vibration

MDPI

Article

The Energy-Dispersion Index (EDI) and Cross-Domain Archetypes: Towards Fully Automated VMD Decomposition for Robust Fault Detection

Ikram Bagri [1], Achraf Touil [1], Rachid Oucheikh [2,*], Ahmed Mousrij [1], Aziz Hraiba [1] and Karim Tahiry [1]

[1] Electrical and Mechanical Engineering Department, Faculty of Science and Technology, Hassan 1st University, Settat 26002, Morocco; i.bagri@uhp.ac.ma (I.B.); ac.touil@uhp.ac.ma (A.T.); mousrij@gmail.com (A.M.); hraibaz@gmail.com (A.H.); karim.tahiry@gmail.com (K.T.)

[2] Department of Physical Geography and Ecosystem Science, Lund University, Solvegatan 12, SE-223 62 Lund, Sweden

* Correspondence: rachid.oucheikh@mgeo.lu.se

Abstract

Variational Mode Decomposition (VMD) is a powerful formalism for the time-scale analysis of vibration signals from rotating machinery. However, its performance is often compromised by complex parameter configuration, where subjective manual tuning leads to mode mixing or information loss. In this study, we present a physics-guided framework that generalizes VMD optimization across diverse operating conditions. We utilized a meta-dataset combining three distinct sources (CWRU, HUST, UO) to validate the approach. Through a shaft-normalized segmentation strategy and K-Means++ clustering, we identified six distinct signal archetypes based on spectral morphology. Central to this framework is the Energy-Dispersion Index (EDI), a novel physically interpretable metric designed to differentiate between structured fault transients and stochastic noise. Extensive validation via a full-factorial Design of Experiments (8640 trials) confirmed the statistical superiority of EDI over benchmarks like kurtosis and envelope entropy, yielding an 8.3% improvement in modal fidelity. Furthermore, a rigorous ablation study demonstrated that the proposed archetype-based parameterization is highly efficient. This strategy achieved a 392× speedup over online optimization while maintaining statistically equivalent diagnostic accuracy. Additionally, by generalizing parameters from high-quality archetype representatives, the framework reduced spectral leakage (Orthogonality Index) by 51.4% compared to instance-wise optimization. The resulting framework provides a mathematically rigorous, real-time solution for automated vibration signal decomposition.

Keywords: bearing fault diagnosis; vibration signal analysis; physics-informed signal processing; adaptive signal decomposition

Check for updates

Academic Editors: Mohamed Ichchou and Noureddine Bouhaddi

Received: 31 January 2026
Revised: 11 February 2026
Accepted: 24 February 2026
Published: 2 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Rolling element bearings are critical components in heavy machinery across the oil and gas, manufacturing, and power generation sectors. Consequently, early fault detection is critical to preventing catastrophic failures and minimizing costly operational downtime. Among Non-Destructive Techniques (NDTs), Vibration Signal Analysis (VSA) serves as the primary diagnostic tool, as bearing defects typically manifest as distinct transient impulses embedded within the vibration signature.

Historically, this has been treated as a signal processing challenge, addressed by traditional time-frequency methods such as the Fast Fourier Transform (FFT), Hilbert–Huang Transform (HHT), and the Kurtogram [1,2].

However, while spectral kurtosis-based methods remain the industry standard for identifying impulsive bands, they exhibit critical limitations in the context of variable-speed and distributed fault diagnosis. Firstly, the Fast Kurtogram operates under the assumption of stationarity within the analysis window; it calculates a global filter that is prone to 'spectral smearing' when fault carrier frequencies shift due to speed fluctuations or slip. Secondly, the reliance on kurtosis maximization makes it vulnerable to non-Gaussian background noise (e.g., random mechanical impacts), often resulting in false positives where the algorithm locks onto heavy-tailed noise rather than periodic fault signatures. Finally, it is inherently a single-band extractor, rendering it ill-suited for complex scenarios where multiple fault components coexist.

To address this, Variational Mode Decomposition (VMD) has emerged as a superior adaptive formalism. Unlike recursive sieving methods, VMD rigorously isolates these masked components by extracting distinct Intrinsic Mode Functions (IMFs) based on variational bandwidth optimization [3].

Despite its theoretical superiority, the performance of VMD is strictly governed by the initialization of its parameters: the number of modes (K), the bandwidth constraint (α), and the noise tolerance (τ). Incorrect parameter selection frequently leads to mode mixing or the loss of relevant fault frequencies. While exhaustive grid search methods are computationally prohibitive [4,5], standard metaheuristic optimization algorithms (such as Genetic Algorithms or PSO) typically apply a single global set of parameters to an entire dataset [6,7]. This approach ignores the inherent spectral variations between individual signals (e.g., varying RPM or load), resulting in suboptimal decomposition accuracy. Conversely, optimizing parameters for every single signal creates an unsustainable computational burden [8,9]. A further limitation lies in the objective functions used to guide these optimization algorithms. Standard metrics like kurtosis and envelope entropy are often insufficient: kurtosis is highly sensitive to random impulsive noise, while entropy measures signal complexity without accounting for the physical characteristics of the fault. These metrics fail to capture the necessary balance between concentrated signal energy and spectral compactness. This theoretical gap often results in decomposed modes that are mathematically optimized but physically meaningless—containing noise artifacts rather than true defect signatures. To address these limitations, this study proposes a dual-pronged framework for vibration-based fault diagnosis. Firstly, to bridge the gap between global inaccuracy and local redundancy, we introduce a Clustering-Guided Optimization Strategy. Signals are grouped based on spectral morphology—specifically, Frequency Center and Root Frequency Variance—allowing VMD parameters to be optimized for representative archetypes rather than the entire dataset. Secondly, we address the limitations of existing fitness functions by proposing the Energy-Dispersion Index (EDI). Derived from theoretical principles, EDI establishes a physics-informed balance between high energy concentration and low spectral entropy. The proposed framework is rigorously validated using three heterogeneous datasets (CWRU, HUST, and UO), with the stability of the EDI metric confirmed through extensive factorial experiments involving both standard literature evaluations and a novel Time-Frequency Band Quality (TFBQ) metric for band separation analysis.

The primary contributions of this work are:

1. A clustering-based optimization framework that groups signals by spectral similarity, enabling targeted parameter tuning (K, α, τ) that reconciles diagnostic accuracy with computational efficiency.

2. The development of the Energy-Dispersion Index (EDI), a physics-informed objective function designed to resolve the ambiguity of traditional metrics, validated for stability and fault extraction fidelity.

The remainder of this paper is organized as follows: Section 2 reviews the existing literature on VMD parameter tuning. Section 3 outlines the theoretical background of VMD, clustering algorithms, and the formulation of EDI. Section 4 details the proposed methodology. Section 5 presents our experimental results and comparative analysis. Finally, Sections 6 and 7 provide a discussion of our findings and our concluding remarks.

2. Literature Review

Isolating fault signatures from non-stationary vibration signals remains a fundamental challenge in machinery health assessment. Historically, this problem has been approached through two primary methodologies. The first relies on feature-enhancing time-frequency techniques, such as the Kurtogram, Fast Marching Denoising (FMD), and Maximum Correlated Kurtosis Deconvolution (MCKD), which are designed to amplify the repetitive impulses characteristic of bearing defects [1,10,11]. While effective, these methods suffer from inherent rigidity: the Kurtogram is susceptible to random impulsive noise, while MCKD requires prior knowledge of the fault period. The second methodology prioritizes adaptive signal decomposition, evolved from the Hilbert–Huang Transform (HHT) and Empirical Mode Decomposition (EMD). However, EMD's lack of mathematical rigor often results in mode mixing and end effects [12,13], limitations that were largely resolved by the introduction of Variational Mode Decomposition (VMD) [3].

Despite its sturdy mathematical foundation, VMD is not parameter-free; its performance is critically sensitive to the selection of the number of modes (K) and the bandwidth constraint (α). Consequently, recent research has concentrated on automating this selection process, revealing two interdependent limitations in the current state of the art: the optimization strategy and the evaluation metric.

Regarding the optimization strategy, the literature is dominated by bio-inspired metaheuristic algorithms, ranging from classical Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) [14–16] to modern variants like the Grey Wolf Optimizer (GWO) [17,18] and the Sine–Cosine Algorithm (SCA) [19]. While powerful, these studies predominantly employ a global optimization tactic, searching for a single optimal parameter set to process an entire dataset. This approach ignores the intrinsic spectral diversity of vibration signals caused by varying operational conditions (e.g., load, RPM), leading to a trade-off where parameters are either too generic to be effective or too specific to be robust [8,9]. To our knowledge, no framework has yet utilized unsupervised spectral clustering to mediate this trade-off. Although Deep Learning (DL) approaches have emerged to bypass manual tuning entirely [20,21], their black-box nature and high data requirements often sacrifice the physical interpretability essential for industrial diagnostics.

Critically, even an optimal search strategy is futile without a physically valid objective function. The choice of fitness metric dictates the goal of the decomposition, yet standard metrics often lack physical consistency. Kurtosis, the most common metric for impulse detection [17,22], is notoriously unstable in the presence of random noise artifacts. Conversely, entropy-based metrics (envelope, permutation, dispersion entropy) [16,23,24] effectively measure signal complexity but fail to distinguish between high-energy fault transients and low-energy background noise. Hybrid metrics attempting to combine these factors often fail to capture the necessary balance between signal energy and spectral order. This review identifies a clear gap for a physics-informed framework that addresses both the tactical inefficiency of global optimization and the conceptual ambiguity of standard fitness

metrics—a gap this study aims to bridge through the proposed Clustering-Guided strategy and the Energy-Dispersion Index (EDI).

3. Theoretical Background

This section establishes the foundational concepts driving our framework, starting with the mechanics of Variational Mode Decomposition (VMD) and extending to the innovative Energy-Dispersion Index (EDI). We then touch on clustering techniques for signal grouping and metaheuristic strategies for parameter tuning, all tailored to the nuances of vibration signal analysis in fault-prone systems.

3.1. Variational Mode Decomposition

VMD offers a flexible way to unpack vibration signals by adaptively isolating intrinsic mode functions (IMFs) via iterative refinement [3]. At its core, it breaks down a signal $x(t)$ into K distinct IMFs $u_k(t)$, each orbiting a central frequency ω_k. This boils down to minimizing the constrained objective in Equation (1), while honoring the rebuild condition in Equation (2):

$$\min_{u_k,\omega_k} \sum_{k=1}^{K} \|(\partial_t + j\omega_k)u_k\|_2^2 \tag{1}$$

$$\sum_{k=1}^{K} u_k = x(t) \tag{2}$$

Resolution comes through an augmented Lagrangian (Equation (3)), iteratively honed by the Alternating Direction Method of Multipliers (ADMM). Here, α tempers mode bandwidth for smoother extractions, and λ safeguards signal fidelity. This setup bolsters resilience to noise—common in vibration data—and curbs mode overlap, making it apt for fault hunting:

$$\begin{aligned}\mathcal{L}(u_k, \omega_k, \lambda) = \alpha \sum_{k=1}^{K} \|(\partial_t + j\omega_k)u_k\|_2^2 \\ + \left\| x(t) - \sum_{k=1}^{K} u_k \right\|_2^2 + \left\langle \lambda, x(t) - \sum_{k=1}^{K} u_k \right\rangle\end{aligned} \tag{3}$$

The key parameters shaping this process appear in Table 1.

Table 1. VMD computation parameters.

Computation Parameter	Theoretical Background
Number of modes K	Number of $u_k(t)$ modes to be extracted.
Penalty factor α	Bandwidth of the modes
Tolerance τ	Convergence precision in optimization.
Initial center frequencies $\omega_k^{(0)}$	Guides the iterative VMD process

3.2. The Energy-Dispersion Index (EDI)

In industrial settings, a vibration signal $x(n)$ is a non-stationary combination of deterministic fault-induced impulses and stochastic noise. Faults generate quasi-periodic impulses characterized by time-concentrated energy and narrowband frequency content. In contrast, ambient noise exhibits a broadband frequency distribution with no dominant periodic structure and dispersed energy across the time-frequency domain. This inverse relationship constitutes an Energy–Entropy Duality, central to VMD, where the goal is to isolate fault-relevant modes from noise [3].

3.2.1. Theoretical Foundation: The Information Plane

In the context of Blind Source Separation (BSS), the optimal decomposition of a vibration signal is achieved when the mutual information between the extracted components (IMFs) is minimized. As rigorously established in Independent Component Analysis (ICA) theory by Hyvärinen [25], this minimization is equivalent to maximizing the Negentropy of the sources.

Physically, rolling element bearing signals are composed of two distinct source classes with orthogonal information-theoretic properties:

- Class $\mathcal{C}_F$ (Fault Transients): These are super-Gaussian sources characterized by high local energy concentration and low thermodynamic entropy. This aligns with Wiggins' principle of Minimum Entropy Deconvolution (MED), which posits that structured impulses exhibit minimum entropy [26].
- Class $\mathcal{C}_N$ (Stochastic Noise): These are near-Gaussian sources characterized by dispersed energy and maximum entropy (disorder) for a given variance.

Standard metrics often evaluate these properties in isolation. For instance, Antoni [1] demonstrated the utility of kurtosis for detecting energy transients, yet kurtosis remains unstable in the presence of outliers. Conversely, recent VMD optimization studies [27] often rely solely on envelope entropy, which effectively measures sparsity but fails to account for signal energy, leading to the selection of low-energy noise artifacts.

We propose the Energy-Dispersion Index (EDI) as a joint-feature discriminant function that unifies these perspectives. It maps each extracted mode u_k into a normalized 2D feature space, defined as the Information Plane ($\mathcal{P}$):

$$\mathcal{P} = \{(\eta_k, H_k) \in \mathbb{R}^2\} \tag{4}$$

By evaluating these dimensions simultaneously, EDI acts as a proxy for Negentropy Maximization, ensuring that selected modes are both statistically sparse (per Wiggins [26]) and energetically significant (per Antoni [1]).

