A Novel Bearing Fault Diagnosis Method Based on Singular Spectrum Decomposition and a Multi-Strategy Enhanced Cuckoo Search–Optimized Extreme Learning Machine

Chengxu Tang *, Yuzhu Ran , Tokunbo Ogunfunmi

Posted Date: 16 October 2025

doi: 10.20944/preprints202510.1308.v1

Keywords: Singular Spectrum Decomposition; Cuckoo Search Algorithm; Multiscale Permutation Entropy; Cauchy Mutation; Adaptive Levy Flight

 doi:10.20944/preprints202510.1308.v1

Article

A Novel Bearing Fault Diagnosis Method Based on Singular Spectrum Decomposition and a Multi-Strategy Enhanced Cuckoo Search–Optimized Extreme Learning Machine

Chengxu Tang [1,*], Yuzhu Ran [2] and Tokunbo Ogunfunmi [1]

[1] Department of Electrical and Computer Engineering, Santa Clara University, Santa Clara, CA 95050 USA
[2] School of Computer Science and Engineering, University of New South Wales, Sydney
* Correspondence: ctang2@scu.edu

Abstract

Large background noise, difficulty in feature extraction, and low parameter-optimization efficiency of diagnosis models are key challenges in rolling bearing fault diagnosis. To address these issues, this paper proposes a fault diagnosis framework that combines Singular Spectrum Decomposition (SSD) with a Multi-Strategy Enhanced Cuckoo Search (MS-CS) algorithm to optimize an Extreme Learning Machine (ELM). First, the raw vibration signal is decomposed via SSD and each intrinsic component's energy contribution is computed; components whose cumulative energy exceeds 90 % are retained and reconstructed, thereby effectively suppressing noise while preserving critical fault features. Next, Multiscale Permutation Entropy (MPE) is extracted from the reconstructed signal to form a high-discriminability feature set. To overcome the traditional Cuckoo Search algorithm's tendency to become trapped in local optima and its slow convergence, Cauchy mutation and adaptive Levy flight strategies are introduced to enhance global exploration and local exploitation. Finally, the improved MS-CS algorithm is employed to optimize the ELM's input weights and hidden-layer biases, yielding a high-precision diagnostic model. Experimental results on benchmark bearing data demonstrate an average fault recognition rate of 96 %, representing improvements of 6.67 % over the conventional CS-ELM and 18 % over the unoptimized ELM. These findings confirm the proposed method's effectiveness and robustness in practical engineering applications.

Keywords: Singular Spectrum Decomposition; Cuckoo Search Algorithm; Multiscale Permutation Entropy; Cauchy Mutation; Adaptive Levy Flight

1. Introduction

Rolling bearings serve as critical support components in a wide range of rotating machinery, and their operational state directly influences equipment safety, reliability, and productivity. Extensive engineering experience demonstrates that bearing failures are among the primary causes of unplanned downtime and serious safety incidents; therefore, accurate and timely fault diagnosis is essential for ensuring the stable operation of industrial systems and reducing maintenance costs. However, vibration signals collected under real-world operating conditions are often contaminated by mechanical friction, oil-film deformation, and environmental disturbances, exhibiting strong nonstationarity and nonlinearity that pose significant challenges to both feature extraction and diagnostic model construction.

Traditional time-domain and frequency-domain analyses can provide preliminary fault discrimination under simple operating conditions but frequently fail in the presence of high noise levels or variable speeds, as fault-induced signal components may be masked or distorted. To address these limitations, researchers have developed numerous adaptive signal decomposition and

 doi:10.20944/preprints202510.1308.v1

denoising techniques. Empirical Mode Decomposition (EMD) adaptively extracts intrinsic mode functions but suffers from mode mixing; Variational Mode Decomposition (VMD) isolates modes by constrained optimization yet requires careful tuning of mode number and penalty parameters and may still degrade under heavy noise; Empirical Wavelet Transform (EWT) improves noise suppression through adaptive band splitting but remains sensitive to initial band allocation and often demands manual intervention. In contrast, Singular Spectrum Decomposition (SSD) constructs a trajectory matrix of the vibration signal and applies singular value decomposition to extract components according to their energy contribution, achieving effective noise attenuation while preserving fault-related features with greater robustness and reduced parameter sensitivity.

