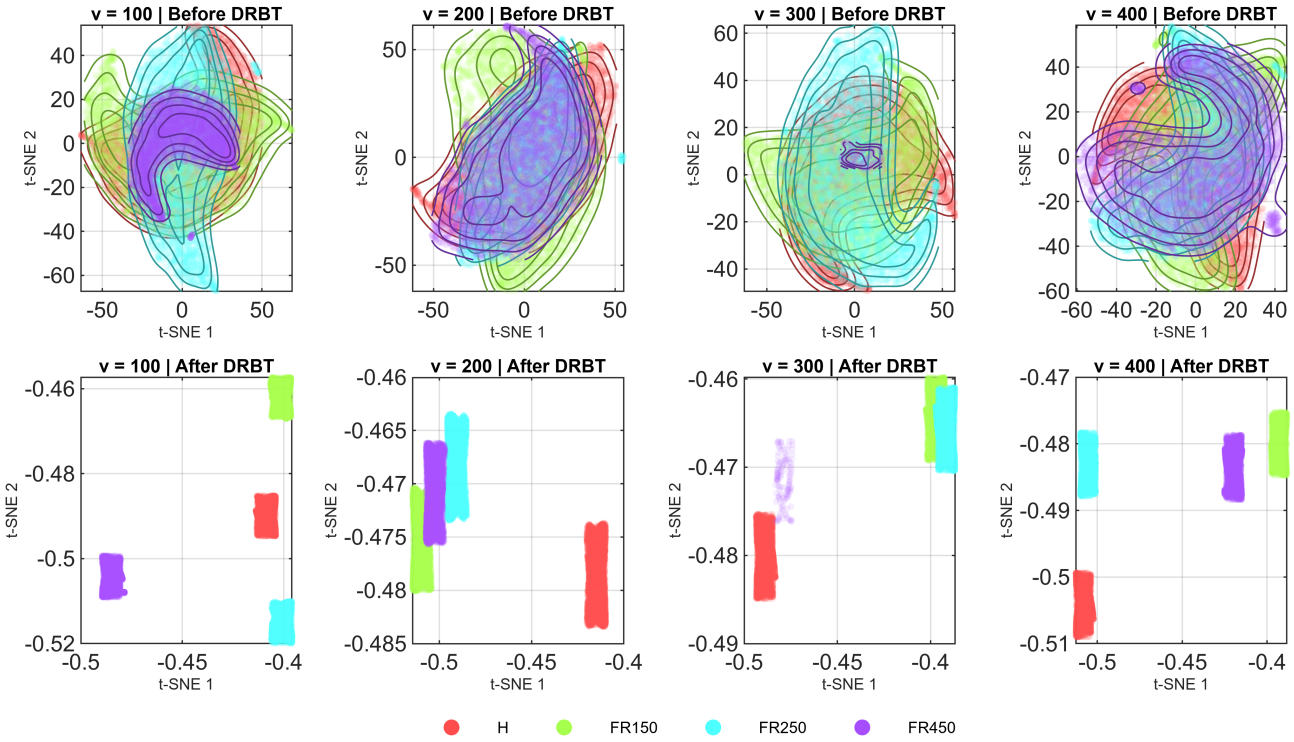


Figure 6.11: The t-SNE embeddings, both prior to and following the DRBT transformation, for the roller fault family (FR) at a load level of **1400 N**, demonstrate that the DRBT consistently achieves compactness and distinct boundary formation across various shaft speeds. This transformation notably enhances the separability of smaller roller defects.

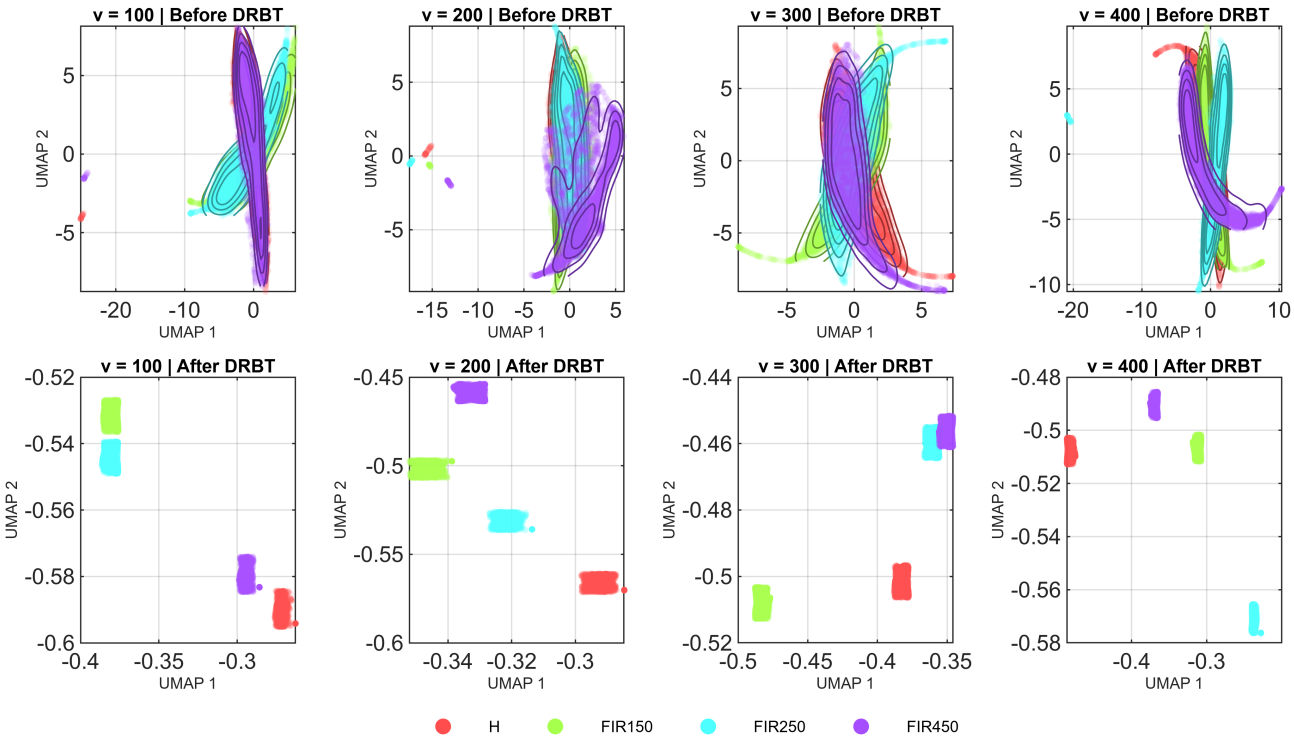


Figure 6.12: UMAP embeddings are depicted before (top) and after (bottom) the DRBT transformation for the inner-race fault family (FIR) at a load level of 1000 N. The DRBT transformation markedly improves fault separability across all shaft speeds (100–400 Hz).

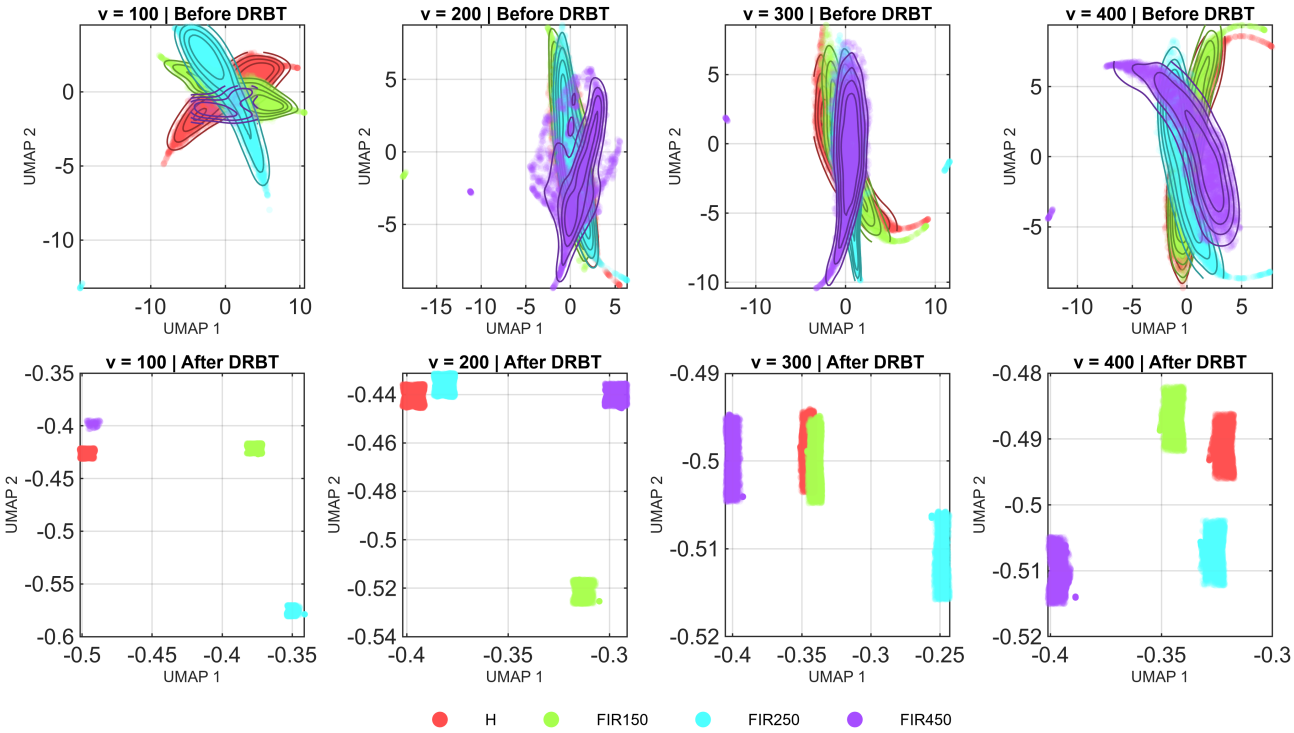


Figure 6.13: UMAP embeddings, both prior to and following the DRBT transformation, are presented for the inner-race fault family (`FIR`) at a load level of 1400 N. A minor overlap between `FIR150` and healthy data is discernible at 300 Hz; however, the overall cluster separability remains substantial.

challenging cases in the CWRU dataset, respectively. Among these, the focus was restricted to the 48kHz drive-end bearing fault recordings, which included inner race, ball, and outer race (centered) defects that were explicitly marked as N1 or N2 by 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 Figure 6.16 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, nearly orthogonal clusters with distinct boundaries. This transformation effectively reduces the curvature and local density bias, allowing the underlying fault structure to become more discernible. Consequently, all four classes became linearly separable, with significantly reduced intra-class spreads and balanced inter-cluster distances.

A similar progression was observed in the UMAP-reduced feature space (Figure 6.17). Prior to DRBT, the UMAP projections displayed curvilinear trajectories and partial overlaps, particu-

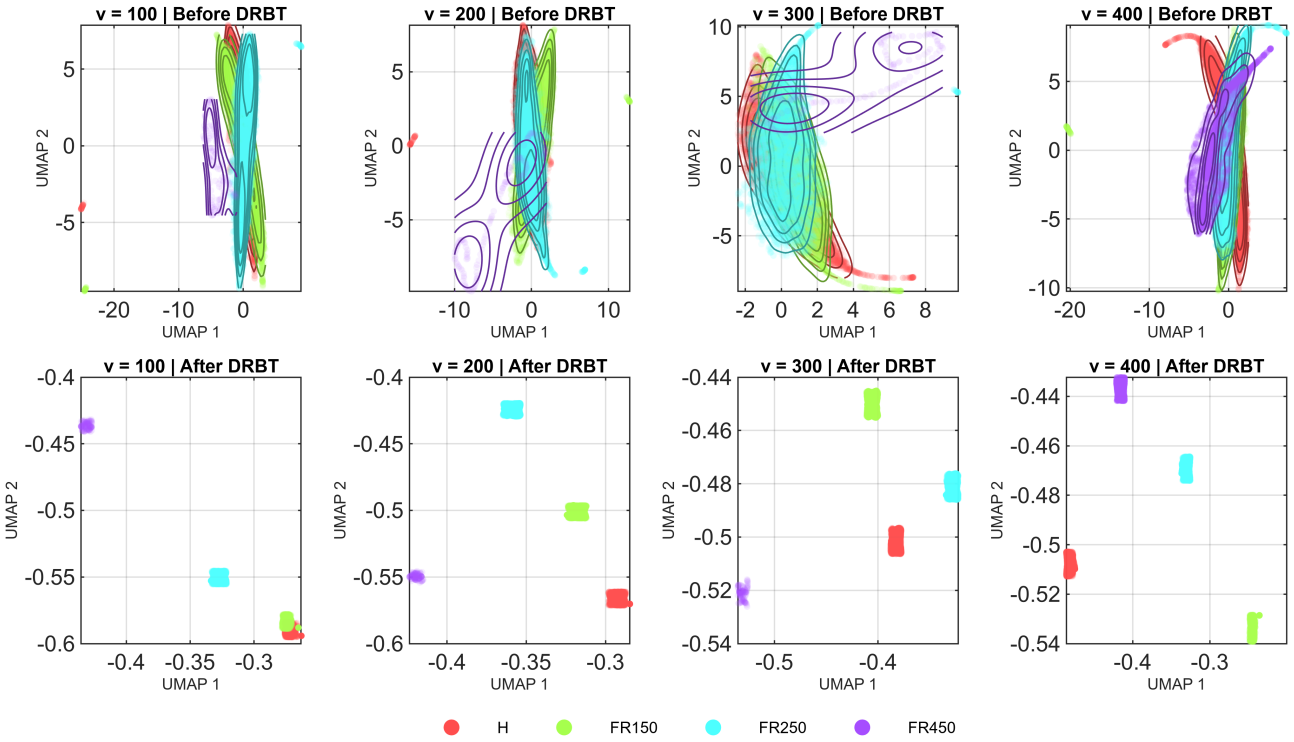


Figure 6.14: UMAP embeddings are depicted prior to (top) and following (bottom) the DRBT transformation for the roller fault family (`FR`) at a load level of 1000 N. The `FR150` and healthy clusters demonstrate minor overlap at 100 Hz, whereas other faults remain distinctly separable post-transformation.

larly among the *Healthy*, *Ball*, and *Outer Race* states. These overlaps result from the continuous mapping nature of UMAP, which preserves the neighborhood topology but not necessarily the inter-manifold 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 transformation.

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, irrespective of the embedding technique. By flattening local density gradients and regularizing inter-cluster spacing, the DRBT mitigates the geometric distortion introduced by nonlinear reductions and enhances class discriminability. These results confirm its suitability as a universal preprocessing stage for unsupervised fault diagnosis, facilitating improved clustering and anomaly detection under variable fan operating speeds.

Quantitative Evaluation The results presented in Tables 6.7 and 6.8 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 across all examined rotational speeds in the CWRU dataset. Prior to the 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

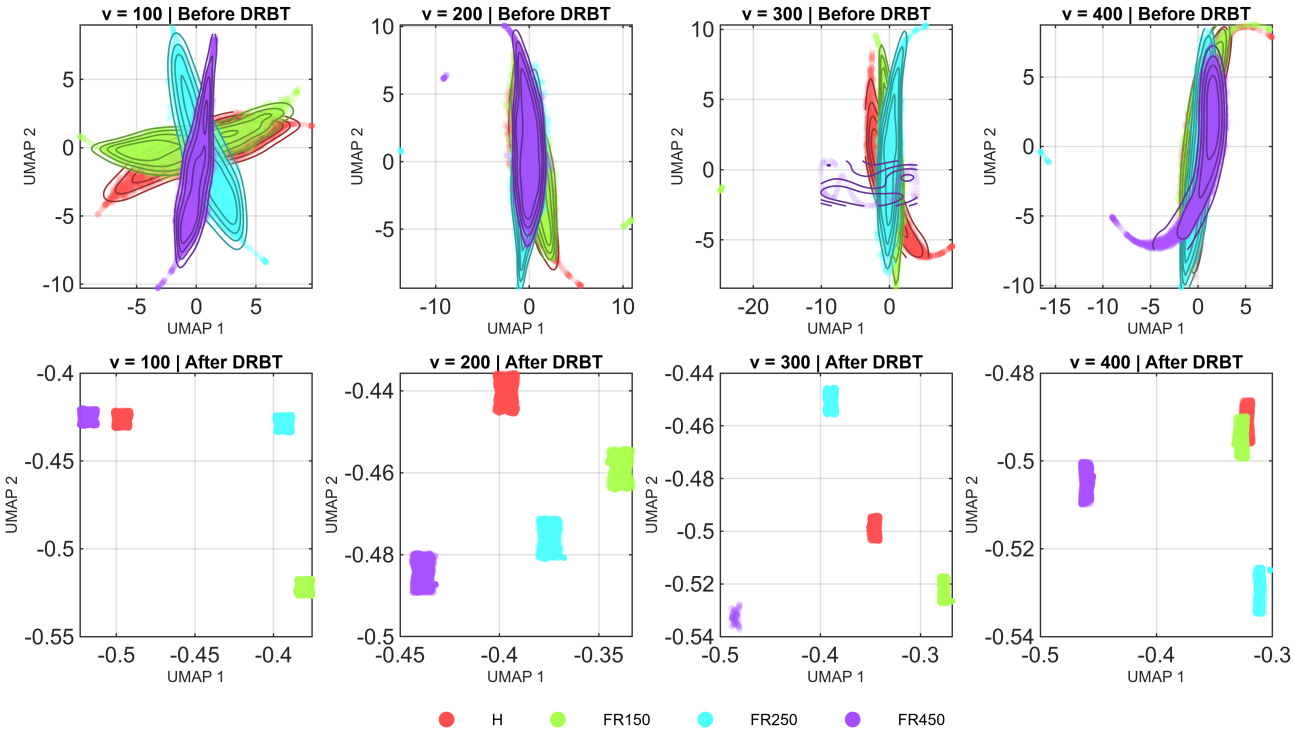


Figure 6.15: UMAP embeddings, both prior to and following the DRBT transformation, are presented for the roller fault family (`FR`) at a load level of 1400 N. The `FR150` class exhibits partial overlap with the healthy cluster at 400 Hz, whereas faults of greater severity (`FR250`, `FR450`) manifest as compact, isolated regions.

t-SNE and UMAP embeddings.

These findings corroborate the visual trends observed in Figures 6.16 and 6.17, where the DRBT transformation effectively enhanced the cluster compactness and separation across the four bearing conditions (healthy, inner race, ball, and outer race fault). The nonlinear mapping minimized the density distortions inherent to the t-SNE and UMAP projections, ensuring that the manifolds corresponding to each fault class remained locally coherent and were 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 both qualitative and quantitative evidence across three distinct datasets, namely 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 necessitating dataset-specific tuning.