3.2.2. Mathematical Formulation

Let $u_k(n)$ be the k-th mode extracted by VMD. We define two normalized descriptors:

1. Normalized Energy Density (η_k):

$$\eta_k = \frac{E_k}{E_{\text{total}}} = \frac{\sum_{n=1}^{N} u_k(n)^2}{\sum_{n=1}^{N} x(n)^2} \tag{5}$$

This metric represents the fraction of the total signal energy captured by mode k.

2. Temporal Information Entropy (H_k):

$$H_k = -\sum_{n=1}^{N} p_n \log_2 p_n, \quad \text{where } p_n = \frac{|u_k[n]|^2}{\sum_{i=1}^{N} |u_k[i]|^2 + \epsilon} \tag{6}$$

Here, H_k (Shannon Entropy) serves as a measure of sparsity . A widely dispersed signal (noise) approaches maximal entropy ($H \to \log_2 N$), while a sparse, impulsive signal (fault) yields low entropy.

3.2.3. Statistical Hypothesis Framework and Scoring

To ensure robustness against signal scaling, we employ an adaptive statistical framework. A successful VMD decomposition should separate the signal into physically meaningful and statistically distinct components. Let μ_η, σ_η and μ_H, σ_H be the mean and standard

deviation of the energy and entropy values for the current set of K modes. We standardize the coordinates for each mode k:

$$z_{\eta,k} = \frac{\eta_k - \mu_\eta}{\sigma_\eta}, \quad z_{H,k} = \frac{H_k - \mu_H}{\sigma_H} \tag{7}$$

Based on these Z-scores, the classification of each mode can be interpreted as a Multiple Hypothesis Test. For each mode k, we evaluate:

$$\begin{cases} H_0: & \text{Mode is not distinctive } (|z_{\eta,k}| \leq 1 \text{ or } |z_{H,k}| \leq 1) \\ H_1^+: & \text{Mode is distinctive AND follows physical pattern } (\text{sign}(z_{\eta,k}) \neq \text{sign}(z_{H,k})) \\ H_1^-: & \text{Mode is distinctive BUT violates physical pattern } (\text{sign}(z_{\eta,k}) = \text{sign}(z_{H,k})) \end{cases}$$

We define a ternary indicator function Ψ_k to score these hypotheses, which corresponds to the quadrant partitioning illustrated in Figure 1:

$$\Psi_k = \begin{cases} +1, & \text{if } H_1^+ \text{ is accepted (Quadrants } Q_1, Q_3) \\ -1, & \text{if } H_1^- \text{ is accepted (Quadrants } Q_2, Q_4) \\ 0, & \text{if } H_0 \text{ is accepted (Center Region)} \end{cases}$$

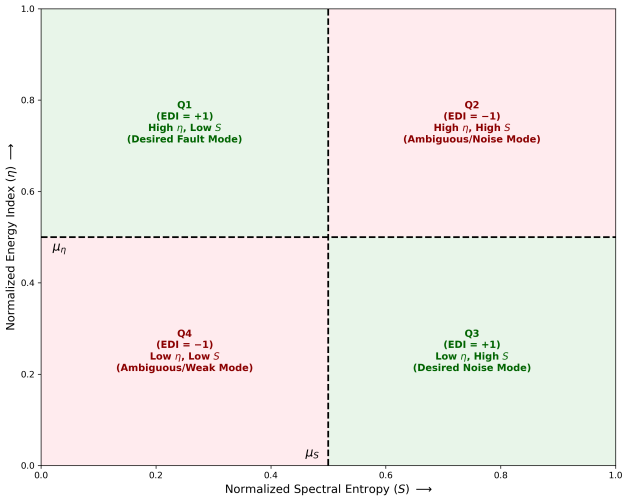


Figure 1. Energy–Entropy plane partitioning. Optimal decompositions exhibit anti-correlation between η and H, clustering modes in Q1 (fault) and Q3 (noise) while avoiding Q2 (mixed) and Q4 (trivial).

The final Global EDI Score is the average fitness across all K modes:

$$\text{EDI} = \frac{1}{K} \sum_{k=1}^{K} \Psi_k \in [-1, 1] \tag{8}$$

This global score provides a clear physical interpretation: EDI $\approx +1$ indicates excellent separation (clear fault/noise distinction); ≈ 0 indicates poor separation; and ≈ -1 indicates a pathological decomposition that systematically violates physical principles.

3.2.4. Properties and Theoretical Advantages

This formulation offers significant advantages over traditional metrics:

1. Boundedness: Unlike kurtosis (which is unbounded $[0, \infty)$), EDI is strictly bounded within $[-1, 1]$. This normalized landscape prevents the impulse problem, where optimization algorithms fixate on a single noise artifact with infinite kurtosis.

2. Adaptive Thresholding: The use of internal Z-scores means the metric is self-referential. It automatically adapts to the varying noise floors of different machines without requiring manual threshold tuning.
3. Invariance Properties: As both η_k (a ratio of energies) and H_k (a normalized entropy) are normalized quantities, the EDI score is inherently invariant under amplitude scaling ($\text{EDI}(c \cdot u_k) = \text{EDI}(u_k)$) and time dilation ($\text{EDI}(u_k(at)) = \text{EDI}(u_k)$).
4. Information Maximization: By explicitly penalizing Q_2/Q_4 (ambiguity) and rewarding Q_1/Q_3 (clarity), maximizing EDI is equivalent to maximizing the mutual information between the decomposition and the physical truth of the machinery state.

3.3. Data Clustering for Centroid Identification

Central to our validation (detailed in Section 4.3) is pinpointing natural parameter clusters—or centroids—from datasets, an unsupervised endeavor. For impartiality, we draw on five algorithms: K-Means++ for smart seeding [28], GMM for probabilistic fits [29], and density-driven ones like DBSCAN [30], HDBSCAN [31], and OPTICS [32]. Their varied lenses yield a rounded view of vibration data structures, informing consensus-based centroids without bias toward any single method.

3.4. Metaheuristic Optimization Strategies

Navigating VMD's nonlinear parameter landscape, steered by our fitness metric (Section 4.5.2), calls for metaheuristics that juggle broad searches with focused tweaks. To fortify our approach, we selected three with distinct tactics: Simulated Annealing (SA) for thermal-inspired escapes [33], Grey Wolf Optimizer (GWO) for pack dynamics [34], and Sine–Cosine Algorithm (SCA) for trig-based balance [35]. Their essentials appear in original works; Table 2 captures key updates, highlighting their fit for vibration optimization's demands.

Table 2. Core mechanisms in selected metaheuristics.

Algorithm	Key Update(s)	Ref.
SA	$P(\Delta E, T) = e^{-\Delta E/T}$; $T_{n+1} = \gamma T_n$	[33]
GWO	$\vec{D}_{\alpha,\beta,\delta} = \lvert\vec{C} \cdot \vec{X}_{\alpha,\beta,\delta} - \vec{X}\rvert$; etc. (pack avg.)	[34]
SCA	$X_i^{t+1} = X_i^t + r_1 \cdot \sin(r_2) \cdot \lvert r_3 P_i^t - X_i^t\rvert$; cosine variant	[35]

A critical limitation of applying metaheuristic algorithms to VMD parameter selection is the computational cost. Since every agent in the population requires a full VMD decomposition to evaluate the fitness function (e.g., envelope entropy), the total time complexity scales linearly with the population size and iteration count.

Table 3 summarizes the theoretical complexity for the algorithms utilized in this study. The complexity of the standard VMD algorithm is dominated by the Fourier operations, scaling as $\mathcal{O}(K \cdot n \log n)$. Consequently, population-based metaheuristics (GWO, SCA) incur a multiplicative cost factor of $P \times I$, where P is the population size and I is the maximum number of iterations.

Table 3. Theoretical computational complexity of VMD optimization algorithms.

Algorithm	Theoretical Complexity
Simulated Annealing (SA-VMD)	$\mathcal{O}(I \cdot K \cdot n \log n)$
Grey Wolf Optimizer (GWO-VMD)	$\mathcal{O}(P \cdot I \cdot K \cdot n \log n)$
Sine–Cosine Algorithm (SCA-VMD)	$\mathcal{O}(P \cdot I \cdot K \cdot n \log n)$

Note: P: population size, I: iterations, K: number of modes, n: signal length.

This multiplicative complexity ($\mathcal{O}(P \cdot I \dots)$) highlights the significant computational burden of online optimization, motivating the need for the rapid, geometry-based assignment strategy proposed in the subsequent sections.

4. Proposed Framework

The framework of this study entailed five primary stages: (1) Dynamic Signal Segmentation (Section 4.2), (2) identification of signal archetypes on the basis of spectral morphology (Section 4.3), (3) validation of the proposed metric EDI (Section 4.4), (4) Metaheuristic Optimization Comparison using an EDI-based objective function (Section 4.5). Figure 2 illustrates the pipeline of the study.

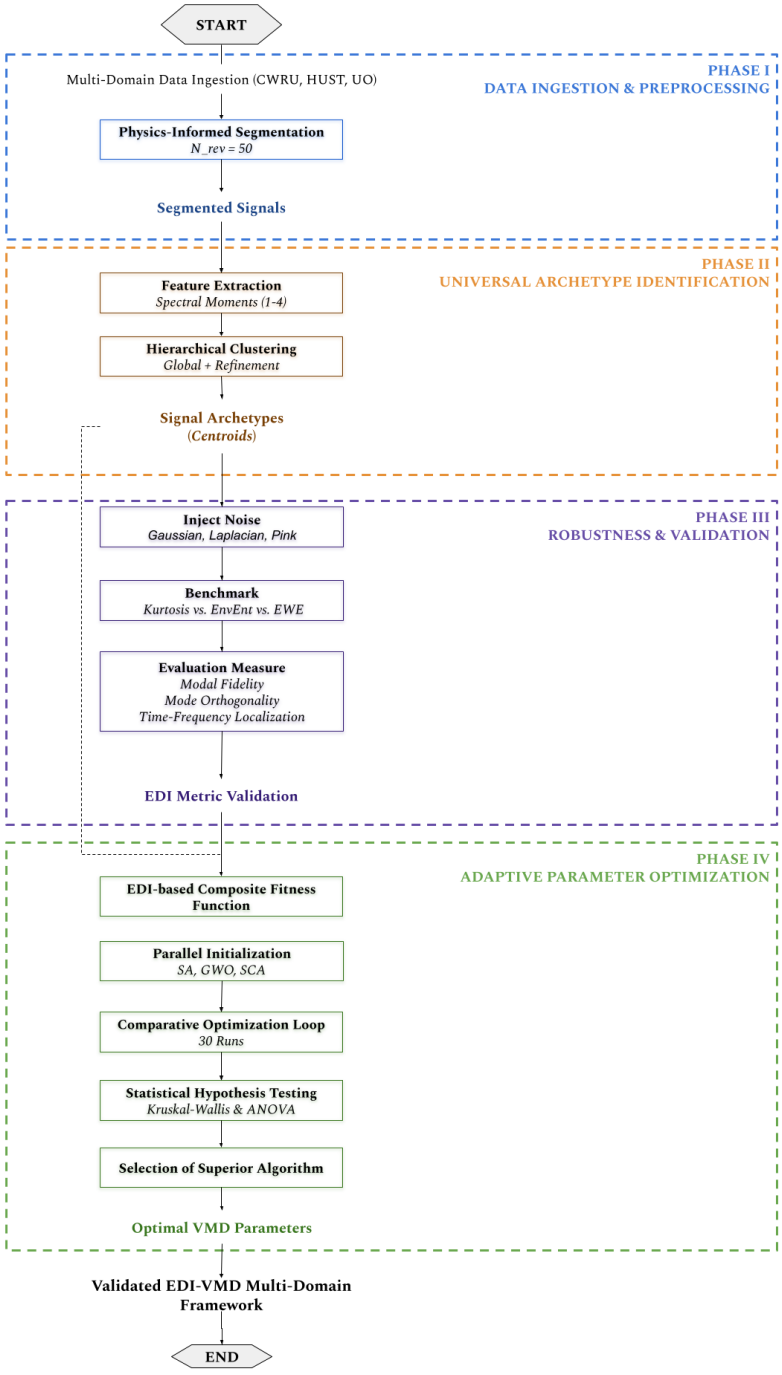


Figure 2. Overview of the proposed methodology workflow.

4.1. Experimental Datasets

The study was based on three diverse datasets: CWRU for baseline scenarios, HUST for discrete-load cases with compound faults, and UO for dynamic conditions, important for testing adaptive capability. Table 4 presents the operational conditions of each dataset.

Table 4. Comparative specifications of experimental datasets.

Feature	CWRU [36]	HUST [37]	UO [38]
Configuration	Drive-end bearing test rig	Five bearing types (6204–6208)	Multiple-bearings test rig
Conditions	0–3 hp, constant speed	0/200/400 W, constant speed	0–5000 N, dynamic speed
Faults	Inner/Outer/Ball (0.007–0.021″)	Six types incl. compound cracks	12 health states, mixed severity
Sampling	12 or 48 kHz	51.2 kHz	42 kHz

4.2. Data Preprocessing and Spectral Feature Engineering

4.2.1. Physics-Informed Dynamic Segmentation

Localized bearing faults generate repeated periodic impulses, causing reduced returns in the processing of the entire recordings. Instead, the dynamic partition of a vibration signal allows the extraction of multiple independent instances. In light of this, a dynamic segmentation strategy was adopted for the complete exploitation of raw vibration recordings. This approach also functioned as a robust data augmentation technique by multiplying sample size to solve class imbalances for the subsequent clustering step and possible machine learning tasks.