At the feature-extraction stage, the nonstationarity and nonlinearity of bearing vibration signals render single-scale or single-domain measures inadequate for fully characterizing fault patterns. Although techniques such as wavelet packet energy entropy and spectral energy density can reflect energy distribution changes to some extent, they are sensitive to noise and require a priori selection of decomposition levels. Entropic measures like singular spectrum entropy enhance information representation but neglect the multi-scale temporal structure of the signal. Multiscale Permutation Entropy (MPE) addresses these shortcomings by reconstructing the signal at multiple scales and computing the permutation entropy at each resolution, thereby quantifying the complexity and regularity of the time series across scales and demonstrating strong noise resistance. MPE thus yields a high-discriminability feature set for fault recognition.

In terms of model construction and parameter optimization, the Extreme Learning Machine (ELM) has attracted considerable attention owing to its single-hidden-layer architecture and rapid training via a closed-form solution. Nevertheless, the performance of ELM heavily depends on the random initialization of input weights and hidden biases, leading to unstable generalization and suboptimal diagnostic accuracy. To enhance ELM robustness and precision, swarm-intelligence algorithms such as Particle Swarm Optimization (PSO) and Cuckoo Search (CS) have been employed, yet both face challenges in high-dimensional search spaces, including loss of population diversity, entrapment in local optima, and slow convergence.

To overcome these issues, this paper proposes a Multi-Strategy Enhanced Cuckoo Search (MS-CS) algorithm that incorporates Cauchy mutation and adaptive Levy flight into the standard CS framework. By dynamically adjusting the flight step size and introducing heavy-tailed mutation, MS-CS achieves a balance between global exploration and local exploitation, thereby efficiently optimizing the input weights and hidden biases of ELM. Combining SSD-based signal reconstruction with MPE feature extraction and MS-CS-optimized ELM, we develop a comprehensive bearing fault diagnosis framework. Experimental validation using benchmark bearing datasets demonstrates that MS-CS-ELM significantly outperforms both the unoptimized ELM and the conventional CS-ELM across multiple fault types, confirming the efficacy and robustness of the proposed methodology.

The remainder of this paper is organized as follows. Section 2 details the principles and implementation of SSD signal reconstruction and MPE feature extraction. Section 3 describes the design rationale and algorithmic steps of the MS-CS method. Section 4 introduces the construction of the MS-CS-ELM diagnostic model. Section 5 presents comparative simulations and performance analysis. Finally, Section 6 summarizes the contributions of this study and outlines directions for future research.

2. Singular Spectrum Decomposition Reconstruction and Multiscale Permutation Entropy Extraction

2.1. Singular Spectrum Decomposition

SSD is employed to isolate fault-related components from noise-corrupted vibration signals by constructing a trajectory (Hankel) matrix from embedded signal vectors and performing spectral decomposition. The principal steps include signal embedding, covariance matrix construction, singular value decomposition (SVD), and component reconstruction, as detailed below.

(1) Signal embedding and trajectory matrix construction.

Let the one-dimensional vibration signal of length N be:

 doi:10.20944/preprints202510.1308.v1

$$x = [x_1, x_2, \ldots, x_N]^{\mathrm{T}} \quad (1)$$

Choosing an embedding dimension (window length) L and a sliding step of one sample, one forms the trajectory matrix:

$$\mathrm{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix}, K = N - L + 1, \quad (2)$$

Where, each column $[\mathbf{X}]_{:,j}$ represents an L-dimensional time-delay embedding at time j. This procedure reconstructs the scalar time series into an L-dimensional phase space, providing a richer representation for spectral analysis.