6.3.3 Comparison with Baseline and State-of-the-Art Methods

To assess the efficacy of the proposed DRBT in comparison to widely utilized distribution-reshaping and density-homogenizing techniques within the reduced space, Figure 6.18 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

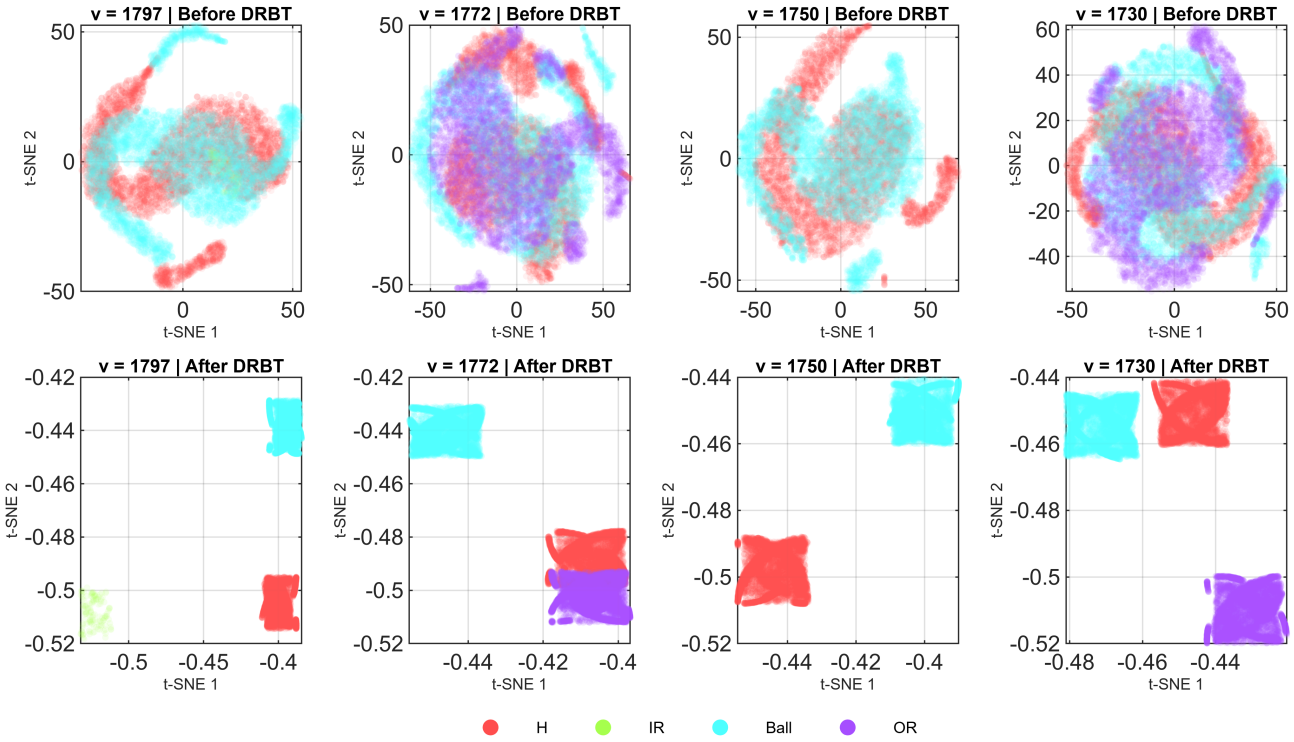


Figure 6.16: The t-SNE embeddings of the *CWRU* dataset at various rotational speeds (v, expressed in rpm), both before and after the application of DRBT, illustrate that the proposed transformation effectively mitigates manifold curvature, equalizes density, and enhances the separability among the four health conditions.

Table 6.7: 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 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

Table 6.8: 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 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

manifold, independent of any variations stemming 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

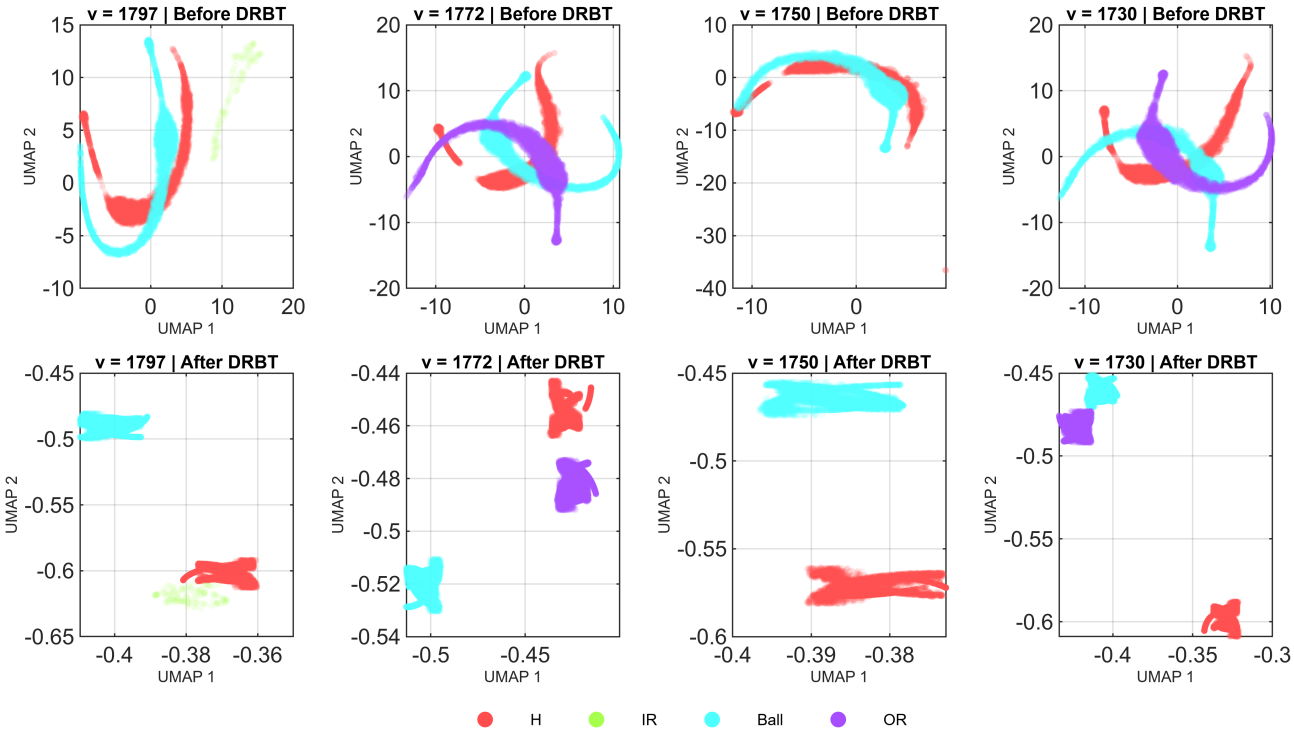


Figure 6.17: UMAP embeddings of the *CWRU* dataset at various rotational speeds (v, expressed in rpm) are depicted both before and after the implementation of DRBT. The DRBT technique successfully converts the initially curved and overlapping manifolds into compact, well-separated clusters with uniform density.

procedure. For reference, the unaltered UMAP embedding is depicted in the top-left panel.

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 Gaussianisation 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.

In contrast, the DRBT generated highly compact, uniformly dense, and distinctly separated clusters for all fault categories and severity levels. In contrast to baseline methods, DRBT eliminates nonlinear curvature, equalizes cluster density, and restores 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 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.

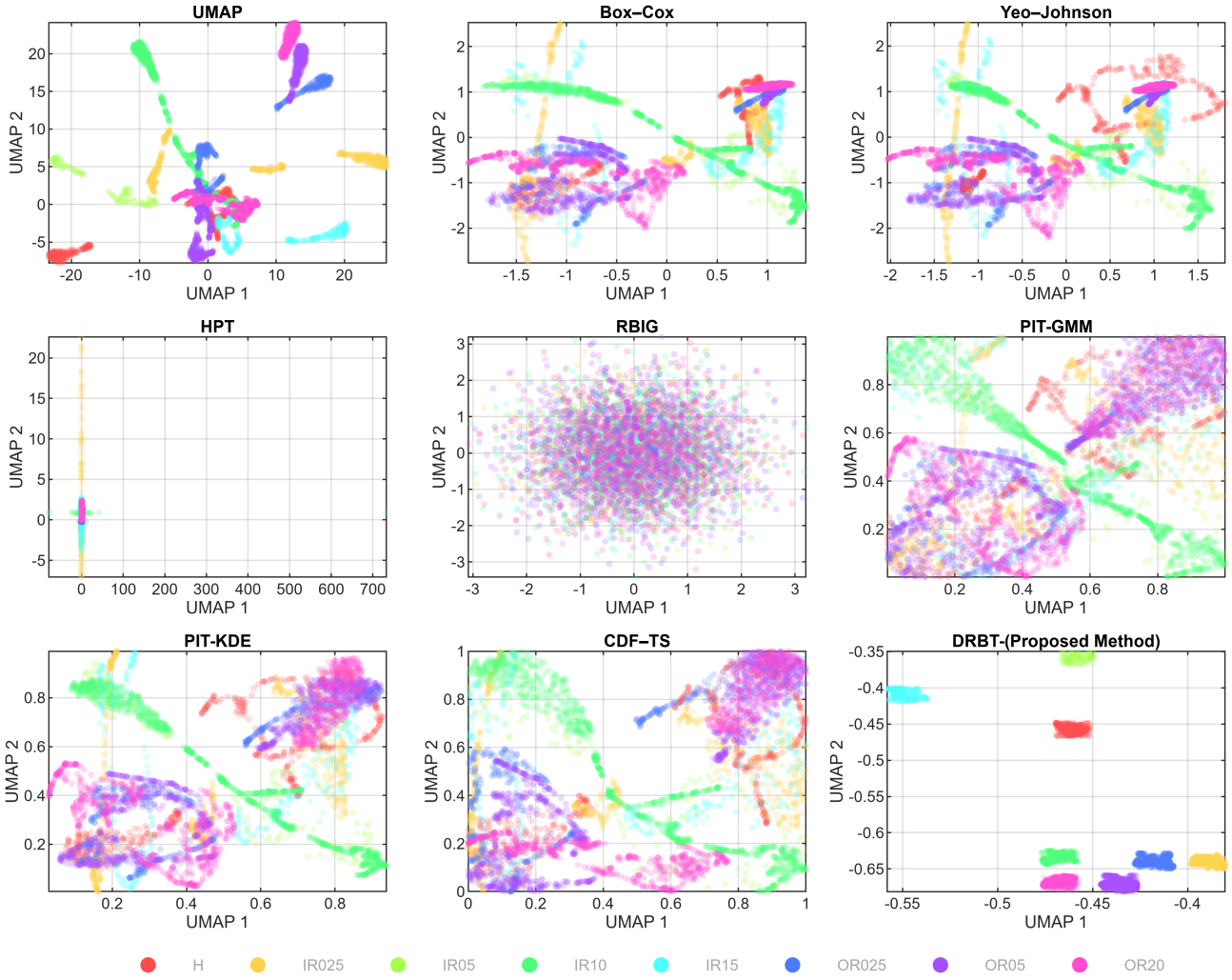


Figure 6.18: The comparative assessment of distribution-reshaping techniques applied directly to two-dimensional UMAP embeddings reveals several key findings. The unprocessed UMAP projection was characterized by elongated trajectories and significant interclass overlap. Traditional power transformations, 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 an excessively stretched embedding with a collapsed class structure, whereas RBIG produces an almost homogeneous point cloud owing to iterative whitening in 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.

6.3.4 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 (Batzner et al., 2024), PaDiM (Defard et al., 2021), and Fully Convolutional Data Description (FCDD) (Liznerski et al., 2020). All models were trained exclusively on healthy data and evaluated under identical experimental conditions to ensure fair and consistent comparisons. In parallel, an Isolation Forest classifier was applied to features obtained through the proposed transformation pipeline, namely *TAIW* $\rightarrow$ *t-SNE* $\rightarrow$ *DRBT*, enabling a direct comparison between fully end-to-end deep learning methods and a lightweight, statistically driven feature transformation framework.

Figure 6.19 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 domain-specific adaptation. FCDD achieves a more balanced detection performance owing to its fully convolutional one-class formulation, which is tailored to describe the anomalies. Notably, the proposed *TAIW* $\rightarrow$ *t-SNE* $\rightarrow$ *DRBT* pipeline combined 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 non-stationary vibration-based bearing-fault diagnosis in industrial environments.

6.3.5 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, which was 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 4.4.4. 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 6.2. 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 Figure 6.20.

The minimum peak prominence parameter, denoted as $p_{\min}$, exhibits 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 signif-

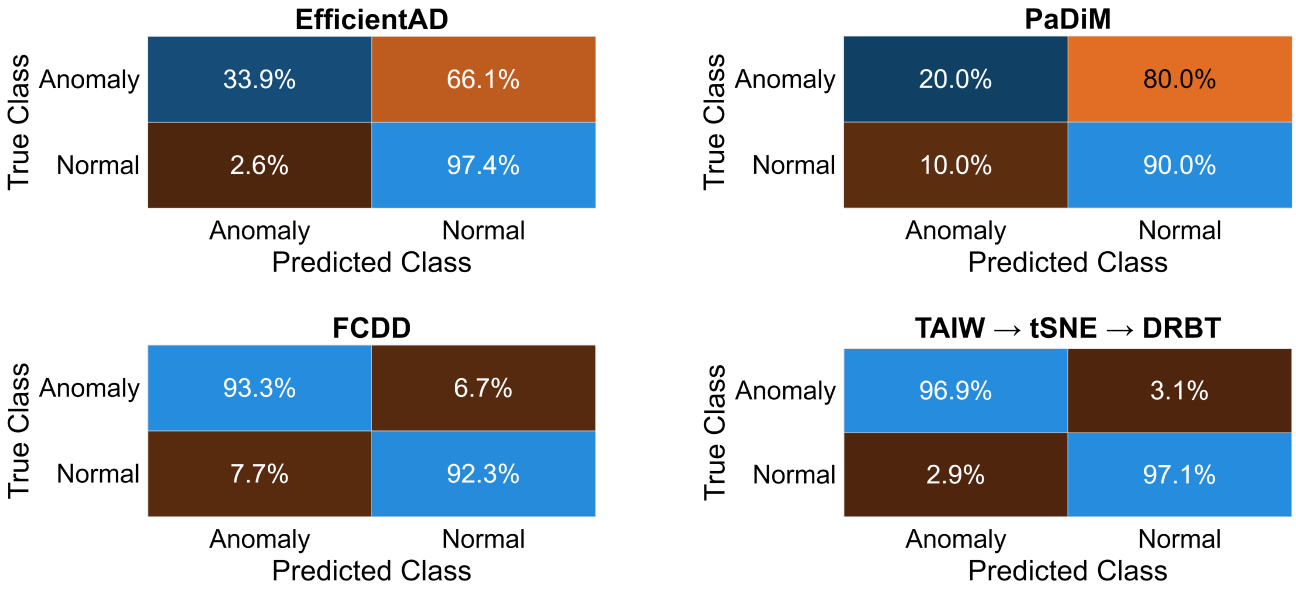


Figure 6.19: Row-normalized confusion matrices comparing anomaly detection performance on the MOIRA–UNIMORE bearing dataset using spectrogram representations derived from a single vibration channel. From top left to bottom right: EfficientAD, PaDiM, Fully Convolutional Data Description (FCDD), and the 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.

icant decrease in the performance. This observation corroborates the notion that excessively stringent prominence thresholds can suppress structurally significant secondary modes in the presence of background noise. In contrast, moderate values effectively differentiated 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 the 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 compactions. 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. This behavior is advantageous in practice because 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 demonstrated that the DRBT functioned within extensive and well-defined stability regions. It was observed that only a limited subset of parameters, particularly $p_{\min}$ and δ, ne-

cessitated a conservative selection to prevent degenerate behavior. In contrast, the remaining parameters exhibited a significant robustness. These results substantiate the hyperparameter rationale outlined in Section 6.1.7 and affirm that the reported values represent typical operating points rather than fragile optima.

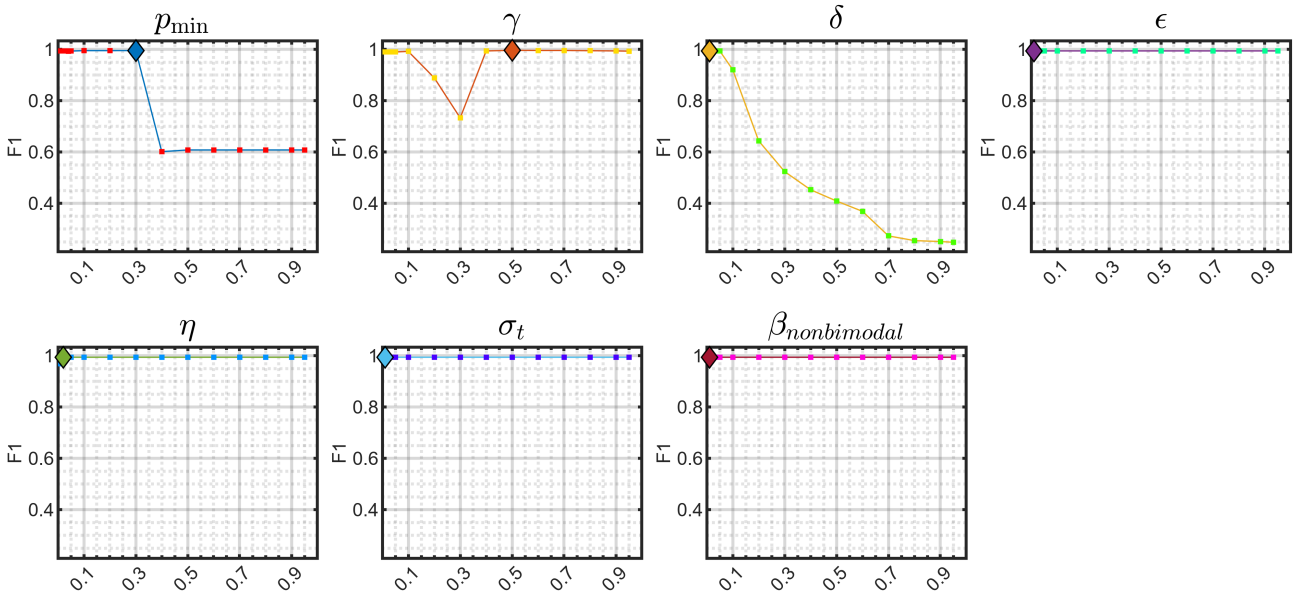


Figure 6.20: One-parameter-at-a-time (OPAT) sensitivity analysis was performed on the DRBT hyperparameters under conditions of noise. The F_1 score of the OCSVM detector is depicted as a function of each parameter, with all other parameters held constant at their nominal values (Table 6.2). Sensitivity 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.

Chapter 7

Entropy-Based Intelligent Spectrogram Window Search Strategies

Nonstationary vibration signals frequently occur in contemporary rotating machinery, industrial drive systems, and motion-controlled platforms, where rapid fluctuations in speed, load, and dynamic interactions result in time-varying spectral behavior (Antoni, 2009; Jablonski & Dziedziech, 2022; Urbanek et al., 2013). The reliable extraction of these evolving components necessitates time–frequency representations capable of resolving both transient and slowly varying phenomena. The short-time Fourier transform (STFT) remains one of the most extensively utilized tools owing to its robustness, interpretability, and strong theoretical foundation (Allen & Rabiner, 1977; Gambardella, 1968). However, the diagnostic efficacy of the spectrogram is critically dependent on the selection of the window length, which determines the balance between the time and frequency resolution. An inappropriate window can obscure impulsive signatures, smear modulated harmonics, or distort structural features, rendering automated and efficient window selection essential for analyzing complex nonstationary vibration signals.