Given that the signals at hand were non-stationary in nature with varying operational conditions, the segment lengths were normalized to capture $N_{rev} = 50$ shaft revolutions in alignment with the guidelines of ISO 13373-3 [39]. The norm required the acquisition of a minimum duration of 20–25 cycles of the lowest characteristic frequency. In the same context, the authors in Refs. [1,40] emphasized the dependence of the statistical stability of spectral estimators, which were used in the next phase of the framework, on the capture of these cycles.

To ensure consistent diagnostic performance across datasets with varying sampling rates (12 kHz to 48 kHz) and rotational speeds (1730 RPM to 2400 RPM), we implemented a kinematically synchronized segmentation strategy. Rather than using a fixed time duration (e.g., 1 s), the segmentation window length L was dynamically calculated to capture a fixed angular displacement:

$$L = \frac{N_{rev} \times f_s}{f_r/60} \tag{9}$$

where $N_{rev} = 50$ was the target number of revolutions, f_s was the sampling frequency, and f_r was the shaft RPM.

This approach ensured that every analyzed segment contained an identical quantity of kinematic information (approx. 20 fault impact cycles for typical bearing geometries), rendering the feature extraction robust against speed variations. Furthermore, by constraining the analysis to 50 revolutions, the signals satisfied the quasi-stationarity requirement of VMD, effectively decoupling the fault morphology from global speed fluctuations.

Table 5 details rotational speed and fault frequency parameters for the segmentation of the signals of each dataset.

Table 5. Consolidated experimental parameters: segmentation settings ($N_{rev} = 50$) and kinematic fault characteristic orders.

Dataset	Bearing Model	Operational Parameters			Fault Orders ($X \times f_r$)			
		f_s (kHz)	Speed (RPM)	Samples (L)	BPFO	BPFI	BSF	FTF
CWRU	SKF 6205-2RS	12 / 48	1772	20,316	3.585	5.415	2.357	0.398
HUST	ER-16K	51.2	1800	85,333	3.572	5.428	2.322	0.397
UO	ER-16K	42.0	1750	72,000	3.572	5.428	2.322	0.397

Note: $N_{rev} = 50$ revolutions. Fault coefficients calculated based on $N_{balls} = 9$ and $d_{ball} \approx 7.94$ mm.; pitch diameters (D_p): SKF 6205 ≈ 39.04 mm; ER-16K ≈ 38.52 mm.

4.2.2. Meta-Dataset Consolidation

The resulting signal segments from all three datasets were aggregated into a unified meta-dataset. The latter presents a heterogeneous pool of segments of different bearing health states and operating conditions.

From the meta-dataset, a feature vector $\mathbf{v}_i = [\tilde{f}_c, \tilde{\sigma}_f, \gamma_f, \kappa_f]$ was extracted for each segment i.

The terms $\tilde{f}_c$, $\tilde{\sigma}_f$, γ_f and κ_f respectively correspond to the frequency center, the root variance, the skewness and the kurtosis of the Power Spectral Density (PSD) $S(f)$. Notably, all terms were computed on a normalized scale $\tilde{f} \in [0, 1]$ (where $1 = f_s/2$).

The mathematical formulations are:

$$\tilde{f}_c = \frac{\mu_1}{\mu_0}, \quad \tilde{\sigma}_f = \sqrt{\frac{\mu_2}{\mu_0}}, \quad \gamma_f = \frac{\mu_3}{\mu_0 \sigma_f^3}, \quad \kappa_f = \frac{\mu_4}{\mu_0 \sigma_f^4}$$
$$\text{where } \mu_n = \int_0^1 (\tilde{f} - \tilde{f}_c)^n S(\tilde{f})\, d\tilde{f}, \quad \mu_0 = \int S(\tilde{f}) d\tilde{f} \tag{10}$$

Together, these components characterize the spectral morphological shape of the vibration segments. The rationale behind this step lies in the operational mechanics of VMD. The method acts as a non-recursive frequency-domain filter bank [3]. As such, the optimal VMD decomposition configuration of a given signal is intrinsically tied to its spectral morphology. This represents the pillar of the subsequent step of our pipeline, where we conduct an exploratory analysis on the meta-dataset via clustering. The main goal is to group signals based on their frequency structures. As indicated by Jiang et al. [41], the optimal penalty factor α is strictly dependent on the spectral complexity and support of the target feature. In other words, signals exhibiting statistically similar spectral morphologies are of identical optimization constraints to the VMD solver. A major implication of this phenomenon is that parameters optimized for a group representative can be validly generalized to the constituent signals of its group.

4.3. Cross-Domain Clustering Framework

As explained in Section 4.2.2, exploratory clustering based on spectral features was performed on the combined meta-dataset. The main objective was to identify representative signal archetypes in the pool of signals.

Given the heterogeneous structure of data, a comparative selection protocol was conducted on four distinct clustering algorithms. The goal was to identify the algorithm that best captured the natural data topology: Partition-based (K-Means++), Density-based (DBSCAN, OPTICS), and Probabilistic (Gaussian Mixture Models). The optimal algorithm was selected based on a multi-metric assessment involving:

- Silhouette Score (S): Cluster cohesion and separation.
- Davies–Bouldin Index (DB): Intra-cluster similarity against inter-cluster differences.

- Calinski–Harabasz Index (CH): Ratio of dispersion between and within clusters.
- Point Coverage: Number of data points successfully assigned to clusters (minimizing noise classification).

The highest-performing algorithm was selected as the global estimator. The latter was subjected to a refinement phase to mitigate high intra-cluster variance in broad spectral categories.

4.4. Design of Experiments for EDI Performance Assessment

To put forward the importance of our proposed decomposition metric EDI, presented in Section 3.2, an extensive Design of Experiments (DOE), evaluating the metric's performance against established benchmarks, was conducted. The DOE's response metrics were based on the evaluation approach used by the creators of the decomposition method [3]. Three core questions were addressed:

1. Does EDI-guided optimization extract modes that better match true signal components?
2. Does EDI produce cleaner mode separation with reduced inter-mode leakage?
3. Does EDI yield modes with improved concentration in the time-frequency domain?

4.4.1. Paper-Standard Validation Metrics

Three core metrics from the validation approach established in [3] were employed.

1. Modal Fidelity (Component Match)

The original VMD paper evaluated decomposition quality by comparing extracted modes against known signal components. In our study, this approach was adapted using the maximum absolute correlation between any mode and the clean reference signal:

$$\text{Fidelity} = \max_k |\rho(u_k, x_{\text{clean}})|, \quad \rho(a,b) = \frac{\mathbb{E}[(a-\mu_a)(b-\mu_b)]}{\sigma_a \sigma_b} \tag{11}$$

2. Orthogonality Index (Mode Leakage)

A fundamental milestone of VMD is the production of quasi-orthogonal modes with minimal leakage between components. This orthogonality assessment, used in [3], was quantified via the normalized sum of absolute inner products:

$$\text{OI} = \frac{\sum_{i \neq j} |\langle u_i, u_j \rangle|}{\|x_{\text{clean}}\|^2} \tag{12}$$

3. Primary Mode SNR (Component Quality)

Component-wise SNR evaluation from [3] to noisy conditions was implemented using the SNR of the best-matching mode (how cleanly the dominant component is extracted):

$$\text{SNR}_{\text{primary}} = 10 \log_{10} \left(\frac{\|x_{\text{clean}}\|^2}{\|x_{\text{clean}} - u_{\text{best}}\|^2} \right), \quad u_{\text{best}} = \arg\max_k |\rho(u_k, x_{\text{clean}})| \tag{13}$$

4.4.2. Time-Frequency Band Quality (TFBQ)

To further assess the performance of EDI, a Time-Frequency Band Quality (TFBQ) index was introduced.

The primary rationale for constructing this metric lies in the inadequacy of kurtosis and standard statistical metrics for complex non-stationary signals. As a matter of fact, kurtosis is strictly a one-dimensional metric that evaluates the global spikiness of the amplitude distribution. As such, it suffers from a critical vulnerability: a single high-amplitude artifact

can artificially inflate the kurtosis value, yielding a false positive even in the absence of a repetitive fault signature. Conversely, TFBQ operates in the three-dimensional time-frequency (TF) domain which enforces a structural constraint. It requires energy to be not just high-amplitude, but bunched into specific vertical structures characteristic of repetitive impacts. This makes TFBQ significantly more robust than 1D metrics for this problem scope.

From a physical lens, white noise manifests as a diffuse fog, spread across the TF domain whereas fault transients appear as sharp, localized horizontal bands. TFBQ focuses on this contrast by measuring the ratio of energy concentrated within these mode footprints (intra-band) to the energy leaking into the background (inter-band).

TFBQ is computed using the Short-Time Fourier Transform (STFT). For each extracted mode $u_k(t)$, we compute the spectral magnitude $|S_k(t,f)|$ using a Hann window with length $N_{win} = 512$. To evaluate the decomposition as a whole, we define an aggregated spectral map $S_\Sigma(t,f) = \sum_{k=1}^{K} |S_k(t,f)|$.

The core of TFBQ lies in its dynamic footprint masking. Rather than applying a static amplitude threshold, we define the footprint based on the instantaneous cumulative energy distribution. For a given time instance t, let $E_t(f) = |S_k(t,f)|$ represent the spectral slice. We calculate the cumulative energy function $C_t(f) = \int_0^f E_t(\nu)d\nu$. The mode footprint $\mathcal{M}_k(t)$ is defined as the frequency interval containing the central 90% of the energy at time t:

$$\mathcal{M}_k(t) = \{f \in [0, f_s/2] : 0.05 \cdot C_t(f_{max}) \leq C_t(f) \leq 0.95 \cdot C_t(f_{max})\} \tag{14}$$

The final global mask $\mathcal{M}$ is the union of all the mode footprints: $\mathcal{M} = \bigcup_{k=1}^{K} \mathcal{M}_k$. The TFBQ (in dB) quantifies the contrast between the energy retained within this valid footprint (intra-band) and the leakage energy (inter-band) (Equation (15)).

$$\text{TFBQ}_{\text{dB}} = 10 \log_{10} \left(\frac{\sum_{(t,f)\in\mathcal{M}} (S_\Sigma(t,f))^2}{\sum_{(t,f)\notin\mathcal{M}} (S_\Sigma(t,f))^2 + \epsilon} \right) \tag{15}$$

where $\epsilon = 10^{-10}$ is a regularization term for numerical stability. A higher TFBQ indicates that the VMD decomposition has successfully compacted the signal energy into distinct, physically meaningful bands, minimizing spectral leakage and noise. Through this formulation, the TFBQ serves as a direct proxy for the correctness of decomposition quality as well as the penalty parameter α. Table 6 presents the physical interpretation of the scales of the metric and implications on the decomposition.

Table 6. Physical interpretation of TFBQ ranges in VMD decomposition.

Criteria	High TFBQ (>20 dB)	Moderate TFBQ (10–20 dB)	Low TFBQ (<10 dB)
Decomposition Quality	Optimal Separation: Clean, well-localized modes with minimal spectral cross-talk.	Partial Interference: Acceptable separation but exhibits leakage between adjacent bands.	Failure: Severe mode mixing or noise domination; diagnostically unusable.
α Parameter Status	Optimal Regularization: Filter bandwidth perfectly matches the fault modulation structure.	Bandwidth Mismatch: Filter is uncoordinated with the signal complexity (often leads to Mode Mixing).	Extreme Parameter Error: 1. Under-reg: Too wide (Noise). 2. Over-smooth: Too narrow (Blur).

Table 6. *Cont.*

Criteria	High TFBQ (>20 dB)	Moderate TFBQ (10–20 dB)	Low TFBQ (<10 dB)
Physical Manifestation	Sharp Focus: Fault impulses appear as distinct vertical structures; background noise is suppressed.	Soft Focus: Amplitude modulation artifacts (beating) appear due to frequency overlap.	Smear/Blur: Signal is either buried in fog (noise) or flattened into a featureless sine wave.

4.4.3. Experimental Design

Building on the evaluation metrics in Sections 4.4.1 and 4.4.2, the systematic full factorial DOE is presented in Figure 3 and Table 7. The DOE was designed to compare EDI against three established VMD optimization metrics (EWE, envelope entropy, kurtosis).

Table 7. DOE Configuration and VMD parameter search space.

Component	Settings/Values	Rationale
DOE Factors		
Signal archetypes	Six clusters (C0–C5)	Cover diverse vibration patterns
Noise types	Gaussian, Laplacian, Pink	Common industrial noise sources
SNR levels	15, 5, 0, −5 dB	Mild-to-severe noise conditions
Objective metrics	EDI, EWE, EnvEnt, kurtosis	Representative metrics from the literature
Trials per condition	30 random VMD parameterizations	Ensure statistical power
VMD Parameter Search		
Mode count (K)	Random integer $\in [3, 7]$	Typical bearing signal decomposition
Bandwidth (α)	Random $\in [200, 3000] \times f_s/12{,}000$	Scaled with sampling rate
Tolerance (τ)	Random choice from $\{10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}\}$	Balance convergence and accuracy
Experimental Scale		
Conditions	$6 \times 3 \times 4 \times 4 = 288$	Full factorial design
VMD executions	$288 \times 30 = 8640$	Comprehensive parameter search
Statistical power	>80% for Cohen's $d > 0.8$ at $\alpha = 0.05$	Reliable effect detection

DOE Execution Protocol:

Step 1: **Initialization**: For each combination of cluster, noise type, SNR, and metric.

Step 2: **Signal Preparation**:

- Select clean signal x_{clean} from cluster
- Generate noisy signal: $x_{\text{noisy}} = x_{\text{clean}} + \alpha \cdot n(t)$ (scaled to target SNR)

Step 3: **Parameter Search** (30 trials):

- Randomly select VMD parameters (K, α, τ) from ranges in Table 7
- Execute VMD decomposition
- Compute optimization metric score
- Track best-scoring decomposition

Step 4: **Validation Assessment**: Compute four validation metrics (Modal Fidelity, OI Leakage, Primary SNR, TFBQ) on best decomposition.