(2) Covariance matrix and singular value decomposition.

The normalized covariance matrix of X is computed as:

$$\mathrm{C} = \frac{1}{K}\mathrm{X}\mathrm{X}^{\mathrm{T}} \quad (3)$$

Where, $\mathrm{C} \in \mathbb{R}^{L \times L}$ encodes energy couplings among delayed components. Applying SVD to C yields:

$$\mathrm{C} = \mathrm{U}\Lambda\mathrm{U}^{\mathrm{T}} \quad (4)$$

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_L), \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0, \quad (5)$$

Where, $\mathrm{U} = [\mathrm{u}_1, \mathrm{u}_2, ..., \mathrm{u}_L]$ is the orthonormal eigenvector matrix and λ is the corresponding eigenvalues. Large eigenvalues correspond to principal signal structures (including fault features), whereas smaller eigenvalues primarily reflect noise or random fluctuations.

(3) Component grouping and time series reconstruction.

Based on either cumulative energy ratio or the "elbow" criterion of the eigenvalue spectrum, the first r eigenvectors are assigned to the signal subspace and the remaining $L-r$ to the noise subspace. The projection matrix onto the signal subspace is:

$$\mathrm{P} = \sum_{i=1}^{r} \mathrm{u}_i \mathrm{u}_i^{\mathrm{T}} \quad (6)$$

Applying P to the trajectory matrix produces the signal component matrix:

$$X_s = \mathrm{PX} \quad (7)$$

Finally, diagonal averaging of X_s reconstructs the one-dimensional signal x_r via:

$$x_r(t) = \frac{1}{N_t} \sum_{i+j=t+1} [X_s]_{i,j}, t = 1, 2, ..., N, \quad (8)$$

Where, N_t normalizes by the number of elements on the (i,j) anti-diagonal. This procedure yields a multi-component time series reconstruction based on energy-spectrum selection.

(4) Component selection and fault feature preservation.

The truncation index r is chosen such that the cumulative energy ratio of the first r singular values satisfies:

$$\frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{L} \lambda_i} \geq \eta \quad (9)$$

Where, the threshold η is typically set between 0.8 and 0.95 to ensure retention of major signal structures.

Through these steps, SSD effectively suppresses random noise and enhances the signal-to-noise ratio of fault-related information, thus establishing a solid foundation for subsequent multiscale permutation entropy feature extraction.

2.2. Multiscale Permutation Entropy Extraction

MPE quantifies the complexity and regularity of nonstationary time series over multiple temporal resolutions. Unlike single-scale entropy measures, MPE employs a coarse-graining

 doi:10.20944/preprints202510.1308.v1

procedure that captures signal dynamics at various scales, thus providing a more comprehensive characterization of fault signatures.

Let the reconstructed one-dimensional signal be $x=[x_1,x_2,...,x_N]$. For a given scale factor τ, the coarse-grained series $y^{(\tau)}=[y_1^{(\tau)},y_2^{(\tau)},...,y_{N_\tau}^{(\tau)}]$ of length $N_\tau=[N/\tau]$ is defined by:

$$y_j^{(\tau)}=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau}x_i,\ j=1,2,...,N, \tag{10}$$

Where, τ controls the window size, averaging each block of τ samples to suppress high-frequency noise and emphasize lower-frequency components.

For each coarse-grained sequence $y^{(\tau)}$, select an embedding dimension m and time delay l to construct the vector series:

$$v_i=[y_i^{(\tau)},y_{i+l}^{(\tau)},\ldots,y_{i+ml}^{(\tau)}],\ i=1,2,\ldots,N_\tau-(m-1)l. \tag{11}$$