Various methodologies have been proposed to address the challenges associated with the analyses of nonstationary vibration signals. Classical spectrogram-based approaches, which rely on the Short-Time Fourier Transform (STFT), remain prevalent and have been effectively applied to gearbox diagnostics, bearing health assessment, and internal combustion engine monitoring (Yan & Gao, 2009; Yesilyurt, 2004; H. Yu et al., 2006). However, these methods typically rely on manually selected window parameters, which are often determined by experience or trial and error. To mitigate this subjectivity, adaptive and data-driven windowing strategies have been investigated, including time-frequency concentration maximization (Varela et al., 2005), empirical modeling combined with constant-Q transforms (Nisar et al., 2016), and other forms of adaptive resolution control. Recently, the *Intelligent Spectrogram* (Jablonski & Dziedziech, 2022) introduced a blind, statistics-based framework that evaluates multiple candidate windows using contrast and noise-related quality coefficients and selects the window that optimizes the representation of the time-frequency variations. Although highly effective, this approach necessitates computing spectrograms for an entire predefined set of window lengths, which is computationally expensive for long-duration recordings or high-resolution analyses.

Despite these advancements, a fundamental limitation persists in most existing methodologies: the requirement to evaluate a vast array of candidate window lengths to ascertain an appro-

priate time-frequency resolution. Techniques such as the Intelligent Spectrogram (Jablonski & Dziedziech, 2022) offer fully automated selection mechanisms; however, they necessitate computing spectrograms for every window in a predefined list. This exhaustive approach results in computational costs proportional to the number of candidate windows, which becomes prohibitive for long-duration signals, high sampling rates, and real-time monitoring applications. This challenge is exacerbated in practical industrial environments, where multinational production facilities operate hundreds of similar machines distributed across sites worldwide. Even if vibration and process data from each machine are recorded for only short intervals per day, the cumulative volume spanning multiple channels, machines, and multi-year archives becomes enormous. Processing such datasets with exhaustive window sweeps is computationally impractical and often infeasible within the standard maintenance cycles. Moreover, industrial systems frequently exhibit rapid speed variations, transient impacts, and multicomponent interactions that require dense time-frequency grids or wide window ranges. These considerations underscore the need for significantly more efficient window-selection strategies that preserve diagnostic fidelity while drastically reducing the number of spectrogram evaluations required.

This chapter introduces an entropy-based exhaustive search alternative to the coefficient of variation (CV) method (Jablonski & Dziedziech, 2022) and addresses the limitations of exhaustive search methods by introducing two novel entropy-based algorithms for intelligent spectrogram window selection. Both methods exploit the characteristic structure of time-frequency entropy across window scales to circumvent the exhaustive evaluation of all candidate lengths. The main contributions of this study are as follows.

- A comprehensive entropy-based framework has been established to assess window quality, effectively capturing both time-domain and frequency-domain concentration through an entropy product measure.
- A rapid *Ternary-Search Window Selection* algorithm is proposed, which utilizes the approximate unimodal behavior of the entropy product to reduce the search complexity from $\mathcal{O}(m)$ to approximately $\mathcal{O}(\log m)$ evaluations.
- A complementary *Bidirectional V-Shape Window Selection* algorithm is introduced, which exploits local monotonicity patterns of entropy across window scales to enable early termination, thereby reducing computational costs in practical applications.
- Both Ternery-search and Bidirectional V-shape algorithms achieve significantly lower computational cost than exhaustive search methods (coeeficient of variation and Entropy based full sweep search), yielding substantial savings on both synthetic and real vibration datasets.

These contributions offer a computationally efficient alternative to existing window selection strategies and are particularly well-suited for long-duration, multichannel, and large-scale vibration-monitoring applications.

7.1 Background and Problem Formulation

7.1.1 Short-Time Fourier Transform and Spectrogram

The short-time Fourier transform (STFT) is a widely used time–frequency analytical tool for examining nonstationary vibration signals. For a discrete-time signal $x[n]$ and an analysis window $w[\cdot]$ of length L, the STFT is defined as

$$S[n,k] = \sum_{m=0}^{L-1} x[m+n]\, w[m]\, e^{-j2\pi km/L}. \tag{7.1}$$

The magnitude squared of the STFT yields a spectrogram that depicts the time-varying distribution of spectral energy. The window length L significantly affects the trade-off between the time and frequency resolutions: smaller windows enhance temporal localization, whereas larger windows provide improved frequency resolution. Consequently, the diagnostic efficacy of the spectrogram is highly dependent on the selection of L, particularly for signals with transient impulses, modulated harmonics or speed-varying components. In practice, the maximum window length is constrained to $L_{\max} = N/4$, where N denotes the total signal length, to prevent excessive temporal averaging and numerical burdens. Thus, selecting an appropriate window length is a critical and complex step in spectrogram-based analysis.

7.1.2 Window-Length Selection in Time–Frequency Analysis

Traditional methods often employ fixed windows selected through heuristics or practitioner experience, which may be inadequate under varying operational conditions. More systematic approaches involve evaluating spectrograms across multiple candidate window lengths and selecting the one that optimizes the chosen criteria. Representative metrics include time–frequency concentration, energy contrast, local smoothness, and noise suppression (Jablonski & Dziedziech, 2022; Nisar et al., 2016; Varela et al., 2005). Among these, the Intelligent Spectrogram (Jablonski & Dziedziech, 2022) introduces a blind, statistics-driven approach that computes the coefficient of variation along both the time and frequency axes and selects the window length by identifying the extremum of the product of these two measures. Although highly effective, this technique necessitates the evaluation of all window lengths in a predefined list, resulting in computational costs proportional to the number of candidates. Exhaustive window sweeps are computationally prohibitive for applications involving long-duration signals, high sampling frequencies, or large-scale monitoring deployments.

7.1.3 Spectral Entropy as a Window-Quality Measure

Spectral entropy provides a systematic methodology for quantifying the structural complexity of a spectrogram. For a specified window candidate, denoted by w, and its corresponding spectrogram magnitude $A^{(w)} = |S^{(w)}|$, the normalization in the time domain is expressed as

$$A_t^{(w)}(i,j) = \frac{A^{(w)}(i,j)}{\sum_i A^{(w)}(i,j) + \varepsilon}, \tag{7.2}$$

which yields the time entropy

$$H_t^{(w)}(j) = -\sum_i A_t^{(w)}(i,j)\, \frac{\log A_t^{(w)}(i,j)}{\log N_f}, \tag{7.3}$$

and its temporal mean

$$E_t(w) = \mathrm{mean}_j \, H_t^{(w)}(j), \tag{7.4}$$

where N_f signifies the number of frequency bins. Analogously, the normalization in the frequency domain is defined by

$$A_f^{(w)}(i,j) = \frac{A^{(w)}(i,j)}{\sum_j A^{(w)}(i,j) + \varepsilon}, \tag{7.5}$$

resulting in the frequency entropy

$$H_f^{(w)}(i) = -\sum_j A_f^{(w)}(i,j) \, \frac{\log A_f^{(w)}(i,j)}{\log N_t}, \tag{7.6}$$

and its frequency-averaged value

$$E_f(w) = \mathrm{mean}_i \, H_f^{(w)}(i), \tag{7.7}$$

where N_t represents the number of time bins.

The product

$$E_{tf}(w) = E_t(w) \, E_f(w), \tag{7.8}$$

serves as a scalar quality metric that concurrently encapsulates the temporal and spectral dispersion for the window candidate indexed by w. Empirical observations indicate that $E_{tf}(w)$ frequently exhibits approximately unimodal behavior or forms a characteristic V-shape across window candidates (or equivalently, across their associated lengths L_w), with the optimal window corresponding to the minimum value of $E_{tf}(w)$. This characteristic renders entropy-based criteria particularly effective for automated and robust window selection.

7.1.4 Problem Statement

Let $K = \{k_1, k_2, \ldots, k_m\}$ represent a discrete set of window exponents corresponding to candidate window lengths, defined as

$$L_w = 2^{k_w}, \qquad w = 1, \ldots, m, \tag{7.9}$$

subject to the constraint $L_w \leq N/4$ for all w, where N denotes the total signal length. The window-selection problem can be expressed as

$$\hat{w} = \arg \min_{w \in \{1,\ldots,m\}} E_{tf}(w). \tag{7.10}$$

A comprehensive evaluation of this objective requires computing a Short-Time Fourier Transform (STFT) for each of the m candidate windows. Given that the computational cost of a single STFT with window length L_w is approximately $\mathcal{O}(N \log L_w)$, where N represents the signal length, an exhaustive search incurs a total cost on the order of $\mathcal{O}(m \, N \log L_{\max})$. This becomes increasingly prohibitive when m is large, spectrograms must be recomputed repeatedly for multichannel measurements, or when processing long-duration recordings as part of large-scale monitoring infrastructure. The objective of this study is to develop more efficient optimization strategies that leverage the structural behavior of entropy across window scales, thereby reducing the number of required spectrogram evaluations while maintaining optimal or near-optimal window selection performance.

7.2 Proposed Entropy-Based Window Selection Algorithms

The entropy-product criterion introduced in Section 7.1 offers a scalar measure of window quality, denoted as $E_{tf}(w)$, which concurrently reflects both temporal and spectral dispersion for a window of length $L_w = 2^{k_w}$. Consequently, the task of window selection is reduced to a discrete one-dimensional optimization over the exponent set $K = \{k_1, k_2, \ldots, k_m\}$, as formalized in (7.10). A basic solution involves evaluating $E_{tf}(w)$ for each $k_w \in K$, necessitating the computation of m spectrograms and incurring a total computational cost of the order of $\mathcal{O}(m\, N \log L_{\max})$. Although straightforward, this method becomes increasingly onerous for long signals, high sampling frequencies, or multichannel monitoring, where numerous spectrograms must be recalculated.

The central concept of the proposed framework is to leverage the characteristic structure of the entropy curve $E_{tf}(w)$ across different window scales. Empirical observations indicate that $E_{tf}(w)$ generally exhibits either a unimodal trend or a locally V-shaped profile, suggesting that evaluating all m candidates is not necessary. Exploiting this structure allows for a significant reduction in the number of spectrogram computations. This section first formalizes the entropy-based exhaustive strategy as a benchmark and subsequently introduces two accelerated optimization methods: (i) a ternary-search-inspired approach that capitalizes on approximate unimodality and (ii) a bidirectional search that identifies V-shaped curvature to facilitate early stopping.

7.2.1 Entropy-Based Exhaustive Window Selection

In accordance with the entropy formulation delineated in Section 7.1, the most straightforward selection strategy involves calculating $E_{tf}(w)$ for each window length in the candidate set. For a window of length $L_w = 2^{k_w}$ and an overlap fraction ρ, the corresponding spectrogram magnitude $A^{(w)} = |S^{(w)}|$ is determined, followed by normalization in both the time and frequency domains to derive the entropy components $E_t(w)$ and $E_f(w)$. The product of these components, $E_{tf}(w) = E_t(w)\, E_f(w)$, functions as a metric of window quality. The optimal window, identified through exhaustive evaluation, is thus $\hat{w}_{\mathrm{E}} = \arg\min_{w\in\{1,\ldots,m\}} E_{tf}(w)$, $\hat{L}_{\mathrm{E}} = 2^{k_{\hat{w}_{\mathrm{E}}}}$. The entire procedure is encapsulated in Algorithm 7. Although this exhaustive approach maintains the blind, statistics-driven characteristic of the Intelligent Spectrogram—utilizing entropy in place of the coefficient of variation—it necessitates the evaluation of all m window candidates and the recomputation of the STFT m times. Consequently, its computational cost scales linearly with the number of window candidates, thereby serving as a benchmark against which the efficiency of the proposed accelerated search strategies is evaluated.

7.2.2 Ternary-Search Window Selection

The initial algorithm, delineated in Algorithm 8, employs a discrete ternary search on the ordered exponent list $k_1 < k_2 < \cdots < k_m$. Empirical observations suggest that the entropy curve $E_{tf}(w)$ frequently displays a singular dominant minimum or narrow valley when considered as a function of $\log_2 L_w$. Given this approximate unimodality, the ternary search method effectively reduces substantial portions of the search interval without necessitating the evaluation of each candidate.

The active search interval is defined by two bounds, (k_L, k_R). In each iteration, two interior

exponents are calculated:

$$k_a = \left\lfloor \frac{2k_L + k_R}{3} \right\rfloor, \qquad k_b = \left\lceil \frac{k_L + 2k_R}{3} \right\rceil. \tag{7.11}$$

The corresponding window lengths are

$$L_{w_a} = 2^{k_a}, \qquad L_{w_b} = 2^{k_b}, \tag{7.12}$$

and the entropy values are determined by

$$E_{tf}(w_a), \qquad E_{tf}(w_b). \tag{7.13}$$

If $E_{tf}(w_a) \leq E_{tf}(w_b)$, the minimum cannot reside to the right of k_b; thus, the interval is reduced to $[k_L,\, k_b]$. Conversely, the search interval is adjusted to $[k_a, k_R]$.

This iterative procedure persists until $k_R - k_L \leq 3$, at which point an exhaustive search is conducted on the limited subset $K = \{k_L, k_{L+1}, \ldots, k_R\}$.

A caching structure C retains tuples of the form

$$(k,\, E_t(w),\, E_f(w),\, E_{tf}(w)), \tag{7.14}$$

ensuring that each window length is evaluated no more than once. Non-finite values are substituted with large sentinel constants to preserve numerical stability.

The final selected exponent is determined as

$$\hat{k} = \arg\min_{k \in \mathcal{K}} E_{tf}(w), \tag{7.15}$$

and the corresponding optimal window length is $\hat{L} = 2^{\hat{k}}$. Since each iteration of the ternary search reduces the interval by one-third, the number of STFT evaluations increases as $\mathcal{O}(\log m)$, as opposed to $\mathcal{O}(m)$ for an exhaustive scan. Incorporating the STFT cost, the total complexity becomes $\mathcal{O}\big((\log m)\, N \log \hat{L}\big)$, representing a significant enhancement over $\mathcal{O}(m\, N \log L_{\max})$. This renders the approach particularly advantageous for dense window grids, long-duration vibration datasets, multichannel acquisitions, and large-scale condition-monitoring infrastructure.

7.2.3 Bidirectional V-Shape Window Selection

The second algorithm, as outlined in Algorithm 9, is specifically designed to leverage locally V-shaped or valley-like patterns in the entropy product across various window scales. Rather than presuming a global unimodal nature of the curve $E_{tf}(w)$, this method is predicated on a less stringent and empirically substantiated condition: near an optimal window index w^*, the entropy product generally diminishes to a minimum before increasing, thereby forming a local V-shape when plotted over the discrete index set $w \in \{1, \ldots, m\}, \quad k_w \in K$, where each exponent k_w determines the corresponding window length $L_w = 2^{k_w}$. This local characteristic facilitates early termination through straightforward three-point comparisons, thereby minimizing the need to evaluate all potential window candidates.

Algorithm 7 Entropy-Based Exhaustive Window Selection

Require: Signal $x[0:N-1]$, sampling frequency f_s, exponent set $K = \{k_1, \ldots, k_m\}$, overlap fraction $\rho \in (0,1)$

Ensure: Selected window length $\hat{L}_{\mathrm{E}}$, entropy curves Q_t, Q_f, and Q_{tf}

Initialize vectors $Q_t \in \mathbb{R}^m$, $Q_f \in \mathbb{R}^m$

for $w = 1$ to m **do**

  $L_w \leftarrow 2^{k_w}$ ▷ Window length in samples

  $L_w \leftarrow \min(L_w, N/4)$ ▷ Enforce maximum window length

  $L_w \leftarrow$ next even integer if needed

  $L_{\mathrm{ov}} \leftarrow \lfloor \rho\, L_w \rfloor$ ▷ Overlap in samples

  Compute STFT

$$S^{(w)} \leftarrow \mathrm{STFT}\big(x; \mathrm{Hamming}(L_w), L_{\mathrm{ov}}, f_s\big)$$

  $A^{(w)} \leftarrow \big|S^{(w)}\big| + \varepsilon$ ▷ Magnitude spectrogram

  ▷ Time-axis normalization and entropy

  Normalize each time slice:

$$A_t^{(w)}(:, j) \leftarrow \frac{A^{(w)}(:, j)}{\sum_i A^{(w)}(i, j) + \varepsilon}$$

  for each time index j **do**

    $H_t^{(w)}(j) \leftarrow \mathrm{Entropy}\big(A_t^{(w)}(:, j)\big)$

  end for

  $Q_t(w) \leftarrow \mathrm{mean}_j\, H_t^{(w)}(j)$

  ▷ Frequency-axis normalization and entropy

  Normalize each frequency bin:

$$A_f^{(w)}(i, :) \leftarrow \frac{A^{(w)}(i, :)}{\sum_j A^{(w)}(i, j) + \varepsilon}$$

  for each frequency index i **do**

    $H_f^{(w)}(i) \leftarrow \mathrm{Entropy}\big(A_f^{(w)}(i, :)\big)$

  end for

  $Q_f(w) \leftarrow \mathrm{mean}_i\, H_f^{(w)}(i)$

end for

$Q_{tf} \leftarrow Q_t \odot Q_f$ ▷ Element-wise product

$\hat{w}_{\mathrm{E}} \leftarrow \arg\min_w Q_{tf}(w)$

$\hat{L}_{\mathrm{E}} \leftarrow 2^{k_{\hat{w}_{\mathrm{E}}}}$

return $\hat{L}_{\mathrm{E}}, Q_t, Q_f, Q_{tf}$

The algorithm functions directly within the discrete index domain $w \in \{1, \ldots, m\}$. Two pointers, denoted as L and R, are initialized at the leftmost and rightmost indices. The routine EVAL assesses the entropy product $E_{tf}(w)$ and computes a new spectrogram only if the value is not already cached. Similar to the ternary search method, all entropy values are derived from the spectrogram magnitudes using the definitions provided in Section 7.1, and nonfinite outputs are substituted with a large sentinel constant. The sequence of the evaluated indices is recorded in a log $\mathcal{J}$, and the total number of spectrogram computations is monitored using a counter c.