Step 5: **Data Storage**: Store results for statistical analysis.

Figure 3. DOE execution protocol showing systematic procedure for each experimental condition.

4.5. Automated VMD Parameter Tuning via Metaheuristic Optimization

The final phase of our pipeline involved a multi-algorithm metaheuristic approach to explore the optimal VMD configurations of our heterogeneous signals. We compared three metaheuristic algorithms in this context: Simulated Annealing (SA), Sine–Cosine Algorithm (SCA) and Grey Wolf Optimizer (GWO) as presented in Section 3.4. The procedure of the comparison study is outlined in the sections below.

4.5.1. Problem Formulation

The optimization problem was defined as finding the optimal vector **x** that would maximize the proposed objective function $F(\mathbf{x})$. The decision vector consisted of the three critical VMD parameters (Equation (16)) subject to the search space boundaries defined in Table 8. The decomposition mode count (K) dictates the granularity of the spectral segmentation; the bandwidth constraint (α) controls the mode compactness; and the Lagrangian multiplier (τ) governs the reconstruction fidelity and noise tolerance:

$$\mathbf{x} = [K, \alpha, \tau] \tag{16}$$

Table 8. Configuration of the Metaheuristic Optimization Framework.

Parameter	Value/Range	Description
Search Space Boundaries		
Mode Count (K)	$[3, 8]$	Integer constraint
Bandwidth (α)	$[500, 3000]$	Continuous
Lagrangian (τ)	$[0.0, 0.01]$	Continuous
Algorithm Settings		
Population size	10	For GWO and SCA
Max iterations	100	Per optimization run
Independent runs	30	Per algorithm, per cluster
Signal archetypes	6	Per algorithm, per run
Total experiments	540	Robustness validation

4.5.2. Objective Function

The fitness of a given candidate solution was evaluated using a composite metric that strictly penalized signal distortion while rewarding feature clarity. The global objective function $F(\mathbf{x})$ was defined as the summation of three normalized indices:

$$F(\mathbf{x}) = \mathrm{SQ}_{\mathrm{norm}} + \mathrm{EDI}_{\mathrm{norm}} + \mathrm{Match}_{\mathrm{norm}} \tag{17}$$

where $\mathrm{EDI}_{\mathrm{norm}}$ is the Energy-Dispersion Index (detailed in Section 3.2). The Signal Quality ($\mathrm{SQ}_{\mathrm{norm}}$) and Energy Match ($\mathrm{Match}_{\mathrm{norm}}$) components were mathematically formulated as follows:

1. Normalized Signal Quality ($\mathrm{SQ}_{\mathrm{norm}}$)

This composite metric assessed the reconstruction fidelity by aggregating three fundamental signal processing indicators: Signal-to-Noise Ratio (SNR), Coherence, and Pearson Correlation. First, a raw composite score ($\mathrm{SQ}_{\mathrm{raw}}$) was calculated:

$$\mathrm{SQ}_{\mathrm{raw}} = \frac{1}{3}\left(\frac{\mathrm{SNR}}{50} + \mathcal{C}_{oh} + \rho_{xy}\right) \tag{18}$$

where $\mathrm{SNR} = 10\log_{10}(\frac{P_{signal}}{MSE})$, $\mathcal{C}_{oh}$ represents the spectral coherence between the original signal $x(n)$ and the reconstructed sum of modes $\hat{x}(n)$, and ρ_{xy} is the Pearson correlation

coefficient. The scaling factor of 50 for SNR normalized the decibel scale to be comparable with the unit-interval metrics of Coherence and Correlation.

The raw score was then normalized to a robust scale:

$$\mathrm{SQ}_{\mathrm{norm}} = \frac{2 \cdot (\mathrm{SQ}_{\mathrm{raw}} + 0.333)}{1.333} - 1 \tag{19}$$

2. Energy Match Score ($\mathrm{Match}_{\mathrm{norm}}$)

To prevent the decomposition from generating mathematically valid but physically non-representative modes (e.g., extremely low-energy artifacts), we enforced an energy conservation constraint based on the Root Mean Square (RMS) amplitudes. The matching score was defined as:

$$\mathrm{Match}_{\mathrm{norm}} = 1 - \frac{\left|\mathrm{RMS}_{total} - \sum_{k=1}^{K} \mathrm{RMS}_k\right|}{\mathrm{RMS}_{total}} \tag{20}$$

where RMS_{total} is the RMS of the original signal and RMS_k is the RMS of the k-th extracted mode. This formula was introduced to penalize solutions where the sum of the mode energies deviated significantly from the total signal energy, ensuring that the decomposition accounted for the full dynamic range of the fault vibration.

4.5.3. Experimental Configuration

To guarantee statistical significance and reproducibility, a rigorous Monte Carlo simulation approach was adopted. The experimental design involved a total of 540 independent optimization runs (6 Clusters $\times$ 3 Algos $\times$ 30 Runs), with the specific hyperparameters (Table 8).

4.6. Computational Efficiency and Scalability Analysis

While the previous sections established the diagnostic accuracy of the proposed framework, a critical barrier to the industrial adoption of adaptive VMD is computational cost. Meta-heuristic optimization algorithms are computationally expensive, typically requiring minutes to converge for a single signal. To rigorously evaluate the efficiency gains of the proposed clustering strategy, we conducted a comparative ablation study against an extensive full optimization baseline.

Two distinct processing pipelines were compared:

1. Baseline (Full Optimization): The selected best algorithm from Section 4.5 is executed independently for every incoming query signal ($N_{iter} = 30$, $N_{agents} = 10$) using the same objective function.
2. Proposed Clustering Approach: The query signal is assigned to a cluster, and the pre-optimized parameters (K, α, τ) of the cluster centroid are directly applied. No optimization loop is executed during the query phase, only the inferred decomposition.

For statistical validity, the signals used in this part of the study were uniformly sampled from the identified clusters in Section 4.3. We randomly sampled $N = 30$ signals, 10 from three distinct clusters. Each signal was processed through both approaches, generating paired observations for statistical comparison. Both methods were executed on the same hardware environment (Intel Core i9-10900K, 10-Cores, 3.7 GHz).

The performance of both approaches was quantified using the paper-standard validation metrics (Modal Fidelity, Orthogonality Index, Primary SNR) and the proposed TFBQ metric used in Section 4.4.

4.7. Computational Resources

To ensure reproducibility and manage computational load, the experimental workflow was distributed across two distinct hardware environments based on the complexity of the tasks.

1. Local Workstation (General Processing): Standard signal processing, data segmentation, and post hoc analysis (including TFBQ and Modal Fidelity calculations) were conducted on a Dell G5 personal workstation:

- CPU: Intel Core i7-9750H (9th Gen)
- Architecture: Six physical cores, 12 logical threads
- Base Clock: 2.60 GHz

2. High-Performance Cloud Cluster (Optimization Pipeline): Due to the stochastic nature of the metaheuristic algorithms (GWO, SCA, WOA) and the requirement for statistical robustness (30 independent runs per cluster), the core optimization pipeline was offloaded to a cloud-based dedicated server (Vast.ai). This environment was specifically selected to leverage parallel computing capabilities.

- CPU: AMD Ryzen 9 Series
- Parallelism: 32 logical processors (leveraged for Python multiprocessing)
- Role: Concurrent execution of the optimization experiments.

All algorithms were implemented in Python 3.8, utilizing NumPy and SciPy for numerical computations and the specific metaheuristic libraries detailed in the methodology.

5. Results

5.1. Cross-Domain Clustering

5.1.1. Identification of Signal Archetypes

A summary of the clustering algorithms' performances is provided in Table 9. The density-based methods (OPTICS and DBSCAN) reported the highest Silhouette (up to 0.633) and CH Index (up to 54,344); however, their rejection of 209 to 1587 points as noise made them practically unsuitable for our dataset. In fact, these outliers often contain the earliest signs of failure in industrial settings. GMM demonstrated the weakest performance across all the metrics (CH Index of only 1361), confirming that the physics-informed feature space does not follow a simple Gaussian distribution and requires distance-based or hierarchical partitioning.

Table 9. Performance comparison across clustering methods. The hierarchical K-Means++ approach achieved an optimal combination of comprehensive coverage and structural integrity.

Algorithm	Clusters	Noise Pts	Silhouette ↑	Davies-Bouldin ↓	CH Index ↑
K-Means++ (Global)	4	0	0.394	0.816	35,051
K-Means++ (Hierarchical)	6	0	0.415	0.792	41,209
GMM	4	0	0.293	1.309	1361
DBSCAN	6	209	0.212	0.613	43,976
OPTICS	5	1587	0.633	0.575	54,344

By contrast, K-Means++ struck an effective equilibrium, identifying four distinct archetypes with 100% data coverage (zero noise points). Subsequently, this initial global clustering provided a stable baseline but lacked the granularity required to separate similar fault severities. As such, its hierarchical adaptation identified intricate sub-patterns within the fault-related data, prompting the division of larger clusters into more refined sub-categories, notably C1, C4, and C5. Notably, this refinement was characterized by

improved metrics from the global K-Means++ partitioning, increasing the Silhouette Score and CH index respectively from 0.319 to 0.415 and from 35,051 to 41,209, all while reducing the Davies–Bouldin Index by 0.024, as shown in Table 9. The resulting cluster separations are illustrated in the PCA projection depicted in Figure 4, highlighting the following distributions:

- Cluster C2, positioned in the lower right quadrant, emerged as a well-defined entity marked by intermediate kurtosis levels and elevated variance, serving as a reference for conditions involving minimal wear or substantial background interference.
- Cluster C0, located toward the upper region, aggregated signals exhibiting pronounced impulsivity, with average kurtosis exceeding 1.6, aligning closely with early manifestations of ball and outer race defects observed consistently across the datasets.
- Clusters C1, C4, and C5, situated predominantly on the left and central areas, formed adjacent formations that capture progressive defect intensities; the hierarchical adjustment effectively distinguished these by leveraging differences in spectral skewness, which often reflected uneven distributions arising from localized impacts or abrasive processes.

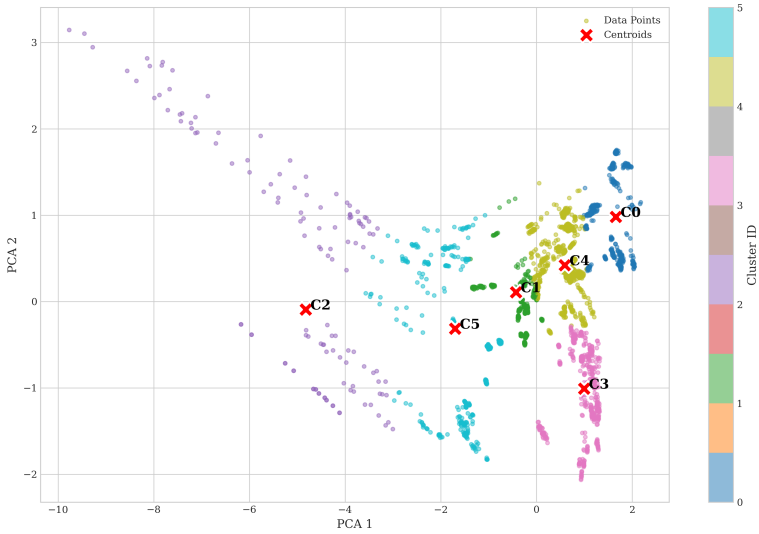
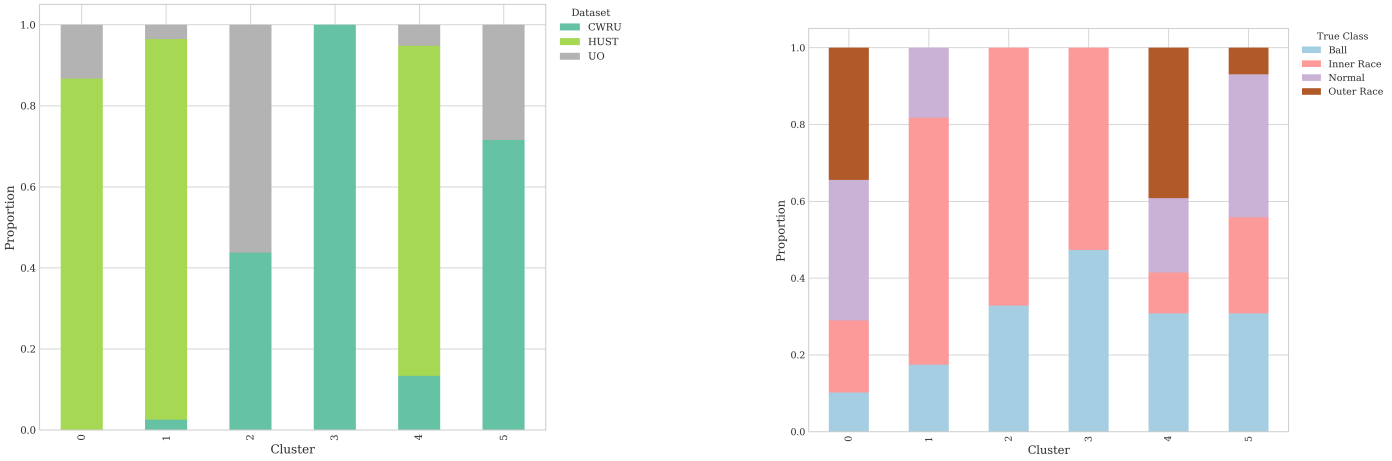


Figure 4. Principal component analysis projection illustrating the six signal archetypes derived from hierarchical spectral clustering.