Each vector v_i is mapped to a permutation pattern π of length m according to the relative ordering of its elements. Denoting by $P=N_\tau-(m-1)l$ the total number of vectors, one obtains the probability distribution $\{p_j\}_{j=1}^{m!}$ of all $m!$ possible patterns, satisfying:

$$\sum_{j=1}^{m!}p_j=1 \tag{12}$$

The permutation entropy at scale τ is then defined as:

$$\mathrm{PE}^{(\tau)}(m,l)=\sum_{\pi\in\Pi}p(\pi)\ln p(\pi) \tag{13}$$

which, upon normalization by $\ln(m!)$, yields:

$$\mathrm{NPE}^{(\tau)}=\frac{\mathrm{PE}^{(\tau)}}{\ln(m!)}\in[0,1], \tag{14}$$

Where, values approaching zero indicate greater regularity and values approaching one indicate higher randomness.

By varying τ from 1 to a maximum scale $\tau_{\max}$ and computing $\mathrm{NPE}^{(\tau)}$ at each level, one constructs the multiscale feature vector:

$$\mathrm{F}=[\mathrm{NPE}^{(1)},\mathrm{NPE}^{(2)},...,\mathrm{NPE}^{(\tau\max)}] \tag{15}$$

Which, encapsulates the signal's complexity across multiple time scales. Applying MPE to the SSD-reconstructed signal thus yields a high-discriminability input set for the subsequent population-based optimization and diagnostic modeling.

3 Multi-Strategy Enhanced Cuckoo Search Algorithm (MS-CS): Design Principles and Implementation

3.1 Cuckoo Search Algorithm

The traditional CS algorithm is a population-based optimization method inspired by the brood parasitism of cuckoo species and Levy flight random walks. Each candidate solution is represented by a *d*-dimensional real vector $\mathrm{x}_i\in\mathbb{R}^d$, and a population of n such vectors is initialized by uniform sampling within the search bounds *L* and U:

$$\mathrm{x}_i^{(0)}=\mathrm{L}+(\mathrm{U}-\mathrm{L})\odot\xi_i,\ \ \xi_i\sim u(0,1)^d \tag{16}$$

Where, $\odot$ denotes element-wise multiplication, ensuring a uniform spread of initial solutions.

During each iteration, new solutions are generated via Levy flights to simulate the long-range, heavy-tailed jumps of cuckoos. For the *i*-th individual at iteration *t*, the update rule is:

$$\mathrm{x}_i^{(t+1)}=\mathrm{x}_i^{(t)}+\alpha\mathrm{Levy}(\lambda)(\mathrm{x}_i^{(t)}-\mathrm{x}_*), \tag{17}$$

Where, $\alpha>0$ is a step-size scaling factor, x_* is the current global best solution, and Levy(λ) denotes a random step drawn from a Levy distribution characterized by:

$$p(u)\propto u^{-(1+\lambda)},1<\lambda\le 3. \tag{18}$$

which balances local exploitation and global exploration.

 doi:10.20944/preprints202510.1308.v1

To mimic the host bird's discovery of alien eggs, a fraction p_a of the worst solutions is replaced by new ones drawn by uniform sampling. Specifically, if $I \subset \{1,...,n\}$ is a randomly selected index set of size $|I| \approx p_a n$, then for each $i \in I$,

$$\mathrm{x}_i^{(t+1)} = \mathrm{L} + (\mathrm{U} - \mathrm{L}) \odot \xi_i^{'}, \quad \xi_i \sim u(0,1)^d \tag{19}$$

The remaining individuals retain their Lévy-flight updates. The new solutions are evaluated and compared to the current ones, and the better of the two is retained. This process of Lévy flights, parasitic replacement, and greedy selection iterates until a convergence criterion or maximum iteration count is met, yielding the global optimum.