Algorithm 8 Ternary-Search Window Selection

Require: $x \in \mathbb{R}^N$, $f_s \in \mathbb{R}_+$, $K = \{k_1 < \cdots < k_m\} \subset \mathbb{Z}$, $\alpha \in (0,1)$
Ensure: $\hat{E} = 2^{\hat{k}}$ where $\hat{k} \in K$ minimizes $E_{tf}(k) = E_t(k)\,E_f(k)$; vectors E_t, E_f, E_{tf} for evaluated k; log $\mathcal{J}$

```
(C, J, k_L, k_R) ← (∅, ∅, k_1, k_m)
function EvalOrCache(k)
    if ∃r ∈ C : r_1 = k then return (r_2, r_3, r_4)
    end if
    w ← 2^k, o ← ⌊αw⌋, S ← spectrogram(x, Hamming(w), o, nfft=auto, f_s)
    A ← |S| + ε
    A_t ← A/(1^⊤ A + ε),
   H_t(j) ← −Σ_i A_t(i,j) log A_t(i,j)/ log N_f,
   E_t ← mean_j H_t(j)
    A_f ← A/(A1 + ε),
   H_f(i) ← −Σ_j A_f(i,j) log A_f(i,j)/ log N_t,
   E_f ← mean_i H_f(i)
    E_tf ← E_t E_f
    if ¬ isfinite(E_tf) then
        E_tf ← realmax
    end if
    C ← C ∪ {[k, E_t, E_f, E_tf]}
    return (E_t, E_f, E_tf)
end function
while k_R − k_L > 3 do
    k_1 ← ⌊(2k_L + k_R)/3⌋,   k_2 ← ⌈(k_L + 2k_R)/3⌉
    (_, _, e_1) ← EvalOrCache(k_1),
   (_, _, e_2) ← EvalOrCache(k_2)
    J ← J ∪ {[k_1, e_1], [k_2, e_2]}
    if e_1 ≤ e_2 then
        k_R ← k_2
    else
        k_L ← k_1
    end if
end while
K ← {k_L, ..., k_R}
for all k ∈ K do
    (_, _, e) ← EvalOrCache(k)
    J ← J ∪ {[k, e]}
end for
k̂ ← arg min_{k∈K}{ e stored at k }
Ê ← 2^k̂
sort C by column 1;   E_t ← C(:,2),  E_f ← C(:,3),  E_tf ← C(:,4)
return (Ê, E_t, E_f, E_tf, J)
```

Following the initial evaluations at L, $L+1$, R, and $R-1$, the algorithm alternates between examining triplets on the left and right sides of the current interval. On the left side, it considers the triplet $(L,\ L+1,\ L+2)$ with corresponding entropy values $e_L = E_{tf}(L), \quad e_M =$

$E_{tf}(L+1), \quad e_R = E_{tf}(L+2)$. If the middle element satisfies the V-shape condition $e_M \leq e_L + \tau \quad \text{and} \quad e_M \leq e_R + \tau$, for a small tolerance $\tau > 0$, then the window index $w = L + 1$ is selected as a local minimizer, and the procedure is terminated. Conversely, if $e_R < e_M - \tau$, the minimum is likely to lie to the right, and the left pointer is updated via $L \leftarrow L + 1$.

Symmetric logic is applied on the right side using the triplet $(R-2,\ R-1,\ R)$. If a V-shape is detected at the center index $R-1$, the algorithm halts; otherwise, if $e_L < e_M - \tau$, the right pointer is updated using $R \leftarrow R - 1$.

These alternating updates continue as long as the search interval contains sufficient points, that is, $L + 2 \leq R$. If the interval contracts without identifying a clear V-shape, a final refinement step is carried out using a small neighborhood $\mathcal{C} = \big\{ \max(1, L-2), \ldots, \min(m, R+2) \big\}$, and the optimal index is extracted as $\hat{w} = \arg\min_{w \in \mathcal{C}} E_{tf}(w), \qquad \hat{k} = k_{\hat{w}}, \qquad \hat{L} = 2^{\hat{k}}$.

The bidirectional search method circumvents the necessity for a strict unimodality condition, thereby offering enhanced robustness when $E_{tf}(w)$ presents multiple shallow minima or exhibits mild-fluctuation noise. In practical applications, only the indices proximate to the valley region are assessed, resulting in an empirical computational cost ranging from $\mathcal{O}(1)$ to $\mathcal{O}(\sqrt{m})$, contingent upon the sharpness and position of the minimum. This approach significantly reduces computational demands compared to the exhaustive $\mathcal{O}(m)$ method, rendering it a valuable adjunct to the ternary search strategy.

> *Remark. Although entropy is utilized in this chapter to facilitate efficient, non-exhaustive window selection, the proposed search strategies are broadly applicable. Specifically, both the ternary search and bidirectional methods can be directly implemented with any scalar window quality metric that demonstrates similar structural characteristics, including the coefficient of variation criterion employed in the Intelligent Spectrogram framework.*

Table 7.1: A comprehensive summary is provided regarding the number of window candidates assessed by each method—CV, entropy-based exhaustive search, entropy ternary search, and entropy bidirectional search—alongside the corresponding optimal window lengths determined for all datasets and signals.

Dataset	**Signal**	**FS (Hz)**	**Time (s)**	**Windows Evaluated**				**Best Window Length**			
				CV	**E**	**E-T**	**E-Bi**	**CV**	**E**	**E-T**	**E-Bi**
Synthetic	S_1	50000	10	11	11	7	6	2048	256	256	256
	S_2	50000	10	11	11	7	6	2048	256	256	256
	S_3	50000	10	11	11	7	6	65536	256	256	256
	S_4	50000	10	11	11	6	6	4096	256	256	256
MOIRA	M_1	50000	10	11	11	7	6	2048	256	256	256
	M_2	50000	10	11	11	7	6	4096	256	256	256
	M_3	50000	10	11	11	7	6	1024	256	256	256
	M_4	50000	10	11	11	7	6	2048	256	256	256
Torino	T_1	51200	10	11	11	6	6	65536	256	256	256
	T_2	51200	10	11	11	6	6	32768	256	256	256
	T_3	51200	10	11	11	6	6	65536	256	256	256
	T_4	51200	10	11	11	6	6	65536	256	256	256
CWRU	CW_1	48000	5	10	10	5	6	32768	256	256	256
	CW_2	48000	5	10	10	6	6	32768	256	256	256
	CW_3	48000	10	11	11	7	6	65536	256	256	256
	CW_4	48000	10	11	11	7	6	16384	256	256	256

Algorithm 9 Bidirectional V-Shape Window Selection

Require: $x \in \mathbb{R}^N$, $f_s \in \mathbb{R}_+$, $K = \{k_1 < \cdots < k_m\} \subset \mathbb{Z}$, $n_{fft} \in \mathbb{N} \cup \{\varnothing\}$, $\alpha \in (0,1)$, $\tau > 0$
Ensure: $\hat{E} = 2^{\hat{k}}$ where $\hat{k} \in K$ minimizes $E_{tf}(i) = E_t(i)\,E_f(i)$; vectors E_t, E_f, E_{tf}; $\log \mathcal{J}$; c (evaluation count)
Initialize E_t, E_f, E_{tf} as undefined arrays of length m; $\mathcal{J} \leftarrow \varnothing$; $c \leftarrow 0$; $(L, R) \leftarrow (1, m)$
function EVAL(i)
if $E_{tf}(i)$ is defined **then return** $E_{tf}(i)$
end if
$\mathcal{J} \leftarrow \mathcal{J} \cup \{i\}$; $c \leftarrow c + 1$
$w \leftarrow 2^{k_i}$; $o \leftarrow \lfloor \alpha w \rfloor$; $\tilde{n} \leftarrow n_{fft}$ if integer else auto
$S \leftarrow \mathrm{spectrogram}(x, \mathrm{Hamming}(w), o, \tilde{n}, f_s)$; $A \leftarrow |S| + \varepsilon$
$A_t \leftarrow A/(\mathbf{1}^\top A + \varepsilon)$; $H_t(j) \leftarrow -\sum_i A_t(i,j) \log A_t(i,j) / \log N_f$; $E_t(i) \leftarrow \mathrm{mean}_j H_t(j)$
$A_f \leftarrow A/(A\mathbf{1} + \varepsilon)$; $H_f(\ell) \leftarrow -\sum_j A_f(\ell,j) \log A_f(\ell,j) / \log N_t$; $E_f(i) \leftarrow \mathrm{mean}_\ell H_f(\ell)$
$E_{tf}(i) \leftarrow E_t(i)\,E_f(i)$
if $\neg$ isfinite($E_{tf}(i)$) **then** $E_{tf}(i) \leftarrow$ realmax
end if
return $E_{tf}(i)$
end function
if $m = 2$ **then**
$\hat{\imath} \leftarrow \arg\min\{\,\textsc{Eval}(1), \textsc{Eval}(2)\,\}$; $\hat{k} \leftarrow k_{\hat{\imath}}$; $\hat{E} \leftarrow 2^{\hat{k}}$ **return** $(\hat{E}, E_t, E_f, E_{tf}, \mathcal{J}, c)$
end if
EVAL(L); EVAL(L+1); EVAL(R); EVAL(R-1)
while $(L + 2 \le R) \ \wedge \ (R - 2 \ge L)$ **do**
if $L + 2 \le m$ **then** $e_L \leftarrow \textsc{Eval}(L)$; $e_M \leftarrow \textsc{Eval}(L+1)$; $e_R \leftarrow \textsc{Eval}(L+2)$
if $(e_M \le e_L + \tau) \ \wedge \ (e_M \le e_R + \tau)$ **then** $\hat{k} \leftarrow k_{L+1}$; $\hat{E} \leftarrow 2^{\hat{k}}$
return $(\hat{E}, E_t, E_f, E_{tf}, \mathcal{J}, c)$
end if
if $e_R < e_M - \tau$ **then** $L \leftarrow L + 1$; **continue**
end if
end if
if $R - 2 \ge 1$ **then** $e_R \leftarrow \textsc{Eval}(R)$; $e_M \leftarrow \textsc{Eval}(R-1)$; $e_L \leftarrow \textsc{Eval}(R-2)$
if $(e_M \le e_L + \tau) \ \wedge \ (e_M \le e_R + \tau)$ **then** $\hat{k} \leftarrow k_{R-1}$; $\hat{E} \leftarrow 2^{\hat{k}}$
return $(\hat{E}, E_t, E_f, E_{tf}, \mathcal{J}, c)$
end if
if $e_L < e_M - \tau$ **then** $R \leftarrow R - 1$; **continue**
end if
end if
if $L + 2 < R$ **then** $L \leftarrow L + 1$
end if
if $R - 2 > L$ **then** $R \leftarrow R - 1$
end if
end while
$\mathcal{C} \leftarrow \{\,\max(1, L-2), \ldots, \min(m, R+2)\,\}$
$\hat{\imath} \leftarrow \arg\min_{i \in \mathcal{C}} \textsc{Eval}(i)$; $\hat{k} \leftarrow k_{\hat{\imath}}$; $\hat{E} \leftarrow 2^{\hat{k}}$
return $(\hat{E}, E_t, E_f, E_{tf}, \mathcal{J}, c)$

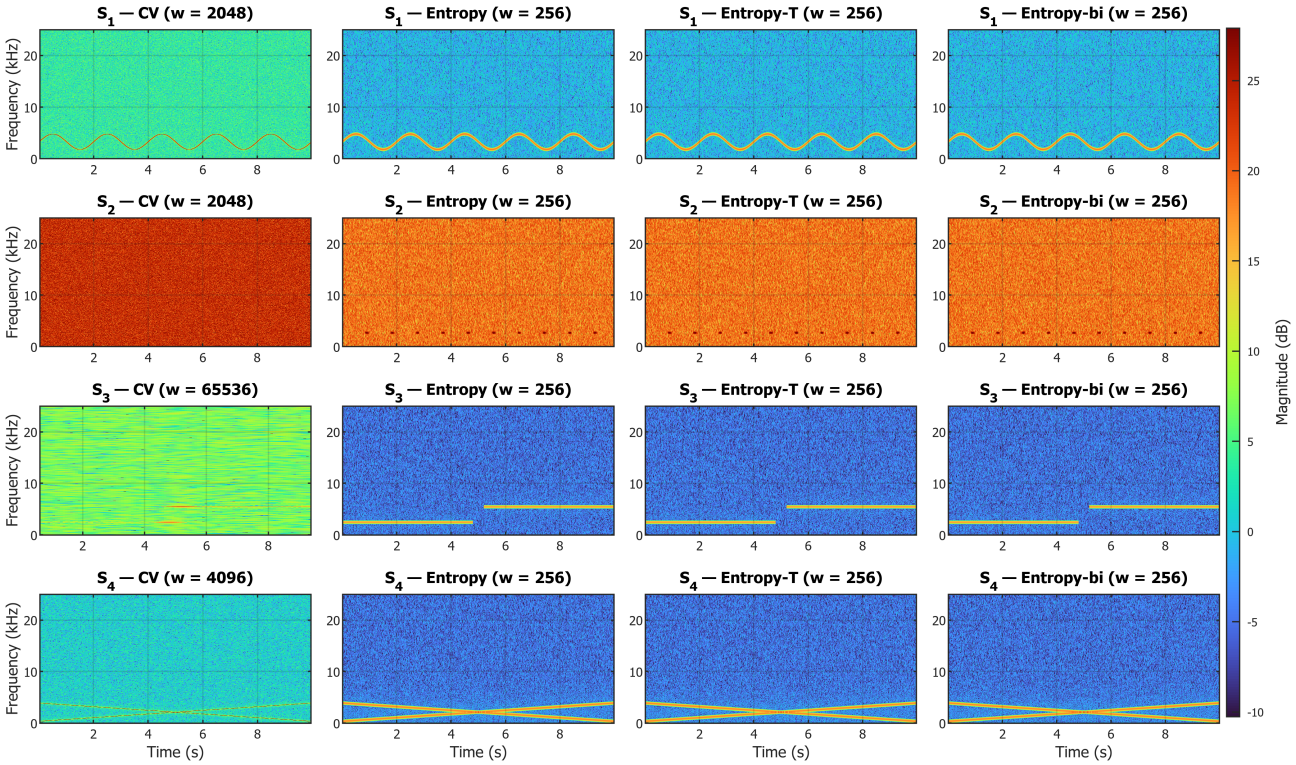


Figure 7.1: Time–frequency comparison of the four synthetic vibration signals using four window-selection algorithms: coefficient of variation (CV), entropy based full sweep search, entropy-based ternary search (Entropy-T), and bidirectional entropy search (Entropy-bi). Each tile displays the spectrogram obtained using the optimal window length returned by the respective algorithm.

7.3 Results

The results include spectrograms computed for both synthetic and MOIRA–UNIMORE vibration signals. Four synthetic one-dimensional test signals were meticulously constructed to encompass a diverse range of time–frequency behaviors characteristic of nonstationary vibration phenomena. All synthetic signals were generated at a sampling frequency of 50 kHz and had a duration of 10 s, ensuring sufficiently high temporal–spectral resolution for evaluating the proposed window-selection algorithms under controlled yet realistic conditions.

- **Signal S_1: Frequency-modulated tone with background noise.** This signal comprises a sinusoid with an instantaneous frequency that oscillates smoothly around a carrier frequency of approximately 3.3kHz, exhibiting a sinusoidal deviation and forming a curved ridge in the time-frequency plane. Additive white Gaussian noise was incorporated to simulate the measurement conditions and assess the robustness of the selection algorithms under mild spectral masking.

- **Signal S_2: Bursty narrowband packets.** The signal comprises broadband Gaussian noise, onto which short high-energy sinusoidal packets are periodically superimposed. A 2.7kHz tone was introduced at evenly spaced intervals throughout the 10s recording, with each burst persisting for approximately 0.2s. Each packet was tapered using a Hann window to mitigate spectral leakage and added with a substantial amplitude relative to the background noise to ensure clear time–frequency localization. The resulting spec-

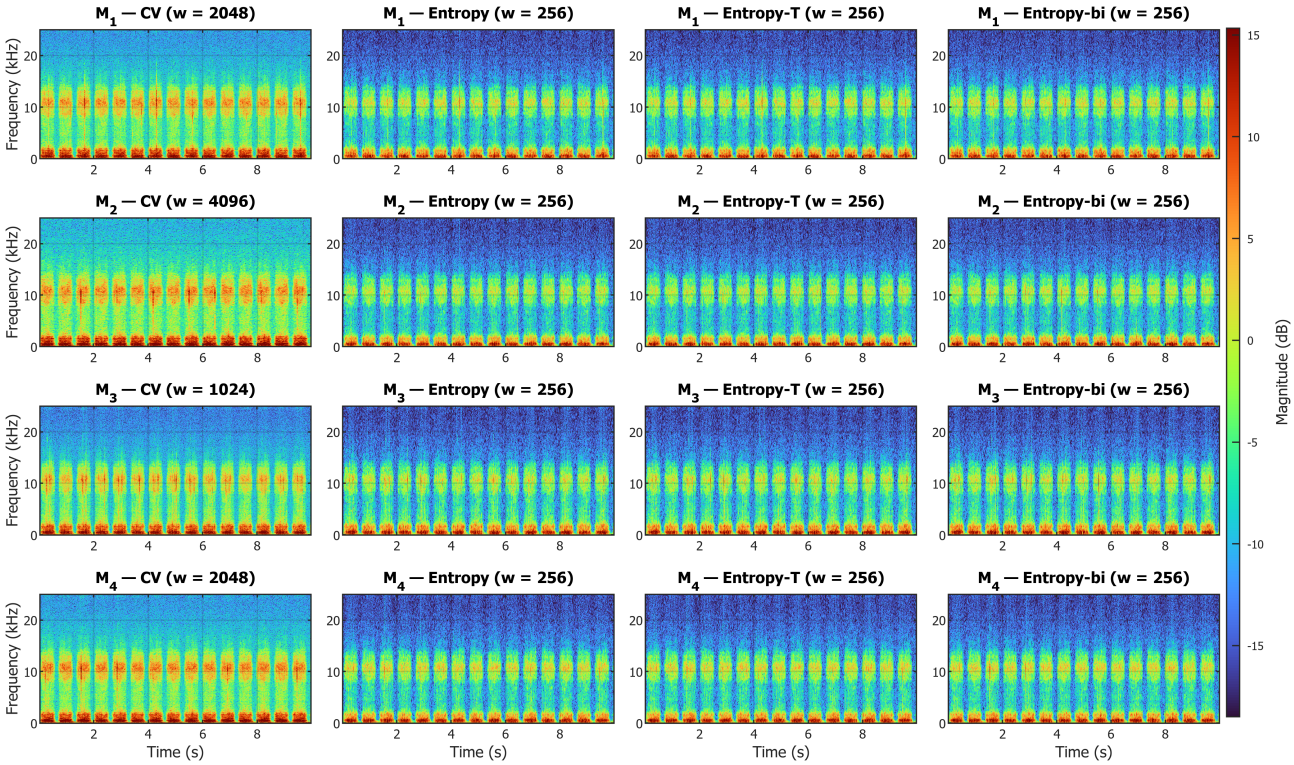


Figure 7.2: The spectrograms of the four MOIRA-UNIMORE vibration channels from four distinct fault classes, M_1–M_4, are analyzed using various window-selection strategies. The first column presents the coefficient of variation (CV) baseline along with its selected window w. The subsequent columns depict the entropy-based exhaustive method (Entropy), the ternary-search strategy (Entropy-T), and the bidirectional search (Entropy-bi), each utilizing their respective optimal window lengths.

trogram displays a sequence of intermittent narrowband bursts that manifest as isolated vertical ridges. This signal was designed to evaluate how various window-selection strategies manage transient, short-duration components, particularly in balancing temporal precision and frequency resolution.