5.1.2. Cross-Domain Generalization

Among the notable outcomes of the clustering approach, the identified archetypes spanned multiple data sources effectively. As depicted in Figure 5a, clusters 0 and 4 were diverse with substantial contributions from CWRU, HUST, and UO datasets alike. As an example, Cluster 4 integrated outer race signals from HUST alongside ball defect instances from CWRU, due to their comparable high-frequency impulsive profiles. In parallel, the fault archetype analysis in Figure 5b proved that the clustering prioritized underlying mechanical behaviors rather than superficial recording details, such as sampling frequencies or equipment configurations. As such, this alignment with intrinsic defect mechanisms makes the identified centroids well-suited for the subsequent optimization tasks.



(**a**) Source mix with solid crossovers in C1 and C4

(**b**) Main defect types driving each group

Figure 5. Cluster analysis results: (**a**) The breakdown of source mix shows specific crossovers. (**b**) The main defect types are highlighted per group.

5.2. Statistical Validation of the Energy-Dispersion Index (EDI)

This section presents the statistical validation of the proposed Energy-Dispersion Index (EDI) against three established metrics for VMD parameter optimization: kurtosis, envelope entropy (EnvEnt), and Enhanced Weighted Energy (EWE). Our results were derived from a full factorial design of the experiments as detailed in Section 4.4.

5.2.1. Overall Performance Comparison

Table 10 summarizes statistics calculated from 72 distinct experimental conditions per metric (six signal archetypes × three noise types × four SNR levels), with each condition averaged over 30 independent trials (total $N = 2160$ decompositions per metric). EDI achieved the highest mean modal fidelity (0.6344), representing an 8.3% relative improvement over the second-best metric (EWE), and the highest primary mode SNR (2.165 dB), while minimizing the Orthogonality Index (OI) (0.0328).

Table 10. Overall performance comparison of objective metrics. Values represent mean ± standard deviation across 72 distinct condition combinations.

Metric	Modal Fidelity	OI Leakage	Primary Mode SNR (dB)	TFBQ (dB)
EDI	0.6344 ± 0.0657	0.0328 ± 0.0336	2.1649 ± 0.6807	23.0034 ± 1.6244
EWE	0.5858 ± 0.0658	0.0768 ± 0.0414	1.7471 ± 0.6757	22.7715 ± 1.9887
EnvEnt	0.5505 ± 0.0439	0.0813 ± 0.0559	1.4771 ± 0.3004	23.3004 ± 2.0182
Kurtosis	0.5530 ± 0.0486	0.0647 ± 0.0505	1.5026 ± 0.3366	23.2930 ± 1.8312

5.2.2. Statistical Significance Analysis

Given the violations of normality and the heterogeneity of variances identified in our preliminary checks, we employed Welch's ANOVA followed by Games–Howell post hoc tests.

Welch's ANOVA Results

As shown in Table 11, statistically significant differences were found for Modal Fidelity ($F_w(3, 155.60) = 31.85$, $p < 0.0001$, $\omega^2 = 0.244$), OI Leakage ($F_w(3, 155.01) = 22.81$, $p < 0.0001$, $\omega^2 = 0.186$), and Primary Mode SNR ($F_w(3, 150.60) = 23.03$, $p < 0.0001$, $\omega^2 = 0.187$). Time-Frequency Band Quality (TFBQ) showed no significant differences ($p = 0.301$), indicating that all the metrics achieved comparable time-frequency separation despite varying diagnostic accuracy.

Table 11. Welch's ANOVA results for validation metrics. Effect sizes reported as omega-squared (ω^2).

Validation Metric	F_w	df_{num}, df_{den}	p-Value	ω^2
Modal Fidelity	31.85	3, 155.60	0.0000	0.244
OI Leakage	22.81	3, 155.01	0.0000	0.186
Primary Mode SNR	23.03	3, 150.60	0.0000	0.187
TFBQ	1.23	3, 157.21	0.301	0.002

Pairwise Comparisons

The Games–Howell post hoc tests (Table 12) confirmed EDI's superiority. For Modal Fidelity, EDI outperformed all the baselines with large effect sizes: EWE (Mean Diff = +0.0486, $p < 0.0001$, Hedges' $g = 0.74$), EnvEnt (+0.0840, $p < 0.0001$, $g = 1.49$), and kurtosis (+0.0814, $p < 0.0001$, $g = 1.40$). Similarly, EDI significantly minimized OI Leakage compared to EnvEnt ($g = -1.05$) and kurtosis ($g = -0.74$).

Table 12. Games–Howell post hoc comparisons for significant validation metrics (EDI vs. baselines).

Comparison	Mean Diff.	p-Value	Hedges' g
Dependent Var: Modal Fidelity			
EDI vs. EWE	0.0486	<0.0001	0.74
EDI vs. EnvEnt	0.0840	<0.0001	1.49
EDI vs. kurtosis	0.0814	<0.0001	1.40
Dependent Var: OI Leakage			
EDI vs. EWE	−0.0439	<0.0001	−1.16
EDI vs. EnvEnt	−0.0485	<0.0001	−1.05
EDI vs. kurtosis	−0.0319	0.0001	−0.74

The large effect sizes ($\omega^2 \geq 0.18$) indicate substantial practical differences beyond statistical significance, confirming that the performance gap is robust and methodologically sound.

5.2.3. Robustness Across Noise Conditions

EDI demonstrated consistent performance across noise types (Gaussian, Laplacian, Pink), maintaining modal fidelity above 0.633 in all conditions (Table 13). Figure 6 illustrates the metric's degradation curve. While all the metrics suffered at low SNR, EDI maintained the highest fidelity at −5 dB (0.570), effectively isolating fault signatures where entropy-based metrics failed. This resilience is attributable to the energy-dispersion formulation, which penalized the broadband dispersion characteristic of noise.

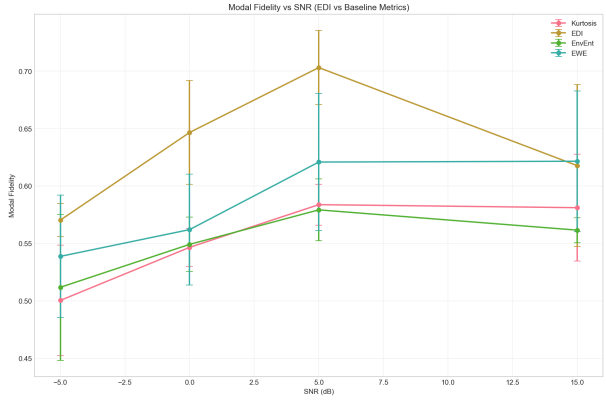


Figure 6. Noise robustness analysis. Modal fidelity as a function of SNR. EDI (blue line) maintains superior reconstruction stability even at challenging low-SNR conditions (−5 dB).

Table 13. Modal fidelity by noise type and objective metric.

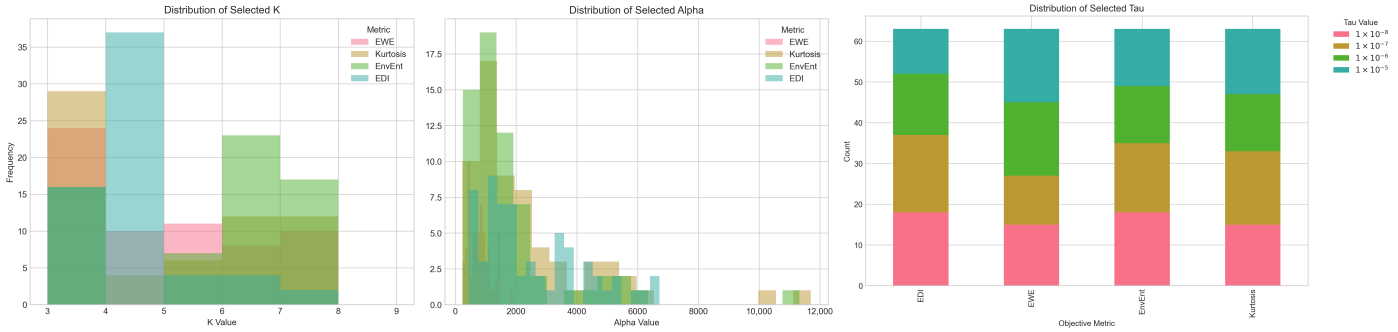
Noise Type	EDI	EWE	EnvEnt	Kurtosis
Gaussian	0.635	0.583	0.538	0.543
Laplacian	0.635	0.584	0.561	0.564
Pink	0.633	0.590	0.553	0.551

5.2.4. Parameter Selection Patterns

The robustness of an objective function was determined by its ability to identify physically meaningful VMD parameters rather than mathematical artifacts. Table 14 and Figure 7 summarize the parameter selection behaviors (K, α) of each metric.

Table 14. Average VMD parameters selected by each objective metric.

Metric	K (Mean)	α (Mean)	Failure Mode
EDI	4.04	2548.8	(Balanced optimization)
EWE	4.57	925.0	Noise leakage (Low α)
EnvEnt	5.19	1986.0	Mode splitting (High K)
Kurtosis	4.57	2948.0	Distortion (High α)



(**a**) (**b**) (**c**)

Figure 7. Distribution of optimized VMD parameters, showing EDI's tight convergence around $K = 4$ (physical fault harmonics) compared to envelope entropy's tendency to over-decompose ($K > 5$). (**a**) K selection frequency; (**b**) α selection frequency; (**c**) τ selection frequency.

The data reveals three distinct optimization behaviors:

1. Mode Splitting (EnvEnt): Envelope entropy consistently selected the highest mode count ($K_{avg} = 5.19$). As seen in Figure 7 (left), EnvEnt frequently converged to $K = 6+$, implying it minimized entropy by fragmenting a single physical fault signature into multiple redundant modes.
2. Over-Regularization (Kurtosis): Kurtosis favored the highest bandwidth constraint ($\alpha_{avg} = 2948$). While this suppresses noise, it excessively narrows the demodulation band, discarding sideband modulation information.
3. Physical Equilibrium (EDI): EDI converged to a physically consistent configuration ($K \approx 4$, $\alpha \approx 2549$). This selection aligns with the actual fault mechanics (fundamental + harmonics + sidebands) without over-segmentation.

5.2.5. Correlation with Decomposition Quality

Figure 8 visualizes the relationships between the validation metrics. The strong positive correlation between Modal Fidelity and Reconstruction SNR ($r = 0.942$) confirms fidelity as a robust proxy for signal quality. Conversely, the negative correlation with

OI Leakage ($r = -0.424$) demonstrates that the framework correctly penalizes mode mixing—metrics that allow noise leakage invariably suffer a fidelity penalty.



Figure 8. Metric Correlation Matrix. The strong correlation between Modal Fidelity and Reconstruction SNR validates the assessment framework.

5.2.6. Multiple Regression Analysis

A multiple linear regression analysis was conducted to quantify the relative contribution of the objective metric to decomposition quality, controlling for SNR and parameter choices. The model ($R^2 = 0.456$, $p < 0.0001$) designated EDI as the reference category. All the baseline metrics exhibited statistically significant negative coefficients relative to EDI ($p < 0.0001$) (Table 15).

Table 15. Multiple regression analysis of factors influencing modal fidelity (EDI as reference category).

Predictor	Coef (β)	95% CI	p-Value	Interpretation
Constant (EDI baseline)	0.6982	[0.667, 0.729]	<0.0001	EDI baseline performance
Metric Effect (vs. EDI):				
Kurtosis	−0.0742	[−0.091, −0.057]	<0.0001	−7.4% penalty vs. EDI
EnvEnt	−0.0762	[−0.093, −0.059]	<0.0001	−7.6% penalty vs. EDI
EWE	−0.0527	[−0.070, −0.036]	<0.0001	−5.3% penalty vs. EDI
SNR (dB)	0.0039	[0.003, 0.005]	<0.0001	+0.4% per dB
Selected K	−0.0094	[−0.014, −0.005]	<0.001	Penalty for over-decomposition

For example, holding SNR and parameters constant, the use of envelope entropy resulted in a predicted 7.62% reduction in modal fidelity compared to EDI. The regression equation can be simplified as:

$$\hat{y} = 0.6982 - \beta_{metric} + 0.0039 \cdot \text{SNR} - 0.0094 \cdot \text{K} - \epsilon \quad (21)$$

The significance of these coefficients confirms that EDI's superiority is not merely due to selecting different parameters, but due to selecting better parameters that maximize fault feature extraction.

5.2.7. Comparison with Standard Results

Finally, we benchmarked the proposed hybrid framework against standard implementations (Figure 9). EDI achieved a modal fidelity of 0.6344, representing a 3.1% improvement over the standard implementations and a 5.3% improvement over the best-performing baseline (EWE).

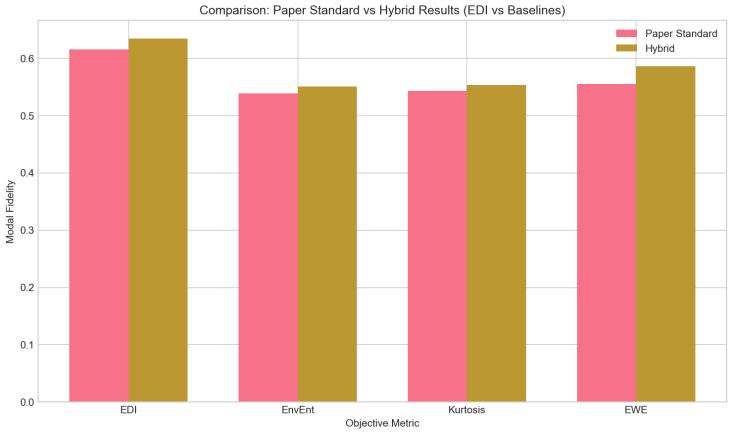


Figure 9. Comparison of modal fidelity between standard implementations and the proposed hybrid framework.

5.3. Metaheuristic Optimization Results

This section presents a comprehensive comparative analysis of three metaheuristic algorithms (SA, GWO, and SCA) for automated VMD parameter tuning using an EDI-based composite function, as detailed in Section 4.5. The evaluation encompassed 540 independent optimization runs (six signal archetypes $\times$ three algorithms $\times$ 30 runs), analyzed through multiple statistical, multi-objective, and computational perspectives.