3.2 Multi-Strategy Enhancement of the Cuckoo Search Algorithm

Although the conventional Cuckoo Search (CS) algorithm excels at global exploration, it may suffer from loss of population diversity and premature convergence in high-dimensional or constrained problems, as well as slow convergence due to inappropriate step-size control. To address these shortcomings, this study integrates Cauchy mutation and adaptive Lévy flight into the CS framework, yielding the Multi-Strategy Enhanced Cuckoo Search (MS-CS) algorithm. After initial population generation, each individual $\mathrm{x}_i^{(t)}$ in generation *t* undergoes Cauchy mutation to enhance diversity:

$$\bar{\mathrm{x}}_j^{(t)} = \mathrm{x}_i^{(t)} + \gamma \mathrm{Cauchy}(0,1)(\mathrm{x}_i^{(t)} - \mathrm{x}_j^{(t)}) \tag{20}$$

Where, $\gamma > 0$ is the mutation strength factor, $\mathrm{Cauchy}(0,1)$ is a standard Cauchy random variable, and $\mathrm{x}_j^{(t)}$ is a randomly selected peer. The heavy tails of the Cauchy distribution facilitate large jumps early in the search, helping to escape local optima.

Building on this mutation, MS-CS introduces an adaptive mechanism for the Lévy flight parameters. Let $r = t/T$ denote the normalized iteration count (with T the maximum number of iterations). The step-size factor α decays exponentially according to:

$$\alpha^{(t)} = \alpha_0 \exp(-kr) \tag{21}$$

while the Lévy distribution parameter λ increases linearly:

$$\lambda^{(t)} = \lambda_{\min} + (\lambda_{\max} - \lambda_{\min})r \tag{22}$$

These adjustments ensure broad exploration in early iterations and fine-grained local search in later stages. Accordingly, each mutated individual $\bar{\mathrm{x}}_i^{(t)}$ is updated via adaptive Lévy flight:

$$x_i^{(t+1)} = \bar{\mathrm{x}}_i^{(t)} + \alpha^{(t)} \mathrm{Levy}(\lambda^{(t)})(\bar{\mathrm{x}}_i^{(t)} - x_*) \tag{23}$$

Where, x_* is the current global best. MS-CS retains the CS probabilistic replacement of a fraction p_a of the worst individuals and applies greedy selection to choose the better solutions. The coordination of Cauchy mutation, adaptive Levy flight, and parasitic replacement enables MS-CS to maintain population diversity while achieving rapid and accurate convergence.

3.3. Optimization of the ELM Network by the Enhanced Cuckoo Search Algorithm

For an ELM with training set $\{(\mathrm{x}_j, \mathrm{t}_j) | j = 1,...,N\}$, where $\mathrm{x}_j \in \mathbb{R}^d$ and $\mathrm{t}_j \in \mathbb{R}^m$, the single-hidden-layer output is given by:

$$y_j = \sum_{i=1}^{L} \beta_i g(\mathrm{w}_i^T \mathrm{x}_j + b_i) \tag{24}$$

Where, L is the number of hidden neurons, $\mathrm{w}_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$ are the input weight vector and bias of neuron i, $\beta_i \in \mathbb{R}^m$ is the output weight, and $g(\cdot)$ is the activation function. Aggregating the network responses yields the hidden-layer output matrix:

$$H = \begin{bmatrix} g(\mathrm{w}_1^T \mathrm{x}_1 + b_1) & \cdots & g(\mathrm{w}_L^T \mathrm{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\mathrm{w}_1^T \mathrm{x}_N + b_1) & \cdots & g(\mathrm{w}_L^T \mathrm{x}_N + b_L) \end{bmatrix} \in \mathbb{R}^{N \times L} \tag{25}$$

Traditionally, $\{\mathrm{w}_i, b_i\}$ are randomly initialized and the output weight matrix $\mathrm{B} \in \mathbb{R}^{L \times m}$ is solved by least squares:

 doi:10.20944/preprints202510.1308.v1

$$\mathrm{B} = \mathrm{H}^{\dagger}\mathrm{T} \tag{26}$$

Where, $\mathrm{T} = [t_1, t_2, ..., t_N]^{\mathrm{T}}$ and $\mathrm{H}^{\dagger}$ is the Moore–Penrose pseudoinverse. Random initialization, however, may yield poorly conditioned H, degrading generalization and stability.