- **Signal S_3: Piecewise-stationary step in frequency with a central gap.** This signal consists of a noisy low-frequency sinusoid (approximately 2.45kHz) in the first half of the record, followed by a silent or low-energy gap around the midpoint, and a higher-frequency sinusoid (approximately 5.45kHz) in the second half. This demonstrates a distinct step change in the dominant frequency and tests the efficacy of different windows in capturing abrupt spectral transitions and quiescent intervals.

- **Signal S_4: Crossing linear chirps.** The final signal contains two simultaneous linear chirps: one sweeping upward from approximately 350Hz to 3.9kHz and the other sweeping downward over the same range. Their superposition generates two crossing time-frequency ridges. This configuration presents a particular challenge for window selection, as it necessitates balancing time and frequency resolution to preserve both chirp trajectories without excessive smearing at the crossing region.

Collectively, these four synthetic cases encompass smoothly varying, transient, piecewise-stationary, and chirp-like behaviors, providing a controlled testbed for evaluating the sensitivity of the pro-

posed entropy- and CV-based window-selection methods to different time-frequency structures.

The spectrograms corresponding to the selected window lengths in Table 7.1 are depicted in Figures. 7.1 and 7.2 for the Synthetic and MOIRA datasets, respectively. These visual comparisons demonstrate that the entropy-based ternary search (E-T) and bidirectional search (E-Bi) yield spectrograms that are qualitatively indistinguishable from those produced by the exhaustive entropy baseline (E) while requiring significantly fewer window evaluations. Conversely, the coefficient-of-variation method (CV) frequently selects considerably larger window lengths, resulting in oversmoothing and a diminished temporal resolution. This substantiates the efficacy of the proposed algorithms in achieving both computational efficiency and high-quality time–frequency representations.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This dissertation presents a comprehensive and coherent framework for unsupervised bearing fault diagnosis within a specialized class of motion-control applications, specifically ICS. These systems represent a class of advanced industrial transport platforms characterized by hybrid translational-rotational dynamics, bidirectional motion, and persistent non-stationarity. The unique behavior of ICS bearings presents challenges that render traditional approaches to rotating machinery insufficient. The contributions of this thesis address these challenges at multiple levels of the diagnostic pipeline, including data acquisition, feature representation, distribution reshaping, time-frequency parameter selection, and computational optimization.

8.1.1 Summary of Contributions

8.1.1.1 Open-Access Dataset and Foundational Analysis

The thesis begins with the introduction of the MOIRA–UNIMORE bearing dataset, which constitutes the first open-access dataset to capture the dynamics of ICS bearings under controlled single-cart and three-cart experimental configurations. This dataset includes a wide range of fault types, severities, and operating modes, providing a crucial foundation for the research community. Classical statistical features and dimensionality reduction techniques, such as PCA and t-SNE, were utilized to assess fault separability. These analyses highlight the limitations of existing representations, particularly in the context of multiple bearings per cart and overlapping fault distributions, thus motivating the development of improved feature transformation methods.

8.1.1.2 Unsupervised Distribution-Reshaping Transformations

Multiple novel transformation frameworks have been proposed to reshape high-variance, multimodal, and density-imbalanced feature distributions without the need for labels or training.

- The first transformation, BDD–T, introduces a lightweight unsupervised mapping designed to address bimodality and improve the separability between healthy and faulty classes. This method was incorporated into a training-free diagnostic framework for ICS, where translational–rotational interactions produce highly nonstationary vibration signa-

tures. It effectively restructures overlapping feature distributions into compact and well-separated clusters within a low-dimensional PCA space. When combined with statistical feature extraction and feature ranking, BDD–T significantly enhances the performance of one-class SVM and Isolation Forest, achieving high recall and F1-scores, particularly for subtle faults such as FIR025 and FOR025, while maintaining robust specificity. This approach is interpretable, adaptable to various domains, and computationally efficient, making it suitable for real-time, label-scarce industrial environments. Its deterministic design facilitates scalable deployment across multiple ICS units, requiring only minor parameter adjustments for different operational profiles.

- The adaptive variant, ADT, enhances BDD–T by integrating a variance-adaptive contraction mechanism, thereby improving robustness across varying noise levels, operational speeds, and multi-cart conditions. This mechanism effectively transforms skewed, bimodal, and highly dispersed feature distributions into compact, well-separated clusters while maintaining the interpretability of fault severity progression. ADT consistently improves separability in PCA and t-SNE embeddings, surpasses traditional and multimodality-aware transforms in performance under one-class SVM and Isolation Forest, and demonstrates resilience to broadband noise, narrowband interference, and structural disturbances. Validation of both single-cart and three-cart configurations confirmed their scalability and practical applicability in label-scarce, high-dimensional, and strongly nonstationary industrial environments.

- A lightweight, unsupervised diagnostic framework was developed to detect bearing faults in independent cart systems operating under nonstationary conditions. Central to this approach, the IWCCT reshapes feature distributions derived from statistical and Hjorth descriptors into compact and well-separated clusters without the need for labels or training. Unlike Gaussianization-based transforms that enforce rigid marginal structures and may suppress discriminative patterns, the IWCCT preserves the local geometry by contracting samples toward weighted compact centers, thereby enhancing separability while retaining informative variability. Across PCA and t-SNE embeddings and multiple anomaly detectors, the IWCCT consistently improved the F1-scores and produced tight, minimally overlapping clusters. Noise stress experiments demonstrated that the framework maintained cluster compactness, highlighting its robustness beyond the clean or quasi-stationary regime. Its deterministic and computationally efficient formulation makes it well suited for real-time, label-scarce industrial applications.

These transformations collectively address the primary limitations of traditional feature descriptors, offering a mathematically substantiated approach for developing compact, geometry-preserving representations for unsupervised learning.

8.1.1.3 Intelligent Time–Frequency Analysis and DRBT Pipeline

To address the nonstationarity of ICS vibrations in the time-frequency domain, this framework introduces the TAIW method. This approach autonomously determines the optimal window length by utilizing the Theil index, thereby eliminating the reliance on practitioner heuristics and ensuring consistent representations across rapidly fluctuating speed profiles. Building on these representations, the DRBT algorithm was proposed to transform highly irregular, density-imbalanced embeddings into compact clusters. DRBT demonstrates superior performance compared to classical power transforms, Gaussianization, kernel-PIT, and CDF-TS by

preserving the underlying class geometry while mitigating density-induced distortions. Cross-dataset validation using the CWRU and Politecnico di Torino datasets confirmed the generality and scalability of the TAIW+DRBT pipeline.

8.1.1.4 Entropy-Based Spectrogram Window Selection

An entropy-driven framework for the automatic selection of spectrogram windows is also introduced in this study. An exhaustive entropy-based search is established as a statistically grounded baseline, and two accelerated strategies, namely ternary and bidirectional searches, are proposed to mitigate computational demands. These accelerated strategies consistently achieve near-optimal window lengths while significantly reducing the number of spectrogram evaluations, thereby facilitating scalable time-frequency analysis across datasets.

8.1.2 Overall Impact

This thesis introduces a comprehensive dataset for ICS-bearing faults, alongside a cohesive set of unsupervised, training-free, and computationally efficient methodologies that collectively address significant challenges in ICS diagnostics throughout all chapters. These challenges encompass:

- absence of publicly available datasets for bearing-fault analysis in ICS,
- nonstationary translational–rotational dynamics,
- severe class overlap and high-dimensional feature spaces,
- label scarcity in industrial environments,
- computational constraints for real-time monitoring,
- robustness requirements under measurement noise and structural interference.

The methodological innovations introduced in this study are distinguished by their flexibility, interpretability, and applicability not only to ICS but also to *conventional rotating machinery*, where bearings undergo purely rotational motion. This cross-domain applicability is substantiated by the consistent performance of the proposed methods on external datasets such as the CWRU and Politecnico di Torino datasets. This dissertation demonstrates that unsupervised diagnostic techniques, when integrated with principled distribution-reshaping transformations, can achieve high reliability, even in complex, nonstationary environments. The contributions presented in this study establish a robust foundation for next-generation industrial monitoring systems and open multiple avenues for continued scientific and technological advancement.

8.2 Future Work

Based on the methodologies established in this dissertation, several promising avenues for future research warrant exploration.

8.2.1 Multiclass and Joint-Class Transformations

The proposed transformations are designed to operate on individual classes, which aligns naturally with models that focus solely on healthy data. Extending these models to simultaneously transform multiple classes would facilitate the following:

- the concurrent compactification of various fault severities,
- enhanced boundary delineation in regions of overlap,
- direct applicability to multiclass unsupervised learning.

8.2.2 Streaming and Online Adaptation

The real-time monitoring of ICS necessitates the development of incremental and adaptive methodologies. Future research may focus on the following areas:

- the creation of online variants of the proposed pipelines,
- the development of rolling-window KDE approximations suitable for low-memory environments.

8.2.3 Fault Localisation, Multichannel Fusion, and Spatial Reasoning

The ICS environment presents distinct diagnostic challenges due to its multi-cart operation, spatially distributed bearings, and the coexistence of translational and rotational dynamics. A significant future research direction involves the development of methods capable of precise fault localization across the following levels:

- *Cart-level localisation:* Identifying which cart within a multi-cart experiment contains the faulty bearing.
- *Bearing-level localisation:* Determining which bearing on the faulty cart is defective, including the three-bearing configuration (two identical upper bearings and one lower bearing).
- *Fault-type discrimination:* Distinguishing among inner-race, outer-race, and rolling-element defects.
- *Fault localization within identical bearing pairs:* Reliably determining which of the two geometrically identical upper bearings (left or right) is faulty, despite their symmetric geometrical placement and similar vibration signatures.

Achieving such spatial resolution will likely require the implementation of *multichannel fusion* and the imposition of cross-sensor consistency constraints. Future research should investigate the integration of vibration, encoder, and current signals, along with a comprehensive understanding of the cart–rail layout and sensor placement, to develop spatially informed diagnostic models. This approach holds significant potential for advancing ICS from system-level anomaly detection to comprehensive fault localization and component-level health assessments.

8.2.4 Integration with Self-Supervised and Few-Shot Models

The field of modern industrial diagnostics increasingly benefits from advancements in representation learning. The proposed transformations can be effectively integrated with:

- contrastive learning to preserve feature geometry,
- masked-signal prediction to maintain temporal coherence,
- few-shot and meta-learning models to minimize data requirements.

8.2.5 Hardware Deployment and Real-Time Optimization

The practical implementation of the proposed methods in industrial settings necessitates their operation within the temporal constraints of embedded controllers. Future research should focus on:

- *GPU and FPGA implementations* of the proposed frameworks to enhance the speed of FFTs, KDE evaluations, and transformation processes for real-time applications,
- *Memory-efficient KDE approximations* (e.g., quantized grids or lookup tables) that are appropriate for devices with limited resources,
- *Low-latency update mechanisms* to ensure that diagnostic outputs are synchronized with cart motion and vibration inputs with minimal delay.

8.2.6 Extensions Beyond Independent Cart Systems

The methodologies developed in this study have the potential to be generalized to a wide range of systems, including:

- robotic joints and collaborative robots,
- high-speed manufacturing lines,
- automotive and aerospace test benches,
- mechatronic and cyber-physical systems subject to nonstationary dynamics.

Overall, the contributions of this thesis provide a robust foundation for future advancements in unsupervised diagnostic algorithms, adaptive signal analysis, and intelligent machinery health monitoring.

Appendix A

File and Data Export Guidelines

A.1 Data Variables Naming Convention

A.1.1 System Variables

- In experiments involving a single cart, variables are not appended with cart identifiers.
- In experiments involving three carts, variables are appended with `_M1`, `_M2`, and `_M3` to denote individual carts.

A.1.2 Vibration Channels

A.1.2.1 Mono-Axial Accelerometers

- **mono_PCB_Top:** This variable refers to a PCB-manufactured accelerometer mounted on the upper section of the guide rail, positioned at the midpoint along the track.
- **mono_ifm_Top:** This variable denotes an ifm-manufactured accelerometer located near the upper section of the guide rail, mounted on an aluminum rod at the midpoint along the track.
- **mono_PCB_Bottom:** This variable indicates a PCB-manufactured accelerometer mounted on the lower section of the guide rail, positioned at the midpoint along the track.

A.1.2.2 Tri-Axial Accelerometers

- These sensors capture vibrations in the X, Y, and Z directions and are mounted on the left and right sides of the guide.

A.1.3 File Naming Conventions for Experimental Data

The file naming convention systematically encodes the details of each experiment, facilitating the identification of the experiment type, fault conditions, speed profiles, and other critical parameters. The file names were categorized into two primary groups: single-mover experiments

Table A.1: Data variable classification and naming convention by experiment type.

Category	Variable Type	Single-Cart Experiments	Three-Cart Experiments
System Variables	Position	ActHwPos	ActHwPos_M1 ActHwPos_M2 ActHwPos_M3
	Following Error	ActFollowingError	ActFollowingError_M1 ActFollowingError_M2 ActFollowingError_M3
	Velocity	ActVelo	ActVelo_M1 ActVelo_M2 ActVelo_M3
	Velocity Error	ActVeloError	ActVeloError_M1 ActVeloError_M2 ActVeloError_M3
	Set Current	SetCurr	SetCurr_M1 SetCurr_M2 SetCurr_M3
Vibration Channels	Mono-Axial	mono_PCB_Top mono_ifm_Top mono_PCB_Bottom	mono_PCB_Top mono_ifm_Top mono_PCB_Bottom
	Tri-Axial	X_Guide_Left Y_Guide_Left Z_Guide_Left X_Guide_Right Y_Guide_Right Z_Guide_Right	X_Guide_Left Y_Guide_Left Z_Guide_Left X_Guide_Right Y_Guide_Right Z_Guide_Right

(Experiment Types 1–4) and three-cart experiments (Experiment Types 5–8). The convention adheres to a structured format:

A.1.3.1 Experiment Type 1–4

<Fault Condition>_<Experiment Category>_<Experiment Description>_<Speed>_<Fault Details>_<Fault Location>_<Trial Number>

A.1.3.2 Experiment Type 5–8

<Fault Condition>_<Experiment Category>_<Experiment Description>_<Separation Details>_<Speed>_<Fault Details>_<Fault Location>_<Trial Number>

A.1.4 Description of File Names

Each component of the file names provides specific information regarding the associated experiment, including the type of experiment, fault conditions, and speed profiles. The meaning of each component is systematically encoded and can be interpreted as follows:

A.1.4.1 Fault Condition

- **H**: No Fault (Healthy).
- **F**: Bearing Fault.

A.1.4.2 Experiment Category

- **SingleMover**: Single-cart experiments.
- **3Movers**: Three-cart experiments.

A.1.4.3 Experiment Description

- **StraightTop**: Straight path back-and-forth movement along the upper section of the guide rail.
- **Curved**: Back-and-forth movement limited to the curved module only.
- **CompleteRotation**: Circular movement along the track.

A.1.4.4 Separation Details

- **150mmSeparation**: The relative distance between two carts is 150 mm.
- **250mmSeparation**: The relative distance between two carts is 250 mm.
- **500mmSeparation**: The relative distance between two carts is 500 mm.
- **750mmSeparation**: The relative distance between two carts is 750 mm.

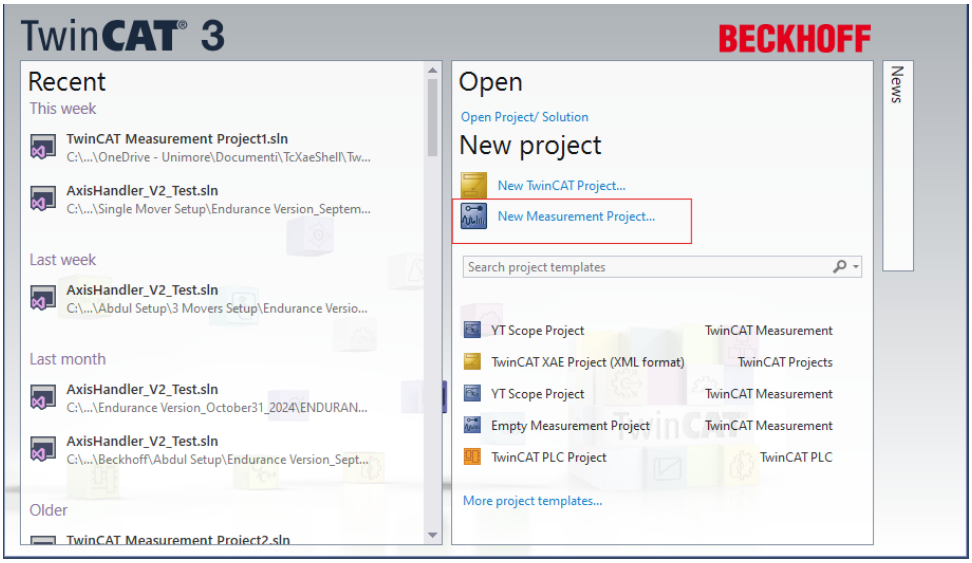
A.1.4.5 Speed

- **1000**: Nominal cart speed profile of 1000 mm/s.
- **2000**: Nominal cart speed profile of 2000 mm/s.

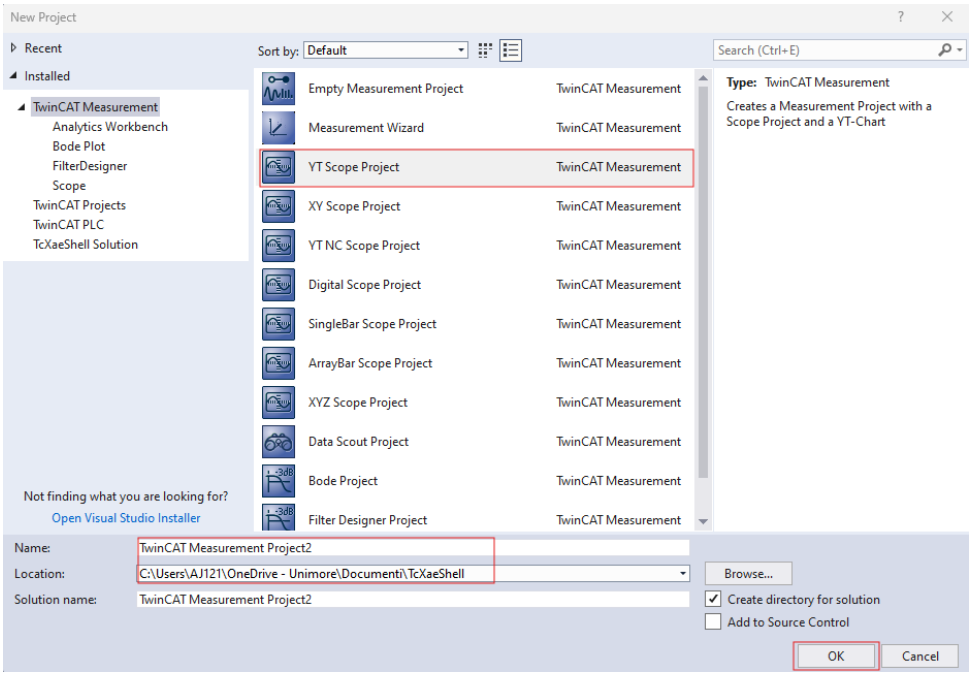
A.1.5 Creating and Exporting a YT Scope Project in TwinCAT XAE Shell

TwinCAT, developed and licensed by Beckhoff, stands for Windows Control and Automation Technology. It is an advanced automation software that transforms nearly any PC-based system into a real-time control platform, supporting multiple runtime PLC, NC, CNC, and robotics systems. The latest version, TwinCAT 3, was used to record the dataset in this study. TwinCAT 3 is freely available for download from the Beckhoff website ((Although it requires a weekly renewal of the license, this process is straightforward and cost-free. Users simply enter an automatically generated code into a dialog box to extend their license.