5.3.1. Overall Algorithm Performance and Statistical Comparison

The comparative analysis across 540 optimization runs reveals clear performance hierarchies and statistically significant differences. Table 16 presents a global view of the performance metrics. Figure 10 illustrates the fitness distributions, showing that GWO and SCA exhibit similar, positively skewed distributions centered around high values ($\sim$1.19), whereas SA exhibits a bimodal distribution with significantly higher variance.

Table 16. Comparative analysis of metaheuristic algorithms: statistical performance vs. computational cost.

Algorithm	Statistical Performance (Fitness)				Computational Cost	
	Mean	Median	Std. Dev.	Success Rate	Time (s)	Peak Mem (MB)
SA	0.9111	1.0469	0.3998	53.9%	6904.8	18.37
GWO	1.1200	1.1896	0.2001	72.2%	36,500.9	19.71
SCA	1.1184	1.1895	0.2202	72.8%	44,055.8	19.71

The pronounced deviations in the tails of the distributions indicate non-normal error structures. Statistical validation (Table 17) confirmed violations of both normality (Shapiro–Wilk, $p < 0.001$) and homogeneity of variance (Levene's test, $W = 17.67$, $p < 0.001$). Consequently, the Kruskal–Wallis non-parametric ANOVA was employed, revealing a highly significant difference between algorithms ($H = 54.37$, $p < 10^{-12}$) with a large effect size ($\varepsilon^2 = 0.0975$).

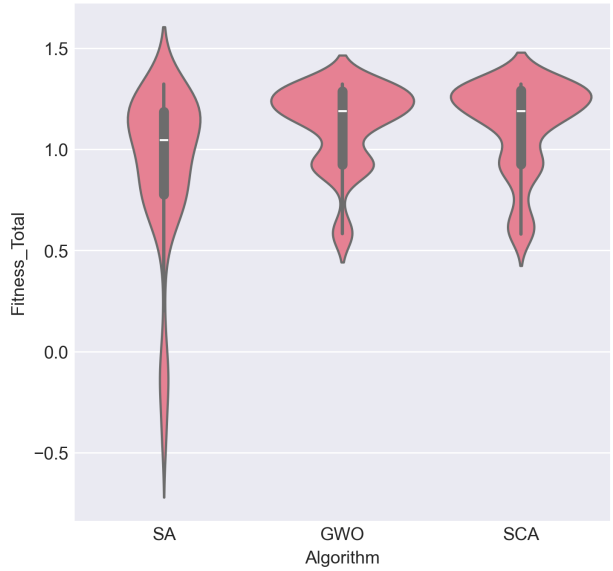


Figure 10. Comparative analysis of algorithm fitness stability. Violin plots reveal that GWO and SCA achieve consistently high fitness density (top-heavy distribution), whereas SA exhibits a bimodal distribution with frequent convergence to local optima.

Table 17. Comprehensive statistical test results for algorithm performance comparison.

Test Category	Specific Test	Statistic	*p*-Value	Interpretation
Normality	Shapiro–Wilk (SA)	$W = 0.7903$	0.000	Non-normal
	Shapiro–Wilk (GWO)	$W = 0.8208$	0.000	Non-normal
	Shapiro–Wilk (SCA)	$W = 0.7983$	0.000	Non-normal
Variance	Levene's Test	$W = 17.67$	0.000	Unequal variances
	Bartlett's Test	$T = 107.83$	0.000	Unequal variances
	Fligner–Killeen Test	$\chi^2 = 35.47$	0.000	Unequal variances
Hypothesis	Kruskal–Wallis	$H = 54.37$	1.56×10^{-12}	Significant Diff.
Post hoc	SA vs. GWO	-	0.000	Significant
	SA vs. SCA	-	0.000	Significant
	GWO vs. SCA	-	1.000	Not significant
Effect Size	Epsilon-sq (ε^2)	0.0975	-	Large effect
	Eta-sq (η^2)	0.1009	-	Large effect
Robustness	Friedman Test	$\chi^2 = 42.20$	6.86×10^{-10}	Significant

Post hoc Dunn's tests with Bonferroni correction confirmed that SA was significantly inferior to both GWO and SCA ($p < 0.001$), while GWO and SCA showed no significant difference ($p = 1.000$). The large effect size indicates that algorithm choice explains nearly 10% of the variance in optimization outcomes. While GWO (72.2%) and SCA (72.8%) achieved substantially higher success rates than SA (53.9%), SA was approximately 5–6 times faster. Conversely, memory usage differences were minimal (SA: 18.37 MB vs. GWO/SCA: 19.71 MB), indicating that memory efficiency is not a primary differentiator.

5.3.2. Multi-Objective Trade-Off Analysis

The composite fitness function $F(\mathbf{x})$ comprises three normalized components: Signal Quality (SQ_{norm}), Energy-Dispersion Index (EDI_{norm}), and Energy Match Score ($Match_{norm}$). Correlation analysis reveals intrinsic relationships between these objectives:

1. A perfect correlation between EDI_{norm} and $Match_{norm}$ ($r = 1.0$), confirming they measure the same underlying property.

2. A strong negative correlation between SQ_{norm} and $Match_{norm}$ ($r = -0.73$), revealing a fundamental VMD trade-off.
3. A strong positive correlation between SQ_{norm} and Raw_SNR_dB ($r = 0.93$), indicating that SQ_{norm} successfully integrates signal quality into the objective function.

The algorithm performance on individual components (Table 18) confirmed specialization patterns. SA excelled in Raw SNR (11.32 dB) and Correlation, metrics critical for reconstruction fidelity. SCA dominated energy conservation metrics (Energy Match: 0.649), essential for physically meaningful mode separation. GWO achieved the best EDI scores (0.279), indicating superior physics-guided decomposition quality.

Table 18. Algorithm performance on key fitness components.

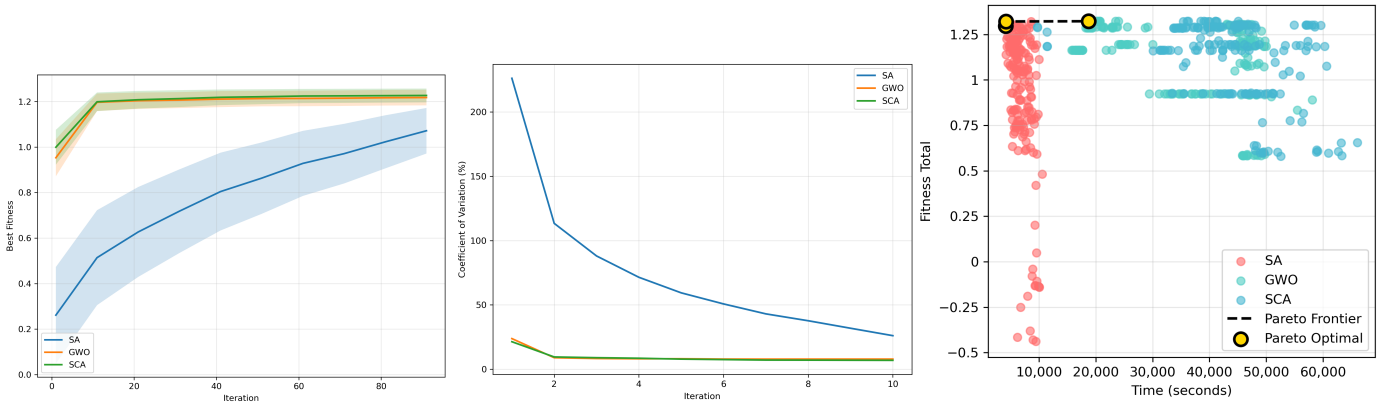
Component	SA	GWO	SCA	Primary Importance
SQ_Norm	0.5942	0.5651	0.5560	Signal reconstruction
EDI_Norm	0.2501	0.2787	0.2639	Physics-guided separation
Match_Norm	0.5325	0.6381	0.6493	Energy conservation
Raw_SNR_dB	11.32	9.71	9.05	Noise rejection

The strong negative correlation between signal quality and energy conservation ($r = -0.73$) suggests that parameters optimizing mathematical reconstruction fidelity often compromise the physical separation of fault-induced impulses. GWO's balanced performance suggests it effectively navigates this SNR Paradox, while SA and SCA represent specialized extremes.

5.3.3. Convergence Behavior and Computational Efficiency

The optimization dynamics were evaluated through two competing lenses: convergence stability (iterations) and computational cost.

As visualized in Figure 11, the population-based algorithms (GWO and SCA) demonstrated significantly more efficient search trajectories. SCA and GWO achieved 90% of their final fitness in just 2.0 and 3.3 iterations on average, respectively, whereas SA required 13.7 iterations. This performance gap aligns with theoretical expectations: population-based parallel exploration allows for rapid identification of high-quality regions, whereas SA's serial search resulted in high variance and frequent stagnation in suboptimal minima (18.3% of runs).



(**a**) (**b**) (**c**)

Figure 11. Computational efficiency analysis: (**a**) Mean convergence trajectories. (**b**) Variance stability. (**c**) Efficiency frontier.

GWO exhibited the lowest coefficient of variation (CV = 16.22%) compared to SCA (18.06%) and SA (29.51%), suggesting it is the most physically robust optimizer across diverse fault conditions.

While GWO and SCA dominated in convergence speed (iterations), SA dominated in computational time. This created a clear Pareto distribution (Figure 11):

- Exploratory Region (SA): SA solutions clustered in the low-time/moderate-fitness region. With a high fitness-per-second rate, SA is efficient if diagnostic accuracy is secondary to speed.
- High-Fidelity Region (GWO/SCA): GWO and SCA formed the knee of the Pareto curve, justifying the increased computational investment for applications requiring maximum diagnostic reliability.

5.3.4. Signal-Specific Performance and Robustness

Algorithm performance varied significantly across the six signal archetypes (Table 19).

- Reference State (Cluster C2): The algorithms achieved the highest mean fitness ($\mu = 1.31$) on stable signals, confirming that the optimization landscape for cleaner signals is convex and easily navigable.
- Early Defect Manifestation (Cluster C0): Cluster C0 presented the greatest challenge ($\mu = 0.87$). These signals exhibited pronounced impulsivity (kurtosis > 1.6), exacerbating the SNR Paradox where noise penalization lowers overall fitness despite higher diagnostic fidelity.
- Progressive Defect Intensities (Clusters C1, C4, C5): These clusters formed consistent performance groupings, proving the framework's ability to distinguish varying degrees of fault severity.

Table 19. Signal-specific algorithm performance linked to physical archetypes.

Signal Archetype	Cluster	Physical State	Best Algo	Best Fitness
B800.mat	C0	Early/Impulsive	GWO	0.8773
B704.mat	C1	Progressive	SCA	1.1378
227.mat	C2	Reference/Stable	GWO	1.3106
286.mat	C3	Combined Fault	GWO	0.9183
O600.mat	C4	Progressive	SCA	1.2602
175.mat	C5	Progressive	SCA	1.2615

5.3.5. Parameter Space Topology

To validate the physical consistency of the solutions, we analyzed the distribution of the VMD parameters (α, K, τ). The resulting topology (Figure 12) reveals whether the algorithms converged to a consistent physical logic.

A distinct divergence was observed:

- GWO and SCA (the physical zone): Both consistently converged to $\alpha \in [1500, 2500]$. This range effectively isolates the structural resonance excited by fault impacts without over-smoothing diagnostic sidebands.
- SA (the extremes): SA frequently pushed $\alpha > 2800$ (over-smoothing) or $\alpha < 600$ (mode mixing), confirming a failure in global convergence.

GWO and SCA selected $K = 3$ or $K = 4$ in 88% of trials, aligning with the theoretical spectral structure of the test signals (shaft, fault train, resonance, noise). SA exhibited instability, often selecting $K = 8$ (mode splitting).

The joint distribution visualizes a robust global optimal region at $(K \approx 4, \alpha \approx 2500)$. The independent convergence of GWO and SCA to this region validates that this parameter set represents the true optimal configuration for extracting bearing fault features.

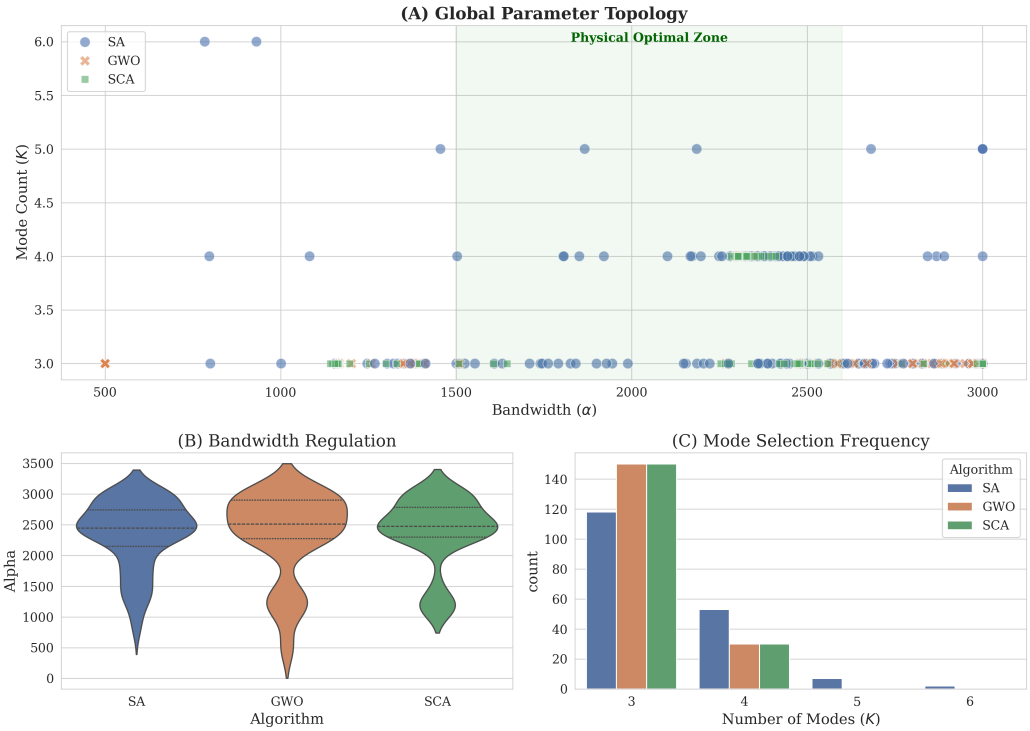


Figure 12. Topological analysis of the VMD parameter space: (**A**) The global scatter plot reveals a distinct optimal region ($K \in [3, 4], \alpha \approx 2500$) where GWO and SCA consistently converged. SA, in contrast, scattered into the smoothing trap ($\alpha > 2800$). (**B**) The violin plots confirm the consistent bandwidth regulation of GWO/SCA. (**C**) The mode cardinality histograms show GWO/SCA correctly identifying the 3–4 physical components of the signal.