To address this, MS-CS is employed to optimize the concatenated parameter vector:

$$\mathrm{z} = [\mathrm{w}_1^{\mathrm{T}}, b_1, ..., \mathrm{w}_L^{\mathrm{T}}, b_L]^{\mathrm{T}} \in \mathbb{R}^{(d+1)L} \tag{27}$$

Using the training error sum of squares as the objective:

$$f(\mathrm{z}) = \left\| \mathrm{H}(\mathrm{z})\mathrm{H}(\mathrm{z})^{\dagger}\mathrm{T} - \mathrm{T} \right\|_F^2 \tag{28}$$

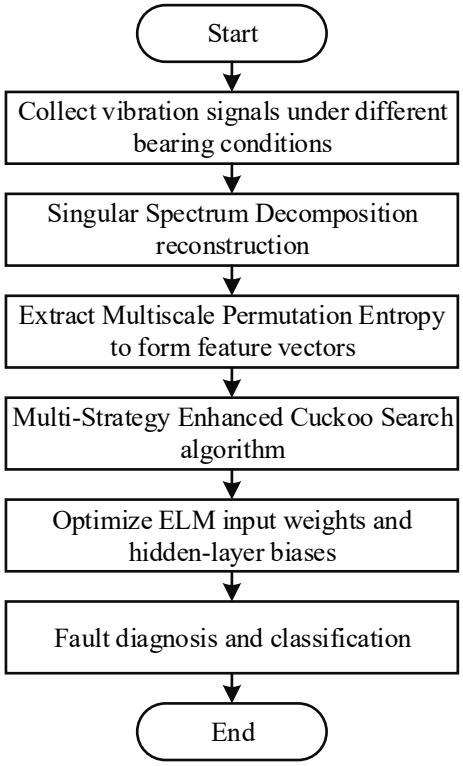
Where, $\|\cdot\|_F^2$ denotes the Frobenius norm and $\mathrm{H}(\mathrm{z})$ emphasizes the dependence of H on z.

During MS-CS optimization, candidate vectors z_i are generated via the multi-strategy mutation and adaptive Lévy flight. For each z_i, one reconstructs $\mathrm{H}(\mathrm{z}_i)$, computes $\mathrm{B}_i = \mathrm{H}(\mathrm{z}_i)^{\dagger}\mathrm{T}$, and evaluates $f(\mathrm{z}_i)$.

MS-CS iteratively updates the population based on these objective values until convergence to the optimal z^*. The resulting ELM, configured by z^*, achieves maximal diagnostic accuracy and robustness. Through this approach, MS-CS-ELM combines the rapid analytic training of ELM with the global search capabilities of the enhanced CS, making it well suited for bearing fault diagnosis under noisy, complex conditions.

4. MS-CS-ELM Fault Diagnosis Model

The MS-CS-ELM fault diagnosis model integrates three core components—signal preprocessing, feature extraction, and intelligent optimization learning—to form a comprehensive pipeline from raw vibration data to fault classification. Initially, the acquired vibration signals are decomposed and reconstructed using SSD to suppress noise and amplify fault-related components. Next, MPE features are computed from the reconstructed signals, yielding feature vectors that capture dynamic complexity across multiple temporal scales. These feature vectors are then fed into an ELM, whose input weights and hidden-layer biases are optimized by the MS-CS algorithm to minimize training error. Finally, the optimized ELM model performs online diagnosis and classification of bearing fault types. A schematic of this diagnostic framework is presented in Figure 1.



 doi:10.20944/preprints202510.1308.v1

Figure 1. MS-CS-ELM fault diagnosis process.