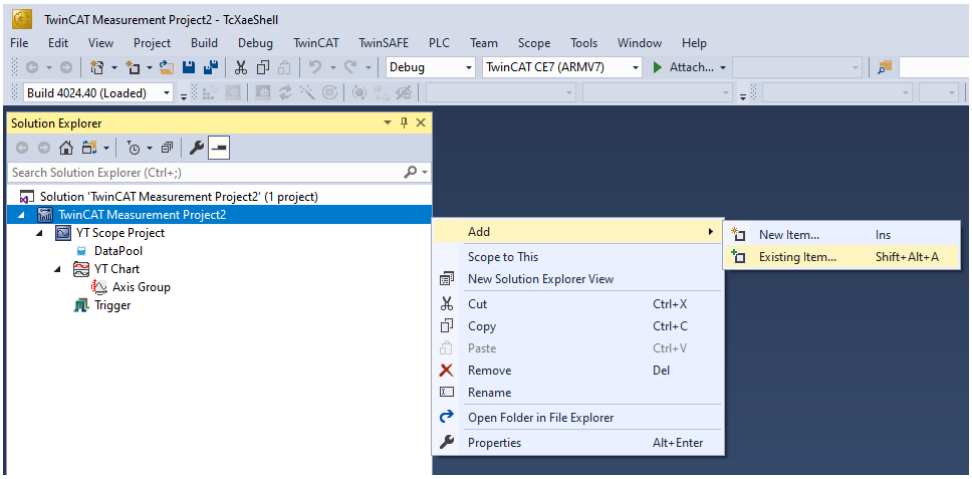
Step 1—Initiate a New Measurement Project: Launch the **TwinCAT XAE shell** and select "New Measurement Project" to commence the process.



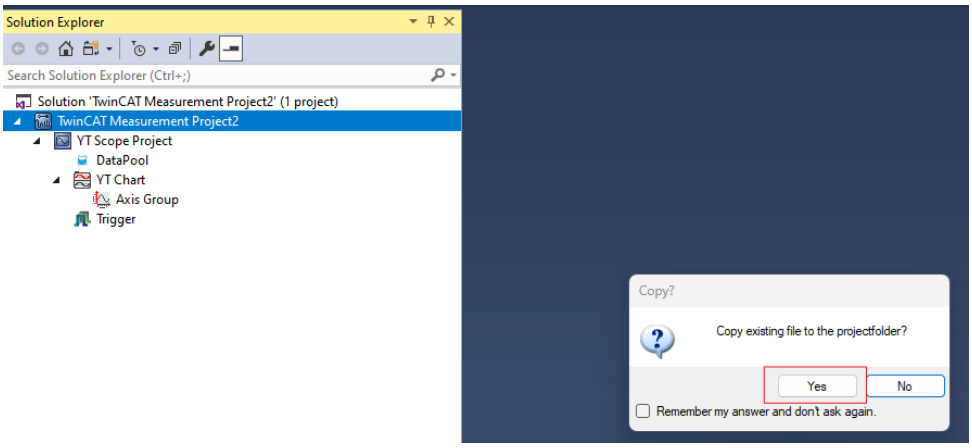
Step 2—Designate and Save the Project: Select the type of measurement project, specifically the **YT Scope Project** in this instance. Assign a name to the project and determine the save directory.



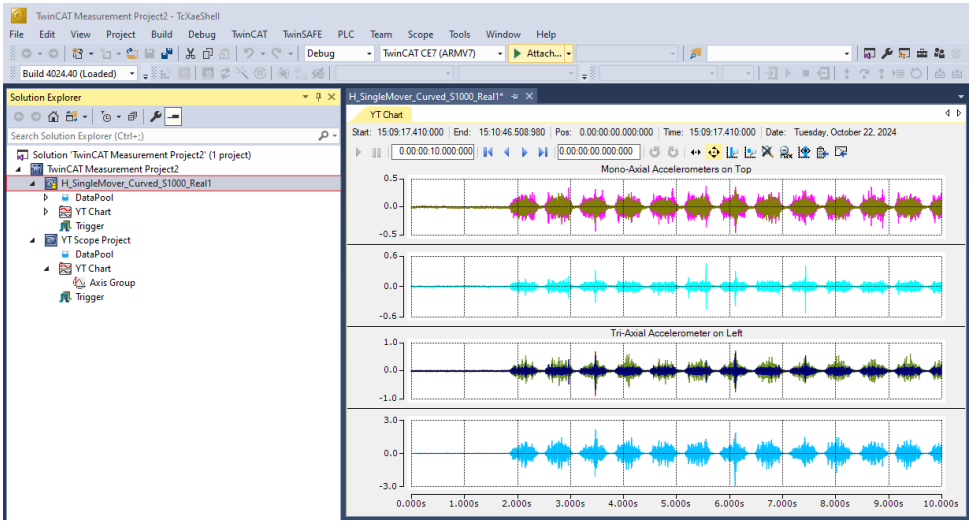
Step 3—Incorporate Existing Files into the Project: Upon project creation, incorporate existing **".svdx"** files from the save directory, as illustrated in the figure. Multiple files may be added concurrently; however, for this example, a single file is loaded.



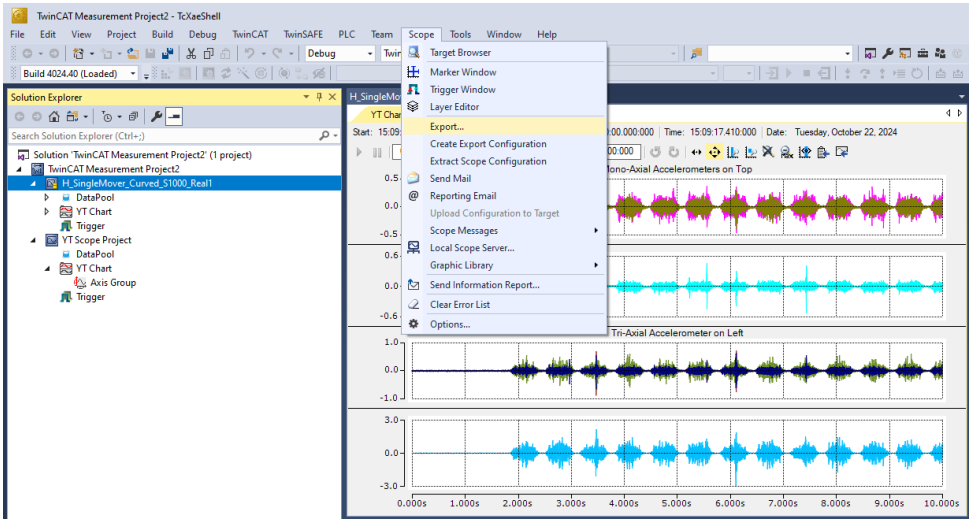
Step 4—Transfer Files to the Project Directory: Decide whether to transfer the selected files to the project directory. This is advisable to do so for more efficient management. Once the files are converted to the desired format, the original **".svdx"** files may be deleted to conserve memory.



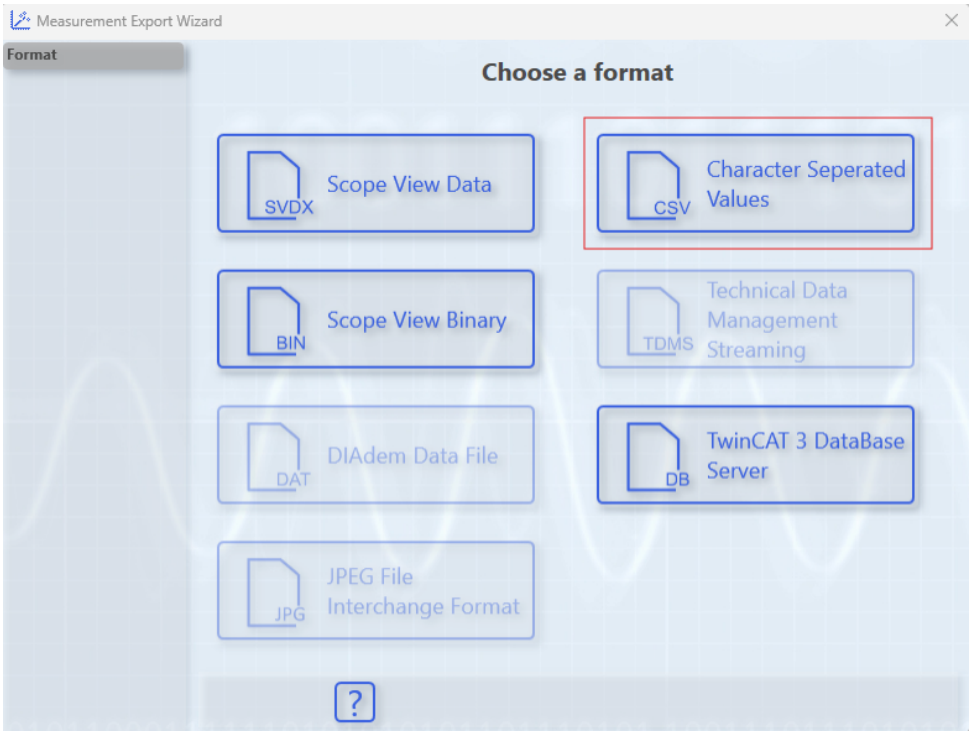
Step 5—Visualize the YT Scope Project: To visualize the data channels, left-click on the **".svdx"** file within the project tree.



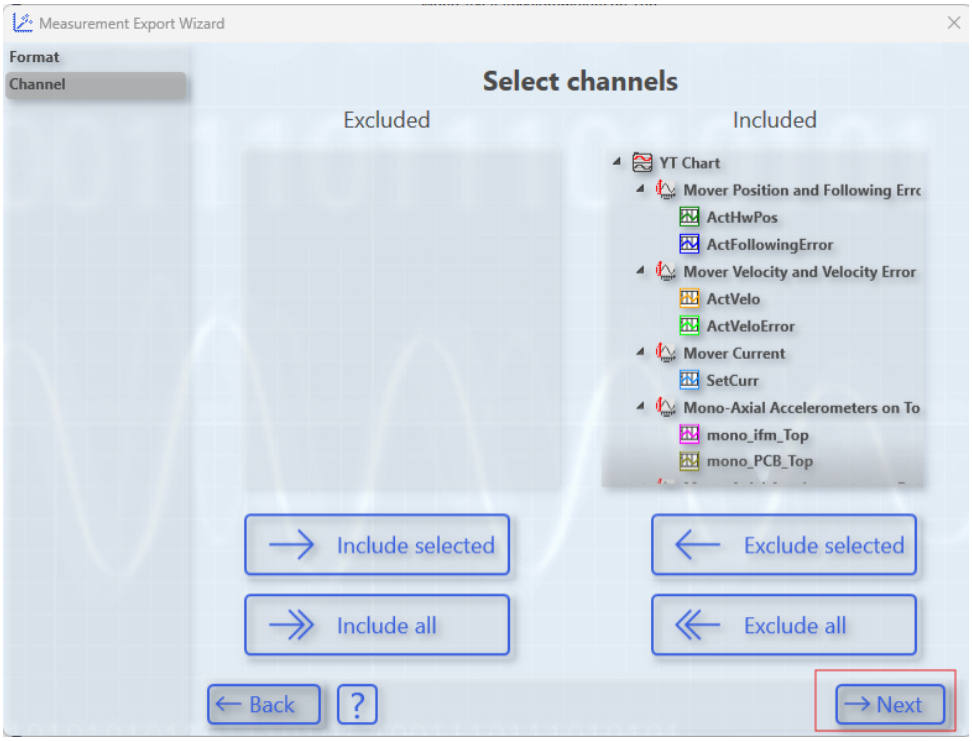
Step 6—Export the Project: To export the project in the desired format, navigate to the Scope tab in the options bar at the top and select **Scope** –> **Export**.



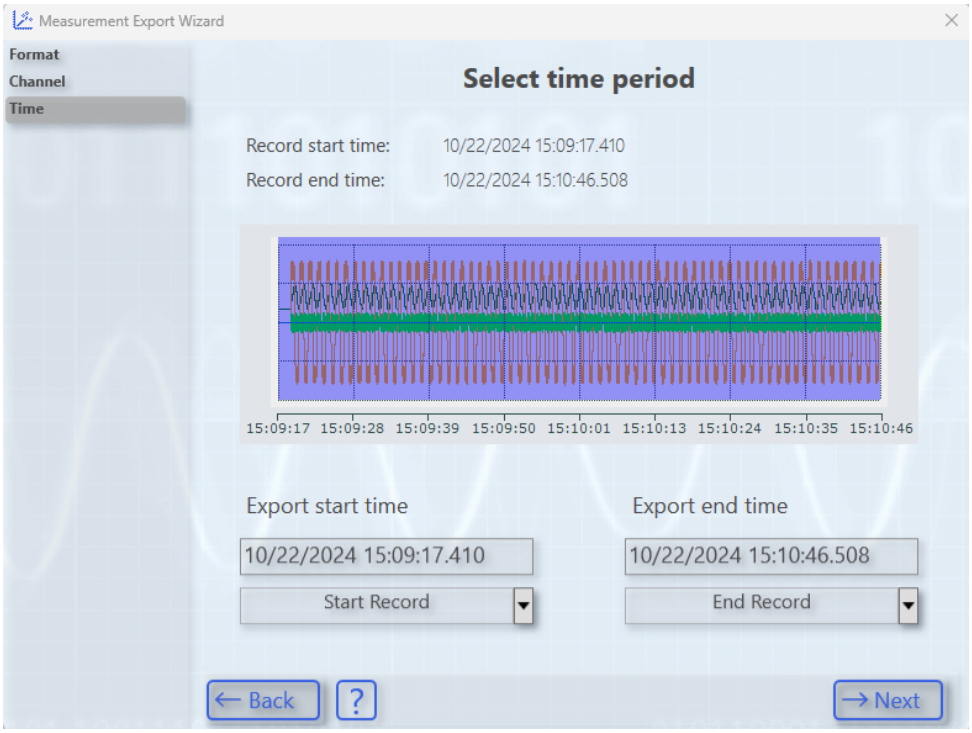
Step 7—Select Format: A new window will appear, guiding the user through the format conversion steps. The desired data format was selected. In this example, the **Character Separated Values (CSV)** format is chosen.



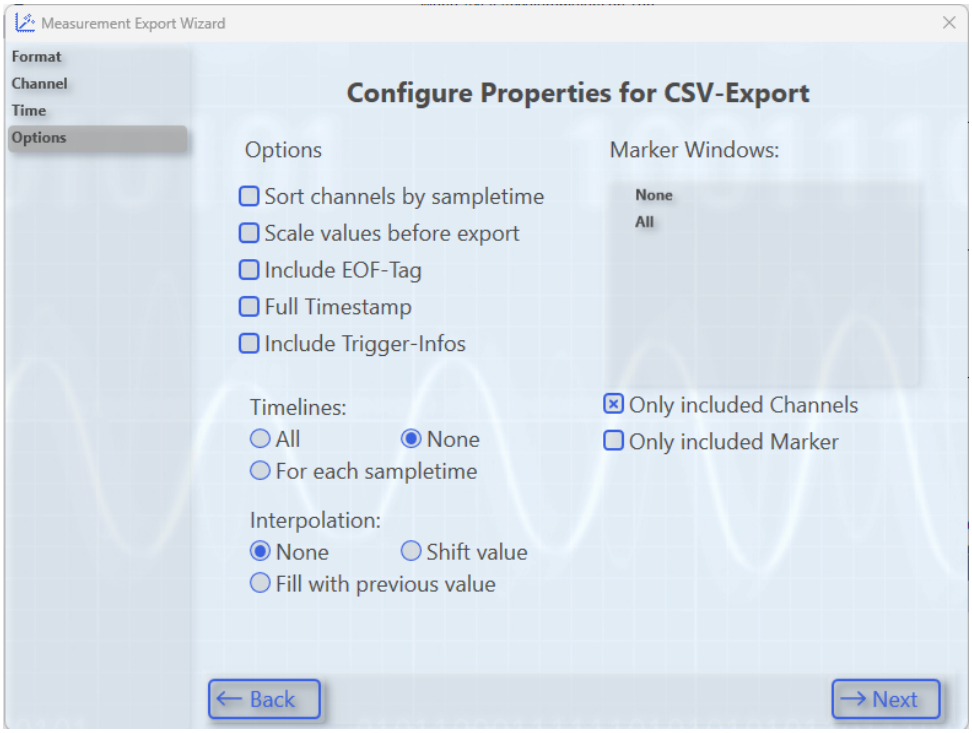
Step 8—Select Channels: Following the selection of the format, identify the data channels intended for export and proceed by clicking **Next**.



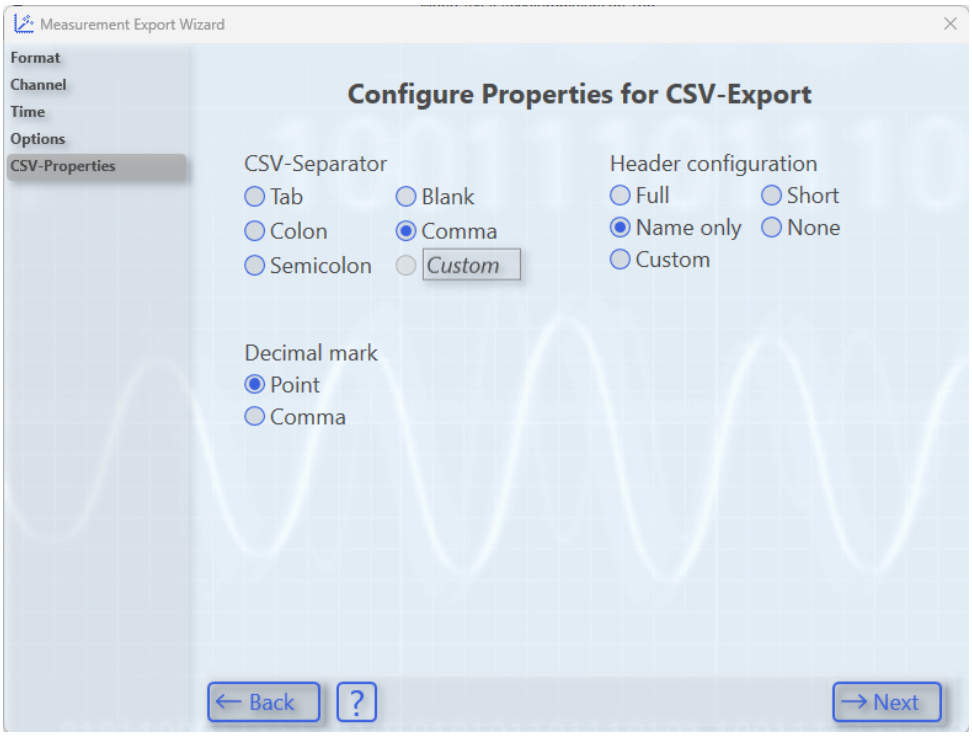
Step 9—Select Time Duration: Define the time range for data conversion. In this instance, **End of Record** is selected.