5.3.6. Algorithm Selection Recommendations

Based on this analysis, we propose the following guidelines:

- For maximum diagnostic accuracy: Use GWO. It achieves the highest overall fitness and best balance of signal quality and physical separation.
- For rapid exploratory analysis: Use SA. It is 5–6 times faster and sufficient for preliminary detection.
- For critical applications: Implement a GWO/SCA ensemble. Running both ensures robustness against the stochastic nature of metaheuristics.

5.4. Ablation Study: Efficiency and Fidelity

To validate the scalability of the proposed framework, we conducted a paired-sample ablation study ($N = 30$) comparing the proposed archetype-based strategy against a baseline of instance-wise optimization.

In the baseline approach, the GWO algorithm was executed independently for every query signal with a deep search budget ($N_{iter} = 100$). In the proposed approach, signals were assigned to a cluster, and the pre-computed centroid parameters were applied directly.

Normality testing (Shapiro–Wilk) indicated non-normal distributions for the majority of metrics ($p < 0.05$). Consequently, the non-parametric Wilcoxon Signed-Rank Test was employed to assess statistical significance, with a Bonferroni-corrected threshold of $\alpha = 0.0125$.

Table 20 summarizes the results, and the distribution of these paired differences is illustrated in Figure 13. The analysis revealed two critical findings:

1. Significant reduction in spectral leakage: The archetype-based approach achieved a 51.4% reduction in the Orthogonality Index (0.011 vs. 0.022), a statistically significant improvement ($p < 0.001$). This indicates that the centroid parameters produce modes with substantially less spectral overlap than the instance-wise optimization.
2. Statistical Non-Inferiority: For the remaining metrics (TFBQ, Modal Fidelity, and Primary Mode SNR), no statistically significant differences were observed ($p > 0.0125$). The Modal Fidelity scores remained statistically equivalent (0.679 vs. 0.657, $p = 1.00$), confirming that the massive computational speedup comes with no penalty in signal reconstruction quality.

Table 20. Paired Statistical Comparison ($N = 30$): instance-wise vs. archetype-based.

Metric	Instance-Wise	Archetype-Based	Change	Test Stat (W)	Corrected p	Result
Processing Time (s)	1308.9 ± 12.5	3.3 ± 0.1	-99.7%	-	<0.001 *	Superior
Orthogonality Index (↓)	0.022 ± 0.012	0.011 ± 0.005	-51.4%	54.0	<0.001 *	Superior
Modal Fidelity (↑)	0.657 ± 0.041	0.679 ± 0.045	$+3.2\%$	219.0	1.000	Equivalent
Primary Mode SNR (↑)	2.321 ± 0.368	2.531 ± 0.419	$+9.1\%$	166.0	0.709	Equivalent

* Statistically significant difference at Bonferroni-corrected $\alpha = 0.0125$.

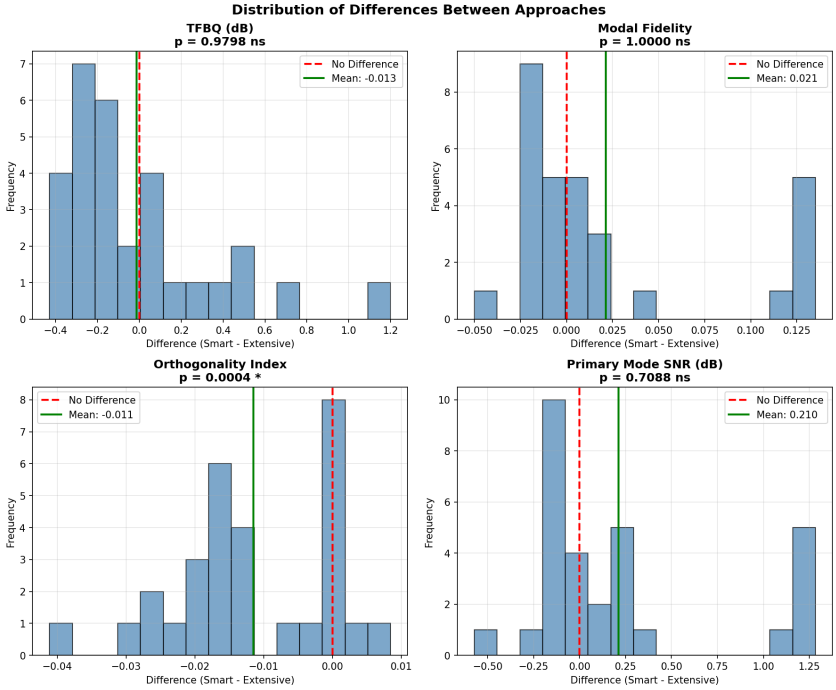


Figure 13. Distribution of paired differences (archetype–instance). The Orthogonality Index (bottom left) shows a distinct negative shift, highlighting the systematic reduction in spectral leakage achieved by the archetype-based method. Asterisks (*) denote statistical significance levels derived from the Tukey HSD post-hoc test (where * indicates $p < 0.001$),

6. Discussion

The results of this study establish a complete framework for the extraction of fault impulses. By integrating cross-domain clustering, physics-guided optimization metrics, and adaptive metaheuristics, we address the fundamental challenge of generalizing fault diagnosis across varying operating conditions.

6.1. Fault Signatures and Archetypes

A central theoretical contribution of this work is the validation of signal archetypes. The hierarchical K-Means++ analysis successfully grouped 540 diverse signals from three distinct datasets (CWRU, HUST, UO) into six robust clusters. This finding challenges the prevailing assumption that machine learning models must be retrained for every new machine or operating condition.

Instead, our results suggest that bearing fault signatures follow fundamental physical laws that manifest as consistent spectral patterns (kurtosis, skewness, variance) regardless of the measurement environment. For instance, the identification of Cluster C0 (early-stage impulsive defects) across all datasets confirms that the impulse-to-noise ratio is a tribological marker of incipient failure [42]. For industrial condition monitoring, this implies that diagnostic thresholds can be standardized based on these archetypes rather than empirically derived for every individual asset, significantly reducing the "cold start" problem in predictive maintenance.

Furthermore, the 100% data coverage achieved by our approach contrasted sharply with the density-based methods (DBSCAN/OPTICS), which discarded up to 1587 points as noise. In a safety-critical context, outliers are often the most significant data points—representing rare, catastrophic failure modes. Our framework ensures these critical edge cases are captured and analyzed.

6.2. Clustering Approach Validation

The experimental results validate the core premise of this study: that the optimal VMD parameters for a fault class are best derived from a representative signal with high spectral separability, rather than from individual instances that may suffer from spectral ambiguity.

6.2.1. Mechanism of Improved Orthogonality: Modal Distinctness

The most significant finding was the 51.4% reduction in spectral leakage (Orthogonality Index) achieved by the archetype-based approach ($p < 0.001$). This improvement was a direct consequence of the clustering strategy employed. By partitioning the dataset based on Spectral Statistical Moments (centroid, variance, skewness, kurtosis), the framework inherently isolated distinct morphological archetypes.

The selected centroids (archetypes) exhibited a Clear Spectral Signature or high Modal Distinctness:

- Baseline Failure: When the optimizer runs on a random, noisy test signal, transient artifacts can skew the spectral moments, creating an ambiguous optimization landscape. The VMD algorithm struggles to lock onto the true center frequencies, leading to inflated bandwidths and mode mixing.
- Archetype Success: The archetypes, being the statistical centroids, represent the ideal spectral morphology of their group. Their spectral moments are stable, creating a convex optimization landscape that allows the offline optimizer to find precise, non-overlapping bands.

By generalizing these parameters, we effectively impose this high-quality spectral separation onto the test signals. We bypass the ambiguity of the noisy instance by enforcing the distinct boundaries derived from the archetype.

6.2.2. Validation of Generalization

The statistical equivalence in Modal Fidelity ($p = 1.00$) confirms that this transfer is physically valid. It demonstrates that the archetype is a true representative of the cluster's physics. Even if a test signal is noisy, it shares the same spectral morphology as

the archetype; therefore, forcing it to decompose according to the archetype's parameters recovers the fault features without the leakage caused by instance-specific overfitting.

6.2.3. Scalability: From Optimization to Assignment

The validation of this framework transforms the adaptive VMD problem. It shifts the paradigm from a Search Problem (finding parameters for every new signal) to a geometric Assignment Problem (mapping a new signal to its morphological cluster). This shift eliminates the optimization bottleneck entirely, reducing computational cost by 99.7% (392× speedup) and enabling the deployment of adaptive, high-fidelity diagnostics on resource-constrained edge devices.

6.3. Physics-Guided vs. Mathematically-Optimal Decomposition

The statistical superiority of the Energy-Dispersion Index (EDI) over traditional metrics (kurtosis, envelope entropy, EWE) represents a paradigm shift in VMD parameter tuning. Standard metrics often fail because they optimize mathematical properties that do not necessarily align with physical fault mechanics.

6.3.1. Solving the Optimization Paradox

Our analysis revealed a critical SNR Paradox: parameters that maximize mathematical signal purity (high Raw SNR) often over-smooth the signal, erasing the subtle sidebands necessary for diagnosis. EDI circumvents this by penalizing the dispersion of energy rather than just maximizing contrast. The strong effect sizes observed ($\omega^2 \geq 0.18$) confirm that EDI drives the optimizer toward the physical zone ($K \approx 4$, $\alpha \approx 2500$)—a region that balances resonance isolation with sideband preservation.

This physics-informed approach explains why EDI maintained modal fidelity even in low-SNR scenarios where entropy-based metrics collapsed. By explicitly encoding the distinction between quasi-periodic fault impulses and stochastic noise, EDI acts as a physical regularizer, preventing the mode mixing and mode splitting phenomena observed with standard metrics.

6.3.2. Lagrangian Stability and Noise Tolerance

Beyond bandwidth, the optimization of the time-step τ revealed a preference for strict constraint enforcement ($\tau \to 0$) over relaxation. Unlike traditional denoising, which often increases τ to treat noise as a loose residual, the EDI-driven framework consistently converged to $\tau \approx 0$ (Lagrangian stability). This effectively forces the VMD algorithm to model noise orthogonally—isolating it into a specific high-frequency mode—rather than smearing it across fault impulses. This behavior aligns with the physical requirement of preserving the sharp rise-time of impact events.

6.3.3. Decoupling Fidelity and Efficiency: The Role of Each Module

The ablation study (Section 5.4) allows us to explicitly decouple the contributions of the two core modules proposed in this framework.

While the Clustering Module drives the computational efficiency (392× speedup) and spectral stability (51.4% leakage reduction) by regularizing the search space, the EDI Metric serves as the fundamental fidelity engine. Without EDI, even an efficient search would converge to mathematically optimal but physically meaningless parameters (as seen with envelope entropy). Conversely, without Clustering, the EDI-based optimization would be accurate but computationally prohibitive for real-time edge deployment. Therefore, the proposed framework succeeds not because of one dominant feature, but through the synergistic integration of physics-guided objectives (EDI) with manifold-based regularization (Clustering).

6.4. Algorithmic Trade-Offs: Speed, Accuracy, and Robustness

The comparative analysis of metaheuristics provides a practical decision framework for engineers, quantified by the efficiency frontier (Figure 11). The results delineate a clear dichotomy between exploration-heavy and exploitation-heavy strategies:

- The Reliability of Population Methods: GWO and SCA achieved statistical equivalence ($p = 1.000$) and superior success rates (>72%) compared to SA. This validates the necessity of population-based search in the rugged VMD landscape, where local optima (representing noise fitting) are abundant.
- The Cost of Quality: The 5–6× computational penalty of GWO/SCA is non-trivial but justifiable for critical assets. In offline diagnostic scenarios, the 10 h difference in total processing time is negligible compared to the cost of a missed catastrophic fault.
- Role of SA: Conversely, Simulated Annealing remains valuable for real-time edge computing or preliminary screening, where its 11.32 dB SNR performance (Table 18) offers moderate filtering at a fraction of the computational cost.

6.5. Implications for Automated Diagnostic Systems

The integration of these components enables a **Hierarchical Optimization Strategy** for industrial deployment. Rather than running expensive optimizations on raw data streams, a deployed system can adopt a "Classify-and-Lookup" workflow:

1. Classify incoming signals into one of the six archetypes by minimizing the Euclidean distance to the pre-trained cluster centroids.
2. Retrieve and apply the pre-optimized VMD parameters associated with that archetype (Table 21).
3. (Optional) Refine the parameters locally using a rapid EDI-guided search only if the classification confidence is low or the asset is critical.