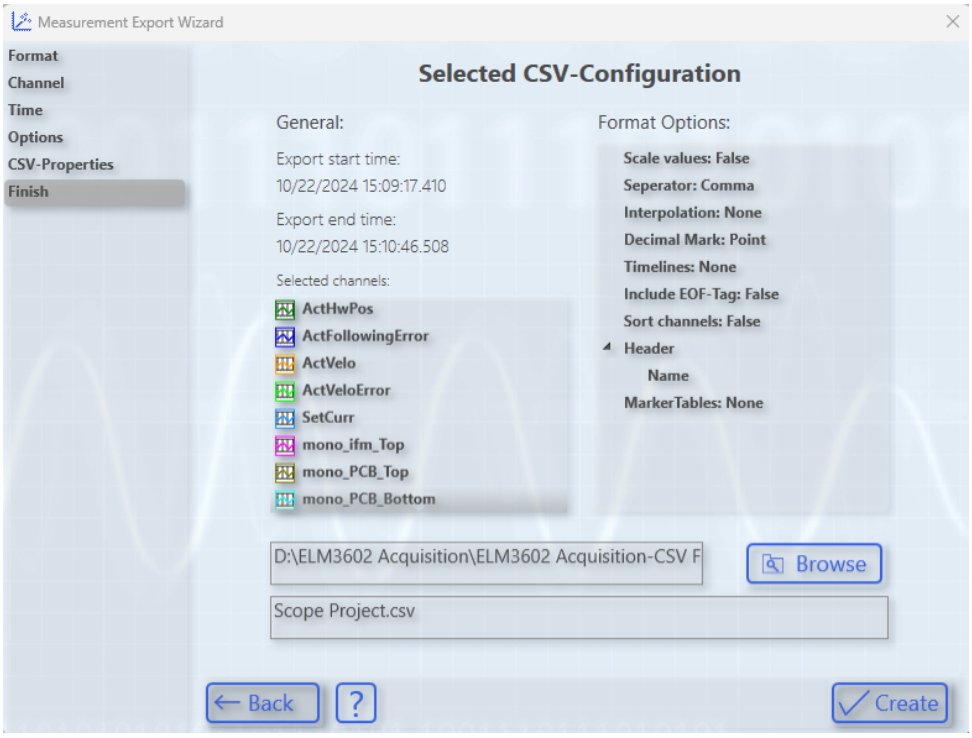
Step 10—Configure CSV Options: Determine the metadata settings for export. In this example, **Only Included Channels** is selected under Marker Windows, with **None** chosen for both Timelines and Interpolation.



Step 11—Configure CSV Properties: Adjust CSV-specific properties, such as the **Separator** and **Decimal Mark**, according to your preferences.



Step 12—Name and Save File: Conclude the export process by assigning a name to the converted file and selecting a save directory.


Measurement Export Wizard
Format
Channel
Time
Options
CSV-Properties
Finish
Selected CSV-Configuration
General:
Export start time:
10/22/2024 15:09:17.410
Export end time:
10/22/2024 15:10:46.508
Selected channels:
ActHwPos
ActFollowingError
ActVelo
ActVeloError
SetCurr
mono_ifm_Top
mono_PCB_Top
mono_PCB_Bottom
Format Options:
Scale values: False
Seperator: Comma
Interpolation: None
Decimal Mark: Point
Timelines: None
Include EOF-Tag: False
Sort channels: False
Header
Name
MarkerTables: None
D:\ELM3602 Acquisition\ELM3602 Acquisition-CSV F
Browse
Scope Project.csv
Back
?
Create

Bibliography

Alhams, A., Abdelhadi, A., Badri, Y., Sassi, S., & Renno, J. (2024). Enhanced bearing fault diagnosis through trees ensemble method and feature importance analysis. *Journal of Vibration Engineering & Technologies*, *12*(Suppl 1), S109–S125. https://doi.org/10.1007/s42417-024-01405-0

Allen, J., & Rabiner, L. (1977). A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, *65*(11), 1558–1564. https://doi.org/10.1109/PROC.1977.10770

An, Y., Zhang, K., Chai, Y., Liu, Q., & Huang, X. (2022). Bearing fault diagnosis under variable working conditions base on contrastive domain adaptation method. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–11. https://doi.org/10.1109/TIM.2022.3200106

Antoni, J. (2009). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, *23*(4), 987–1036. https://doi.org/https://doi.org/10.1016/j.ymssp.2008.10.010

Batzner, K., Heckler, L., & König, R. (2024). Efficientad: Accurate visual anomaly detection at millisecond-level latencies. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 127–137. https://doi.org/10.1109/WACV57701.2024.00020

Bienefeld, C., Becker-Dombrowsky, F. M., Shatri, E., & Kirchner, E. (2023). Investigation of feature engineering methods for domain-knowledge-assisted bearing fault diagnosis. *Entropy*, *25*(9). https://doi.org/10.3390/e25091278

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Bristi, S. D., Tatha, M. J., Ali, M. F., Bhatti, U. A., Sarker, S. K., Masud, M., Ghadi, Y. Y., Algarni, A., & Saha, D. K. (2023). A meta-heuristic sustainable intelligent internet of things framework for bearing fault diagnosis of electric motor under variable load conditions. *Sustainability*, *15*(24), 16722. https://doi.org/10.3390/su152416722

Brito, L. C., Susto, G. A., Brito, J. N., & Duarte, M. A. V. (2021). Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction. *Informatics*, *8*(4). https://doi.org/10.3390/informatics8040085

Cavalaglio Camargo Molano, J., Capelli, L., Rubini, R., Borghi, D., & Cocconcelli, M. (2020). A bearing fault model for independent cart conveyor system and its validation. *Applied Acoustics*, *159*, 107069. https://doi.org/https://doi.org/10.1016/j.apacoust.2019.107069

Chen, B., Zhang, W., Xi Gu, J., Song, D., Cheng, Y., Zhou, Z., Gu, F., & Ball, A. D. (2023). Product envelope spectrum optimization-gram: An enhanced envelope analysis for rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *193*, 110270. https://doi.org/https://doi.org/10.1016/j.ymssp.2023.110270

Chen, M.-C., Hsu, C.-C., Malhotra, B., & Tiwari, M. K. (2016). An efficient ica-dw-svdd fault detection and diagnosis method for non-gaussian processes. *International Journal of*

Production Research, *54*(17), 5208–5218. https://doi.org/10.1080/00207543.2016.1161250

Chen, T., Guo, L., Gao, H., Wang, D., Feng, T., & Yu, Y. (2025). Investigations on improved box-cox sparsity measures for machine condition monitoring. *ISA Transactions*, *157*, 466–480. https://doi.org/https://doi.org/10.1016/j.isatra.2024.12.010

Chen, Y., Zhang, D., Yan, R., & Xie, M. (2025). Applications of domain generalization to machine fault diagnosis: A survey [Early Access]. *IEEE/CAA Journal of Automatica Sinica*, *12*(3), 1–15. https://doi.org/10.1109/JAS.2025.125120

Chu, T., Nguyen, T., Yoo, H., & Wang, J. (2024). A review of vibration analysis and its applications. *Heliyon*, *10*(5), e26282. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e26282

Coble, J., & Hines, J. (2009). Identifying optimal prognostic parameters from data: A genetic algorithms approach [28 September–1 October]. *Proceedings of the Annual Conference of the Prognostics and Health Management Society*.

Cocconcelli, M., Cavalaglio Camargo Molano, J., Rubini, R., Capelli, L., & Borghi, D. (2019). 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* (pp. 211–220). Springer International Publishing.

CWRU. (2005). Case western reserve university (cwru) bearing data center [Accessed on 20 January 2025].

Daga, A. P., Fasana, A., Marchesiello, S., & Garibaldi, L. (2019). The politecnico di torino rolling bearing test rig: Description and analysis of open access data. *Mechanical Systems and Signal Processing*, *120*, 252–273. https://doi.org/10.1016/j.ymssp.2018.10.010

Das, O., Bagci Das, D., & Birant, D. (2023). Machine learning for fault analysis in rotating machinery: A comprehensive review. *Heliyon*, *9*(6), e17584. https://doi.org/10.1016/j.heliyon.2023.e17584

Defard, T., Setkov, A., Loesch, A., & Audigier, R. (2021). 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* (pp. 475–489). Springer International Publishing.

Fan, H., Ren, Z., Cao, X., Zhang, X., & Huang, J. (2024). A gti&ada-act lmcnn method for intelligent fault diagnosis of motor rotor-bearing unit under variable conditions. *IEEE Transactions on Instrumentation and Measurement*, *73*, 3508314. https://doi.org/10.1109/TIM.2024.3351265

Farhat, M. H., Chiementin, X., Chaari, F., Bolaers, F., & Haddar, M. (2021). Order-based identification of bearing defects under variable speed condition. *Applied Sciences*, *11*(9). https://doi.org/10.3390/app11093962

Feng, Z., Chen, X., & Wang, T. (2017). Time-varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions. *Journal of Sound and Vibration*, *400*, 71–85. https://doi.org/10.1016/j.jsv.2017.03.037

Fisher, R. (1925). *Statistical methods for research workers*. Oliver; Boyd.

Fu, S., Wu, Y., Wang, R., & Mao, M. (2023). A bearing fault diagnosis method based on wavelet denoising and machine learning. *Applied Sciences*, *13*(10), 5936. https://doi.org/10.3390/app13105936

Gambardella, G. (1968). Time scaling and short-time spectral analysis. *Journal of the Acoustical Society of America*, *44*(6), 1496–1503.

GFX-Guidance. (2025). Hepco motion. gfx—guidance system for beckhoff xts linear transport system [Accessed on 20 January 2025].

Giantomassi, A., Ferracuti, F., Iarlori, S., Ippoliti, G., & Longhi, S. (2015). Electric motor fault detection and diagnosis by kernel density estimation and kullback–leibler divergence based on stator current measurements. *IEEE Transactions on Industrial Electronics*, *62*(3), 1770–1780. https://doi.org/10.1109/TIE.2014.2370936

Gove, R., Cadalzo, L., Leiby, N., Singer, J., & Zaitzeff, A. (2022). New guidance for using t-sne: Alternative defaults, hyperparameter selection automation, and comparative evaluation. *Visual Informatics*, *6*, 87–97. https://doi.org/10.1016/j.visinf.2022.04.003

Hakim, M., Omran, A. A. B., Ahmed, A. N., Al-Waily, M., & Abdellatif, A. (2023). A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Engineering Journal*, *14*, 101945. https://doi.org/10.1016/j.asej.2022.101945

Hamadache, M., Lee, D., & Veluvolu, K. C. (2015). Rotor speed-based bearing fault diagnosis (rsb-bfd) under variable speed and constant load. *IEEE Transactions on Industrial Electronics*, *62*(10), 6486–6495. https://doi.org/10.1109/TIE.2015.2416673

Hasan, M. J., Islam, M. M. M., & Kim, J.-M. (2019). Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions. *Measurement*, *138*, 620–631. https://doi.org/10.1016/j.measurement.2019.02.075

He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. *Advances in Neural Information Processing Systems*, *18*, 507–514.

Healy, J., & McInnes, L. (2024). Uniform manifold approximation and projection. *Nature Reviews Methods Primers*, *4*, 82. https://doi.org/10.1038/s43586-024-00363-x

Huang, W., Gao, G., Li, N., Jiang, X., & Zhu, Z. (2019). Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *68*(8), 2819–2829. https://doi.org/10.1109/TIM.2018.2868519

Huang, Y., Xu, Z., Cao, L., Hu, H., & Tang, G. (2022). Fractional dimensionless indicator with random forest for bearing fault diagnosis under variable speed conditions. *Shock and Vibration*, *2022*, 1–14. https://doi.org/10.1155/2022/1781340

Islam, M. R., Kim, Y.-H., Kim, J.-Y., & Kim, J.-M. (2019). Detecting and learning unknown fault states by automatically finding the optimal number of clusters for online bearing fault diagnosis. *Applied Sciences*, *9*(11). https://doi.org/10.3390/app9112326

iTRAK. (2025). Independent cart technology, rockwell automation. https://www.rockwellautomation.com

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

Jabbar, A., Cocconcelli, M., D'Elia, G., Borghi, D., Capelli, L., Cavalaglio Camargo Molano, J., Strozzi, M., & Rubini, R. (2025a). Moira-unimore bearing data set for independent cart systems. *Applied Sciences*, *15*(7). https://www.mdpi.com/2076-3417/15/7/3691

Jabbar, A., Cocconcelli, M., D'Elia, G., Borghi, D., Capelli, L., Cavalaglio Camargo Molano, J., Strozzi, M., & Rubini, R. (2025b). Moira-unimore bearing data set for independent cart systems—experiment type 5 and 6. https://doi.org/10.5281/zenodo.14761243

Jabbar, A., Cocconcelli, M., D'Elia, G., Borghi, D., Capelli, L., Cavalaglio Camargo Molano, J., Strozzi, M., & Rubini, R. (2025c). Moira-unimore bearing data set for independent cart systems—experiment type 7. https://doi.org/10.5281/zenodo.14764717

Jabbar, A., Cocconcelli, M., D'Elia, G., Borghi, D., Capelli, L., Cavalaglio Camargo Molano, J., Strozzi, M., & Rubini, R. (2025d). Moira-unimore bearing data set for independent cart systems—experiment type 8. https://doi.org/10.5281/zenodo.14765815

Jabbar, A., Cocconcelli, M., D'Elia, G., Borghi, D., Capelli, L., Cavalaglio Camargo Molano, J., Strozzi, M., & Rubini, R. (2025e). Moira-unimore bearing data set for independent cart systems-experiment type 3 and 4. https://doi.org/10.5281/zenodo.14755761

Jabbar, A., Cocconcelli, M., D'Elia, G., Borghi, D., Capelli, L., Cavalaglio Camargo Molano, J., Strozzi, M., & Rubini, R. (2025). *Moira-unimore bearing data set for independent cart systems—experiment type 1 and 2*. https://doi.org/10.5281/zenodo.14753683

Jabbar, A., Cocconcelli, M., d'Elia, G., Strozzi, M., & Rubini, R. (2023). Results on experimental data analysis of independent cart systems in non-stationary conditions [hal-04165905]. *Surveillance, Vibrations, Shock and Noise*. https://hal.science/hal-04165905

Jabbar, A., D'Elia, G., & Cocconcelli, M. (2024). 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* (pp. 517–530). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-39619-9-38

Jabbar, A., D'Elia, G., & Cocconcelli, M. (2025). Distribution reshaping transformation for bearing fault diagnosis in independent cart systems. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2025.3636190

Jabbar, A., Fonte, C., D'Elia, G., & Cocconcelli, M. (2024). Ball-bearings fault detection for an independent cart system: Experimental campaign and preliminary results. *Proceedings of ISMA 2024 - International Conference on Noise and Vibration Engineering and USD 2024 - International Conference on Uncertainty in Structural Dynamics*, 1698–1711.

Jabbar, A., Mazzonetto, M., Orazi, L., & Cocconcelli, M. (2024). Ultrafast laser damaging of ball bearings for the condition monitoring of a fleet of linear motors. *PHM Society European Conference*, *8*(1), 10. https://doi.org/10.36001/phme.2024.v8i1.4136

Jablonski, A. (2021). *Condition monitoring algorithms in MATLAB®*. Springer. https://doi.org/10.1007/978-3-030-62749-2

Jablonski, A., & Dziedziech, K. (2022). Intelligent spectrogram – a tool for analysis of complex non-stationary signals. *Mechanical Systems and Signal Processing*, *167*, 108554. https://doi.org/https://doi.org/10.1016/j.ymssp.2021.108554

Jang, G.-B., & Cho, S.-B. (2021). Feature space transformation for fault diagnosis of rotating machinery under different working conditions. *Sensors*, *21*(4). https://doi.org/10.3390/s21041417

Jiang, Y., Zhou, J., Wu, X., Liu, T., & Liu, X. (2025). Vision-based bearing fault diagnosis under non-stationary conditions using optimized short-time concentrated transform method. *Reliability Engineering & System Safety*, *262*, 111183. https://doi.org/https://doi.org/10.1016/j.ress.2025.111183

Johnson, J. E., Laparra, V., Piles, M., & Camps-Valls, G. (2021). Gaussianizing the earth: Multidimensional information measures for earth data analysis. *IEEE Geoscience and Remote Sensing Magazine*, *9*(4), 191–208. https://doi.org/10.1109/MGRS.2021.3066260

Jolliffe, I., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, *374*, 20150202. https://doi.org/10.1098/rsta.2015.0202

Jones, M. C., & Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, *96*(4), 761–780. https://doi.org/10.1093/biomet/asp053

Jung, S., Dagobert, T., Morel, J., & Facciolo, G. (2024). A review of t-sne. *Image Processing On Line*, *14*, 250–270. https://doi.org/10.5201/ipol.2024.528

Kitao, A. (2022). Principal component analysis and related methods for investigating the dynamics of biological macromolecules. *J*, *5*, 298–317. https://doi.org/10.3390/j5020021

Kou, L., Liu, C., Cai, G.-w., Zhang, Z., Li, X.-j., & Yuan, Q.-d. (2020). Fault diagnosis for power converters based on random forests and feature transformation. *2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia)*, 1821–1826. https://doi.org/10.1109/IPEMC-ECCEAsia48364.2020.9367970

Kruskal, W., & Wallis, W. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*, 583–621. https://doi.org/10.2307/2280779

Lang, C. I., Sun, F.-K., Lawler, B., Dillon, J., Dujaili, A. A., Ruth, J., Cardillo, P., Alfred, P., Bowers, A., Mckiernan, A., & Boning, D. S. (2022). One class process anomaly detection using kernel density estimation methods. *IEEE Transactions on Semiconductor Manufacturing*, *35*(3), 457–469. https://doi.org/10.1109/TSM.2022.3181468

Laparra, V., Camps-Valls, G., & Malo, J. (2011). Iterative gaussianization: From ica to random rotations. *IEEE Transactions on Neural Networks*, *22*(4), 537–549. https://doi.org/10.1109/TNN.2011.2114351

Laparra, V., Hepburn, A., Johnson, J. E., & Malo, J. (2022). Orthonormal convolutions for the rotation based iterative gaussianization. *2022 IEEE International Conference on Image Processing (ICIP)*, 4018–4022. https://doi.org/10.1109/ICIP46576.2022.9897849

Lee, J., Qiu, H., Yu, G., & Lin, J. (2007). Ims, university of cincinnati. "bearing data set".