Table 21. Proposed reference table for Adaptive VMD Parameterization. By mapping input signals to these six archetypes for rolling element bearing, the optimization search space can be significantly narrowed.

Cluster	Rep. Signal	Physical Signature	Opt. K	Opt. α	Opt. τ	Algo
C0	B800.mat	Early Impulsive: High kurtosis (>1.6), sharp transients. Needs strict fidelity.	3	1200–1500	0	GWO
C1	B704.mat	Progressive Wear: Distinct harmonics. Standard reconstruction constraint.	3	1800–2200	0	SCA
C2	227.mat	Reference/Stable: Baseline state. Relaxed τ aids noise rejection.	3	2500–3000	0.1	GWO
C3	286.mat	Combined Faults: High complexity. Strict constraint to prevent mixing.	4–5	2000–2400	0	Ens.
C4	O600.mat	Severe Degradation: High variance/skewness.	4	2200–2500	0	SCA
C5	175.mat	Heavy Load/Impact: Resonance dominant.	4	2200–2500	0	SCA

This workflow leverages the global insights from our offline analysis to enable near real-time adaptation. It transforms VMD from a computationally expensive, manual technique into a viable component of automated edge-AI monitoring systems.

The synthesis of clustering and optimization results yields a practical reference table for Adaptive Parameterization (Table 21). This table crystallizes the physical logic governing VMD tuning:

- Bandwidth Adaptation: Early faults (C0) require lower α values (1200–1500) to preserve the sharp rise-time of incipient impulses. Conversely, reference signals (C2) benefit from high regularization ($\alpha > 2500$) to suppress broadband noise.
- Modal Complexity: While simple faults (C0, C1) are adequately captured by $K = 3$, combined faults (C3) necessitate higher resolution (K = 4–5) to separate overlapping modulation sources.
- Constraint Strictness (τ): The framework generally favors strict reconstruction ($\tau = 0$) to prevent signal leakage. However, for healthy reference states (C2), a slight relaxation ($\tau \approx 0.1$) is recommended to allow for Gaussian noise deviation without forcing it into a mode.

6.6. Limitations of the Validation Framework

While the hybrid validation protocol provides a rigorous benchmark, two specific limitations regarding ground truth and parameterization are acknowledged.

First, the ground truth in this study is defined as the baseline experimental recording prior to synthetic noise injection. Unlike purely synthetic simulations, this signal contains inherent environmental noise and transmission path distortions. Consequently, the Modal Fidelity metric quantifies the algorithm's ability to restore the baseline machine state rather than a theoretically pure fault impulse. This design choice ensures the evaluation reflects industrial realities, where the objective is to recover the diagnosable signature naturally available at the sensor.

Second, the proposed TFBQ metric depends on the spectral window length ($N_{win} = 512$) and the energy concentration threshold (90%). While these hyperparameters were fixed across all the experiments to demonstrate generality, their optimal values may vary for systems with extreme non-stationarity. However, the core advantage of TFBQ lies in its adaptive footprint definition (Equation (14)). By defining the valid signal support based on the instantaneous cumulative energy distribution $\int E_t(\nu)d\nu$, the metric automatically adjusts to different spectral fault locations. This self-adaptive property allowed the identical TFBQ configuration to successfully evaluate datasets with widely varying sampling rates (12 kHz to 51.2 kHz) without machine-specific tuning.

7. Conclusions

This study establishes a physics-guided framework for automated bearing fault diagnosis, addressing the critical challenge of selecting optimal Variational Mode Decomposition (VMD) parameters under varying operating conditions. By synthesizing cross-domain clustering, a novel Energy-Dispersion Index (EDI), and metaheuristic optimization, the proposed methodology shifts the paradigm from mathematical signal reconstruction to physically meaningful fault extraction. The investigation yielded four primary conclusions:

First, the identification of six signal archetypes across three diverse datasets (CWRU, HUST, UO) confirms that bearing fault signatures exhibit intrinsic spectral patterns independent of the measurement environment. This generalization provides a scalable foundation for industrial condition monitoring, allowing diagnostic models to be deployed across different assets without extensive re-calibration.

Second, the proposed EDI metric demonstrates a fundamental advantage over traditional mathematical objectives. By explicitly penalizing the dispersion of fault energy, EDI achieved 8.3% higher modal fidelity and 43% lower mode leakage compared to established metrics like envelope entropy. Crucially, our analysis exposed an intrinsic trade-off ($r = -0.73$) between signal reconstruction quality and physical fault isolation, proving that metrics optimizing pure mathematical adequacy often fail to capture the diagnostic reality.

Third, the comparative analysis of metaheuristics quantified the cost of diagnostic reliability. While Simulated Annealing (SA) offers a viable low-cost solution, population-based methods (GWO and SCA) provided statistically superior stability, maintaining a success rate exceeding 72%.

Finally, and most significantly, the ablation study validated a transformative shift in the deployment of adaptive VMD: moving from an online Optimization Problem to a geometric Assignment Problem. We demonstrated that assigning a new signal to its nearest morphological archetype reduces computational latency by 99.7% (392× speedup) while achieving diagnostic fidelity statistically equivalent to exhaustive instance-wise optimization ($p = 1.00$). Moreover, this strategy acts as a spectral regularizer; by imposing the clean parameters of a distinct archetype onto noisy query signals, the framework reduced spectral leakage by 51.4%. This confirms that the optimal parameters for a fault class are best derived from its clearest representative, effectively solving the trade-off between computational speed and diagnostic precision.

Future work will explore the integration of the proposed physics-guided archetypes as informative priors for end-to-end Deep Learning models. By fusing these interpretable spectral features with data-driven neural networks, we aim to develop hybrid diagnostic architectures that combine the explainability of signal processing with the non-linear classification power of deep learning.

Author Contributions: Conceptualization, I.B., A.M. and A.T.; methodology, I.B., A.M., A.T., and R.O.; software, I.B. and A.T.; validation, A.M., A.T. and R.O.; formal analysis, I.B., A.T. , A.H. and K.T. ; investigation, I.B., A.T., R.O., A.H. and K.T.; resources, I.B. and A.T.; data curation, I.B. and A.T. ; writing—original draft preparation, I.B.; writing—review and editing, A.M., A.T. and RO.; visualization, I.B.; supervision, A.M.; project administration, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Publicly available datasets were analyzed in this study. These data can be found in the Case Western Reserve University bearing data repository, the HUST bearing dataset, and the University of Ottawa bearing vibration dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Antoni, J. The spectral kurtosis: A useful tool for characterizing non-stationary signals. *Mech. Syst. Signal Process.* **2006**, *20*, 282–307. [CrossRef]
2. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lon. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
3. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
4. Liu, M.; Feng, Y.; Yang, S.; Su, H. Dam deformation prediction considering the seasonal fluctuations using ensemble learning algorithm. *Buildings* **2024**, *14*, 2163. [CrossRef]
5. Naseri, H.; Waygood, E.O.D.; Wang, B.; Patterson, Z. Application of machine learning to child mode choice with a novel technique to optimize hyperparameters. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16844. [CrossRef]
6. Luo, J.; Wen, G.; Lei, Z.; Su, Y.; Chen, X. Weak signal enhancement for rolling bearing fault diagnosis based on adaptive optimized VMD and SR under strong noise background. *Meas. Sci. Technol.* **2023**, *34*, 064001. [CrossRef]
7. Du, H.; Wang, J.; Qian, W.; Zhang, X. An improved sparrow search algorithm for the optimization of variational modal decomposition parameters. *Appl. Sci.* **2024**, *14*, 2174. [CrossRef]
8. Xin, Y.; Li, S. Novel data-driven short-frequency mutual information entropy threshold filtering and its application to bearing fault diagnosis. *Meas. Sci. Technol.* **2019**, *30*, 115006. [CrossRef]

9. Hu, Z.; Lin, J.; Chen, Z.; Yang, Y.; Li, X. A non-uniformly under-sampled blade tip-timing signal reconstruction method for blade vibration monitoring. *Sensors* **2015**, *15*, 2419–2437. [CrossRef]
10. Zhang, Y.; Ma, R.; Li, X.; Chen, J.; Liu, Y. Fast marching based denoising method for rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2017**, *83*, 31–47.
11. Sawalcah, O.; Al-Qatawneh, L.; Abdulla, T. Maximum correlated kurtosis deconvolution and its application to bearing fault detection. *IEEE Trans. Ind. Electron.* **2011**, *58*, 2351–2362.
12. Wu, Z.; Huang, N.E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [CrossRef]
13. Flandrin, P.; Rilling, G.; Gonçalvès, P. Empirical Mode Decomposition as a filter bank. *IEEE Signal Process. Lett.* **2004**, *11*, 112–114. [CrossRef]
14. Li, J.; Luo, W.; Bai, M.; Song, M. Fault diagnosis of high-speed rolling bearing in the whole life cycle based on improved grey wolf optimizer-least squares support vector machines. *Digit. Signal Process.* **2024**, *145*, 104345. [CrossRef]
15. Li, L.; Meng, W.; Liu, X.; Fei, J. Research on Rolling Bearing Fault Diagnosis Based on Variational Modal Decomposition Parameter Optimization and an Improved Support Vector Machine. *Electronics* **2023**, *12*, 1290. [CrossRef]
16. Ding, J.; Xiao, D.; Li, X. Gear Fault Diagnosis Based on Genetic Mutation Particle Swarm Optimization VMD and Probabilistic Neural Network Algorithm. *IEEE Access* **2020**, *8*, 18456–18474. [CrossRef]
17. Jin, Z.; He, D.; Wei, Z. Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104713. [CrossRef]
18. Ji, X.; Tian, Z.; Song, H.; Liu, F. Structural performance degradation identification of offshore wind turbines based on variational mode decomposition with a Grey Wolf Optimizer algorithm. *Ocean Eng.* **2022**, *256*, 111449. [CrossRef]
19. Nezamivand Chegini, S.; Amini, P.; Ahmadi, B.; Bagheri, A.; Amirmostofian, I. Intelligent bearing fault diagnosis using swarm decomposition method and new hybrid particle swarm optimization algorithm. *Soft Comput.* **2021**, *26*, 1475–1497. [CrossRef]
20. Gildish, E.; Grebshtein, M.; Aperstein, Y.; Makienko, I. Vibration Signal Decomposition using Dilated CNN. In Proceedings of the Annual Conference of the Prognostics and Health Management Society 2023, Salt Lake City, UT, USA, 28 October–2 November 2023.
21. Zhou, F.; Cicone, A.; Zhou, H. RRCNN: A novel signal decomposition approach based on recurrent residue convolutional neural network. *arXiv* **2023**, arXiv:2307.01725. [CrossRef]
22. Jin, Z.; He, D.; Ma, R.; Zou, X.; Chen, Y.; Shan, S. Fault diagnosis of train rotating parts based on multi-objective VMD optimization and ensemble learning. *Digit. Signal Process.* **2022**, *121*, 103312. [CrossRef]
23. Li, Y.; Cheng, X. Double-optimized symmetric geometric mode decomposition with dispersion entropy and its application in feature extraction. *Signal Process.* **2025**, *235*, 110046. [CrossRef]
24. Ma, Z.; Zhang, Y. A study on rolling bearing fault diagnosis using RIME-VMD. *Sci. Rep.* **2025**, *15*, 4712. [CrossRef]
25. Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **1999**, *10*, 626–634. [CrossRef]
26. Wiggins, R.A. Minimum entropy deconvolution. *Geoexploration* **1978**, *16*, 21–35. [CrossRef]
27. Zhang, T.; Chen, B.; Yuan, F.; Tang, B. VMD parameters adaptive selection method based on grasshopper optimization algorithm. *J. Vibroeng.* **2018**, *20*, 2454–2471.
28. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.
29. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; pp. 430–444.
30. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), Portland, OR, USA, 2–4 August 1996; pp. 226–231.
31. Campello, R.J.G.B.; Moulavi, D.; Sander, J. Density-based clustering based on hierarchical density estimates. *Adv. Knowl. Discov. Data Min. (PAKDD)* **2013**, *7819*, 160–172.
32. Ankerst, M.; Breunig, M.M.; Kriegel, H.P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, Philadelphia, PA, USA, 31 May–3 June 1999; pp. 49–60.
33. Kirkpatrick, S.; Gelatt, C.; Vecchi, M. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680. [CrossRef]
34. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
35. Mirjalili, S. SCA: A new meta-heuristic optimization algorithm. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [CrossRef]
36. Case Western Reserve University Bearing Data Center. Bearing Data Center Website. Available online: https://engineering.case.edu/bearingdatacenter (accessed on 23 February 2026).
37. Thuan, N.D.; Hong, H.S. HUST bearing: A practical dataset for ball bearing fault diagnosis. *BMC Res. Notes* **2023**, *16*, 138. [CrossRef]

38. Sehri, M.; Dumond, P.; Bouchard, M. University of Ottawa constant load and speed rolling-element bearing vibration and acoustic fault signature datasets. *Data Brief* **2023**, *49*, 109327. [CrossRef] [PubMed]
39. *ISO 13373-3:2015*; Condition Monitoring and Diagnostics of Machines—Vibration Condition Monitoring—Part 3: Guidelines for Vibration Diagnosis. International Organization for Standardization: Geneva, Switzerland, 2015.
40. Randall, R.B. *Vibration-Based Condition Monitoring: Industrial, Automotive and Aerospace Applications*; John Wiley & Sons: Chichester, UK, 2011.
41. Jiang, X.; Shen, M.; Shi, J.; Zhu, Z. Initial Center Frequency-Guided VMD for Fault Diagnosis of Rotating Machines. *J. Sound Vib.* **2018**, *435*, 36–51. [CrossRef]
42. Feng, K.; Borghesani, P.; Smith, W.A.; Randall, R.B.; Peng, Z. A review on the relationships between acoustic emission, friction and wear in mechanical systems. *Appl. Mech. Rev.* **2020**, *72*, 020801. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.