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016). Condition monitoring of bearing damage in electromechanical drive systems (paderborn dataset) [Paderborn University Bearing Dataset].

Li, C., Zhang, S., Qin, Y., & Estupinan, E. (2020). A systematic review of deep transfer learning for machinery fault diagnosis. *Neurocomputing*, *407*, 121–135. https://doi.org/10.1016/j.neucom.2020.04.045

Li, C., Li, S., Zhang, A., He, Q., Liao, Z., & Hu, J. (2021). Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing*, *439*, 197–211. https://doi.org/10.1016/j.neucom.2021.01.099

Li, K., Ping, X., Wang, H., Chen, P., & Cao, Y. (2013). Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis. *Sensors*, *13*(6), 8013–8041. https://doi.org/10.3390/s130608013

Li, L., Ma, Z., Yu, Z., Bai, X., & Li, B. (2025). Bearing fault diagnosis under variable speed conditions based on time series mixup and unsupervised domain adaptation. *Measurement Science and Technology*, *36*(4), 046130. https://doi.org/10.1088/1361-6501/adc4fd

Li, R., Ran, C., Zhang, B., Han, L., & Feng, S. (2020). Rolling bearings fault diagnosis based on improved complete ensemble empirical mode decomposition with adaptive noise, nonlinear entropy, and ensemble svm. *Applied Sciences*, *10*(16), 5542. https://doi.org/10.3390/app10165542

Li, Y., & Xia, H. (2024). Bearing fault diagnosis for variable operating conditions based on deep learning and mutual attention feature fusion. *Proceedings of the 43rd Chinese Control Conference*, 4925–4930.

Liu, D., Cui, L., & Cheng, W. (2023). Flexible generalized demodulation for intelligent bearing fault diagnosis under nonstationary conditions. *IEEE Transactions on Industrial Informatics*, *19*(3), 2717–2728. https://doi.org/10.1109/TII.2022.3192597

Liu, Q., & Wang, Y. (2021). Fault diagnosis of rolling bearing under variable speed conditions using multisynchrosqueezing transform. *Journal of Physics: Conference Series*, *1820*(1), 012124. https://doi.org/10.1088/1742-6596/1820/1/012124

Liu, Z., & Zhang, L. (2020). Naturally damaged wind turbine blade bearing fault detection using novel iterative nonlinear filter and morphological analysis. *IEEE Transactions on Industrial Electronics*, *67*(10), 8713–8722. https://doi.org/10.1109/TIE.2019.2949522

Liu, Z., Peng, D., Zuo, M. J., Xia, J., & Qin, Y. (2022). Improved hilbert–huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. *ISA Transactions*, *125*, 426–444. https://doi.org/https://doi.org/10.1016/j.isatra.2021.07.011

Liznerski, P., Ruff, L., Vandermeulen, R. A., Franks, B. J., Kloft, M., & Muller, K.-R. (2020). Explainable deep one-class classification. *ArXiv*, *abs/2007.01760*.

Ma, H., Li, S., & An, Z. (2019). A fault diagnosis approach for rolling bearing based on convolutional neural network and nuisance attribute projection under various speed conditions. *Applied Sciences*, *9*(8), 1603. https://doi.org/10.3390/app9081603

Ma, J., Wei, J., Li, Q., & Xia, L. (2025). Variable-speed bearing fault diagnosis based on bdvmd, frtsmfrbsie, and parameter-optimized gru-mhsa. *Processes*, *13*(2). https://doi.org/10.3390/pr13020498

Ma, Z., & Guo, H. (2024). Fault diagnosis of rolling bearing under complex working conditions based on time-frequency joint feature extraction-deep learning. *Journal of Vibroengineering*, *26*(7), 1635–1652. https://doi.org/10.21595/jve.2024.24238

Maliuk, A. S., Prosvirin, A. E., Ahmad, Z., Kim, C. H., & Kim, J.-M. (2021). Novel bearing fault diagnosis using gaussian mixture model-based fault band selection. *Sensors*, *21*(19). https://doi.org/10.3390/s21196579

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. https://arxiv.org/abs/1802.03426

McInnes, L., Healy, J., & Melville, J. (2020). Umap: Uniform manifold approximation and projection for dimension reduction. https://arxiv.org/abs/1802.03426

Meng, Y., Zhou, J., Lei, F., Li, D., & Liu, R. (2024). A novel class of non-gaussian system performance assessment and controller parameter tuning methods. *ISA Transactions*, *154*, 199–212. https://doi.org/https://doi.org/10.1016/j.isatra.2024.08.031

Mobley, R. K. (2002). *An introduction to predictive maintenance* (2nd). Butterworth–Heinemann.

Molano, J. C. C., Scurria, L., Source, C., Cocconcelli, M., & Tamarozzi, T. (2020). Virtual training of machine learning algorithm using a multibody model for bearing diagnostics on independent cart system [Paper presented at ISMA 2020 and USD 2020]. *Proceedings of the 2020 International Conference on Noise and Vibration Engineering (ISMA 2020) and 2020 International Conference on Uncertainty in Structural Dynamics (USD 2020)*, 2013–2024.

Molano, J. C. C., Rossi, S., Cocconcelli, M., & Rubini, R. (2017). Dynamic model of an independent carts system. In G. Boschetti & A. Gasparetto (Eds.), *Advances in italian mechanism science* (pp. 379–387). Springer International Publishing.

Montgomery, D. (2017). *Design and analysis of experiments*. John Wiley & Sons.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N., & Varnier, C. (2012). Pronostia: An experimental platform for bearings accelerated life test. *Pro-*

ceedings of the IEEE International Conference on Prognostics and Health Management, 1–8.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear statistical models* (4th). Irwin Press.

Nisar, S., Khan, O. U., & Tariq, M. (2016). An efficient adaptive window size selection method for improving spectrogram visualization. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2016/6172453

Noura, H. N., Allal, Z., Salman, O., & Chahine, K. (2025). Feature engineering for fault detection and diagnosis in power transmission lines using a tree-based approach. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, *12*, 100991. https://doi.org/https://doi.org/10.1016/j.prime.2025.100991

Padrón-Hidalgo, J. A., Laparra, V., & Camps-Valls, G. (2022). Unsupervised anomaly and change detection with multivariate gaussianization. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–10. https://doi.org/10.1109/TGRS.2021.3116186

Peng, B., Bi, Y., Xue, B., Zhang, M., & Wan, S. (2022). A survey on fault diagnosis of rolling bearings. *Algorithms*, *15*(10). https://doi.org/10.3390/a15100347

Peng, D., Liu, Z., Wang, H., Qin, Y., & Jia, L. (2019). A novel deeper one-dimensional cnn with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access*, *7*, 10278–10293. https://doi.org/10.1109/ACCESS.2018.2888842

Peng, D., Teng, W., Gao, C., Tong, B., & Liu, Y. (2023). Cyclic band box-cox sparse measures based blind filtering and its application to bearing fault diagnosis. *Measurement*, *218*, 113054. https://doi.org/https://doi.org/10.1016/j.measurement.2023.113054

Pham, M. T., Kim, J.-M., & Kim, C. H. (2020). Deep learning-based bearing fault diagnosis method for embedded systems. *Sensors*, *20*(23), 6886. https://doi.org/10.3390/s20236886

Pořízka, P., Klus, J., Képeš, E., Prochazka, D., Hahn, D., & Kaiser, J. (2018). On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis: A review. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *148*, 65–82. https://doi.org/10.1016/j.sab.2018.05.030

Qian, Q., Pu, H., Tu, T., & Qin, Y. (2024). Variance discrepancy representation: A vibration characteristic-guided distribution alignment metric for fault transfer diagnosis. *Mechanical Systems and Signal Processing*, *217*, 111544. https://doi.org/https://doi.org/10.1016/j.ymssp.2024.111544

Qin, L., Yang, G., & He, W. (2024). Generalized shannon entropy sparse wavelet packet transform for fault detection of traction motor bearings in high-speed trains. *Structural Health Monitoring*, *24*(2), 998–1028. https://doi.org/10.1177/14759217241245320

Qin, L.-X., & Self, S. G. (2006). A new variance stabilizing transformation for gene expression data analysis. *BMC Bioinformatics*, *7*(1), 254. https://doi.org/10.1186/1471-2105-7-254

Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, automotive and aerospace applications*. John Wiley & Sons Ltd. https://doi.org/10.1002/9781119477631

Rao, B. K. N. (2019). *Handbook of condition monitoring: Techniques and methodology*. Elsevier Science.

Romanssini, M., de Aguirre, P. C. C., Compassi-Severo, L., & Girardi, A. G. (2023). A review on vibration monitoring techniques for predictive maintenance of rotating machinery. *Eng*, *4*(3), 1797–1817. https://doi.org/10.3390/eng4030102

Sa'd, M. A., Jalonen, T., Kiranyaz, S., & Gabbouj, M. (2024). Quadratic time-frequency analysis of vibration signals for diagnosing bearing faults. *ArXiv*, *abs/2401.01172*.

Saidin, S. S., Jamadin, A., Kudus, S. A., Sani, M. S. M., Ariffin, A. K., Zahid, M. S. M., & Ali, M. S. M. (2022). An overview: The application of vibration-based techniques in bridge structural health monitoring. *International Journal of Concrete Structures and Materials*, *16*(69). https://doi.org/10.1186/s40069-022-00557-1

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2019). Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *IEEE Access*, *7*, 122644–122662. https://doi.org/10.1109/ACCESS.2019.2938227

Scheffer, C., & Girdhar, P. (2015). *Practical machinery vibration analysis and predictive maintenance* (2nd). Elsevier Butterworth–Heinemann.

Schelling, B., Miklautz, L., & Plant, C. (2020). Non-linear cluster enhancement: Forcing clusters into a compact shape. *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020)*, *325*, 1451–1458. https://doi.org/10.3233/FAIA200251

Schelling, B., & Plant, C. (2018). Diptransformation: Enhancing the structure of a dataset and thereby improving clustering. *2018 IEEE International Conference on Data Mining (ICDM)*. https://doi.org/10.1109/ICDM.2018.00056

Schelling, B., & Plant, C. (2020). Dataset-transformation: Improving clustering by enhancing the structure with dipscaling [Metadata (volume/pages) to confirm]. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-019-01388-5

Shao, S., McAleer, S., Yan, R., & Baldi, P. (2018). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, *15*, 2446–2455.

Shen, C., Xia, Y., Jiang, X., Chen, Z., Kong, L., & Zhu, Z. (2022). Optimal transport-based multisource student–teacher learning network for bearing fault diagnosis under variable working conditions. *IEEE Sensors Journal*, *22*(16), 16392–16402. https://doi.org/10.1109/JSEN.2022.3190513

Sheskin, D. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC.

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, *64-65*, 100–131. https://doi.org/https://doi.org/10.1016/j.ymssp.2015.04.021

Sohaib, M., & Kim, J.-M. (2018). Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network. *Shock and Vibration*, *2018*, 1–11. https://doi.org/10.1155/2018/2919637

Song, W., Guo, L., Duan, A., Gao, H., Yu, Y., Feng, T., Chen, T., & Ma, W. (2023). Multispectral balanced automatic fault diagnosis for rolling bearings under variable speed conditions. *Structural Control and Health Monitoring*, *2023*, 1–17. https://doi.org/10.1155/2023/9369850

Spina, D. E., de O. Campos, L. F., de Arruda, W. F., Melo, A., de S. Alves, M. F., Rabello, G. L., Anzai, T. K., & Pinto, J. C. (2024). Comparison of autoencoder architectures for fault detection in industrial processes. *Digital Chemical Engineering*, *12*, 100162. https://doi.org/https://doi.org/10.1016/j.dche.2024.100162

Tang, J., You, Y., Zhao, Y., Guo, C., Li, Z., & Yang, B. (2025). Fault diagnosis of hvac system sensors: A method based on box-cox transformation and multi-model fusion. *Energy Reports*, *13*, 3489–3503. https://doi.org/https://doi.org/10.1016/j.egyr.2025.03.012

Tayyab, S. M., Chatterton, S., & Pennacchi, P. (2022). Intelligent defect diagnosis of rolling element bearings under variable operating conditions using convolutional neural network and order maps. *Sensors*, *22*(5), 2026. https://doi.org/10.3390/s22052026

Tian, Y., Ma, J., Lu, C., & Wang, Z. (2015). Rolling bearing fault diagnosis under variable conditions using lmd-svd and extreme learning machine. *Mechanism and Machine Theory*, *90*, 175–186. https://doi.org/10.1016/j.mechmachtheory.2015.03.014

Tiboni, M., Remino, C., Bussola, R., & Amici, C. (2022). A review on vibration-based condition monitoring of rotating machinery. *Applied Sciences*, *12*(3). https://doi.org/10.3390/app12030972

Tong, Z., Li, W., Zhang, B., Jiang, F., & Zhou, G. (2018). Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning. *IEEE Access*, *6*, 76175–76197. https://doi.org/10.1109/ACCESS.2018.2883078

Tsai, A. C., Liou, M., Simak, M., & Cheng, P. E. (2017). On hyperbolic transformations to normality [Open access under CC BY-NC-ND license]. *Computational Statistics and Data Analysis*, *115*, 250–266. https://doi.org/10.1016/j.csda.2017.06.001

Urbanek, J., Barszcz, T., & Antoni, J. (2013). A two-step procedure for estimation of instantaneous rotational speed with large fluctuations. *Mechanical Systems and Signal Processing*, *38*(1), 96–102. https://doi.org/https://doi.org/10.1016/j.ymssp.2012.05.009

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, *9*, 2579–2605.

Varela, P., Silva, A., Manso, M., & Team, A. U. (2005, May). Adaptive window calculation for automatic spectrogram analysis of broadband reflectometry data. In G. Conway (Ed.), *Proceedings of the 7th international reflectometry workshop for fusion plasma diagnostics (irw 7)* (pp. 21–24).

Wang, B., Lei, Y., & Li, N. (2020). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, *69*, 401–412. https://doi.org/10.1109/TR.2018.2882682

Wang, P., Long, Z., Lv, Z., & Wang, Z. (2019). Fault detection for non-gaussian processes using multiple canonical correlation analysis models and box-cox transformation. *IEEE Access*, *7*, 68707–68717. https://doi.org/10.1109/ACCESS.2019.2914960

Wei, Z., He, D., Jin, Z., Liu, B., Shan, S., Chen, Y., & Miao, J. (2023). Density-based affinity propagation tensor clustering for intelligent fault diagnosis of train bogie bearing. *IEEE Transactions on Intelligent Transportation Systems*, *24*(6), 6053–6064. https://doi.org/10.1109/TITS.2023.3253087

Wu, T., Chen, J., & Wang, C. (2012). Characterization of gear faults in variable rotating speed using hilbert-huang transform and instantaneous dimensionless frequency normalization. *Mechanical Systems and Signal Processing*, *30*, 103–122. https://doi.org/https://doi.org/10.1016/j.ymssp.2012.01.022

XTS. (2025). Extended transport systems (xts), beckhoff automation. https://www.beckhoff.com

Yan, R., & Gao, R. X. (2009). Multi-scale enveloping spectrogram for vibration analysis in bearing defect diagnosis. *Tribology International*, *42*(2), 293–302. https://doi.org/https://doi.org/10.1016/j.triboint.2008.06.013

Yang, X., Chi, F., Shao, S., & Zhang, Q. (2021). Bearing fault diagnosis under variable working conditions based on deep residual shrinkage networks and transfer learning. *Journal of Sensors*, *2021*, 1–13. https://doi.org/10.1155/2021/5714240

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. https://doi.org/10.1093/biomet/87.4.954

Yesilyurt, I. (2004). The application of the conditional moments analysis to gearbox fault detection—a comparative study using the spectrogram and scalogram. *NDT & E International*, *37*(4), 309–320. https://doi.org/https://doi.org/10.1016/j.ndteint.2003.10.005

Yu, H., Guo, Q., Hu, J., & Xu, A. (2006). Rolling bearings fault diagnosis based on adaptive gaussian chirplet spectrogram and independent component analysis. *International Conference on Natural Computation*, 321–330.

Yu, K., Wu, Z., Sun, J., Zhang, Y., Xu, Y., Wei, Z., & Zheng, S. (2024). Accelerating hyperspectral anomaly detection with enhanced multivariate gaussianization based on fpga. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–12. https://doi.org/10.1109/TGRS.2024.3476152

Yusoff, M., Mahmud, Y., Azmi, P. A. R., & Sallehud-din, M. T. M. (2025). The improvement of smote-enn-xgboost through yeo johnson strategy on dissolved gas analysis dataset. *Energy Reports*, *13*, 6281–6290. https://doi.org/https://doi.org/10.1016/j.egyr.2025.05.013

Zeng, L., Jian, J., Chang, X., & Wang, S. (2024). A meta-learning method for few-shot bearing fault diagnosis under variable working conditions. *Measurement Science and Technology*, *35*(5), 056205. https://doi.org/10.1088/1361-6501/ad28e7

Zhang, L., Lv, Y., Huang, W., & Yi, C. (2022). Bearing fault diagnosis under various operation conditions using synchrosqueezing transform and improved two-dimensional convolutional neural network. *Measurement Science and Technology*, *33*(8), 085002. https://doi.org/10.1088/1361-6501/ac69b1

Zhang, Q., Lv, Z., Hao, C., Yan, H., & Fan, Q. (2024). Intelligent fault diagnosis of bearings in unsupervised dynamic domain adaptation networks under variable conditions. *IEEE Access*, *12*, 82911–82925. https://doi.org/10.1109/ACCESS.2024.3413087

Zhang, T., Chen, J., Liu, S., & Liu, Z. (2023). Domain discrepancy-guided contrastive feature learning for few-shot industrial fault diagnosis under variable working conditions. *IEEE Transactions on Industrial Informatics*, *19*(10), 10277–10287. https://doi.org/10.1109/TII.2023.3240921

Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., & Chen, X. (2021). Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. *IEEE Transactions on Instrumentation and Measurement*, *70*, 3525828. https://doi.org/10.1109/TIM.2021.3116309

Zheng, H., Wang, R., Yang, Y., Yin, J., Li, Y., Li, Y., & Xu, M. (2019). Cross-domain fault diagnosis using knowledge transfer strategy: A review. *IEEE Access*, *7*, 129260–129290. https://doi.org/10.1109/ACCESS.2019.2939876

Zhu, Y., Ting, K. M., Carman, M. J., & Angelova, M. (2021). Cdf transform-and-shift: An effective way to deal with datasets of inhomogeneous cluster densities. *Pattern Recognition*, *117*, 107977.

Zou, Y., Zhao, W., Liu, T., Zhang, X., & Shi, Y. (2024). Research on high-speed train bearing fault diagnosis method based on domain-adversarial transfer learning. *Applied Sciences*, *14*(19), 8666. https://doi.org/10.3390/app14198666