5. Experimental Results and Comparative Analysis

5.1. Experimental Setup

The validation experiments were conducted on the widely used Xi'an Jiaotong University rolling bearing test rig, which comprises a digitally controlled hydraulic loading system, an AC motor coupled to a shaft through a motor speed controller, and support bearings mounting the test bearing under varying load conditions. A digital force display monitors the applied radial load, while vertical and horizontal accelerometers are mounted on the bearing housing to capture vibration responses. The test bearing itself can be configured in one of four distinct states— inner race fault, outer race fault, or cage fault—by introducing controlled defects of known dimensions and locations.

During each run, the motor speed controller maintains a constant rotational speed while the hydraulic system applies predetermined radial loads. Vibration signals are recorded simultaneously from both accelerometers at a high sampling rate to ensure accurate capture of transient and steady-state features. This setup enables systematic collection of multi-condition bearing data under repeatable and well-controlled experimental conditions. The resulting dataset thus provides a robust benchmark for evaluating the proposed SSD–MPE–MS-CS–ELM diagnostic framework across normal and various fault scenarios.

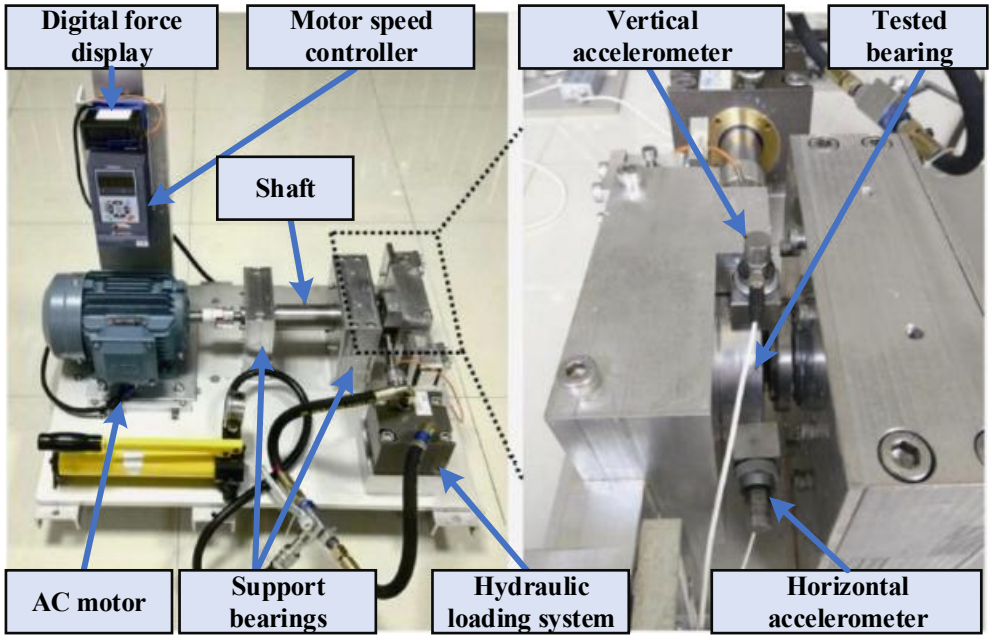


Figure 2. Diagram of experimental device.

The test bearing selected for this study is the LDK UER204 rolling bearing, whose specifications are listed in Table 1. Three operating conditions were designed: Condition a at 2100 r/min and 12 kN radial load; Condition b at 2250 r/min and 11 kN; and Condition c at 2400 r/min and 10 kN. For each condition, five bearings were tested. This rig, under well-controlled speed and load parameters, emulates real-world operating influences on bearing vibration, thereby providing a reliable data foundation for subsequent fault diagnosis model development.

Table 1. Specifications of the LDK UER204 Bearing.

Parameter	Value	Parameter	Value
Inner raceway diameter (mm)	29.30	Ball diameter (mm)	7.92
Outer raceway diameter (mm)	39.80	Number of balls	8

 doi:10.20944/preprints202510.1308.v1

Bearing pitch diameter (mm)	34.55	Contact angle (°)	0
Basic dynamic load rating (N)	12820	Basic static load rating (kN)	6.65

5.2. SSD Signal Decomposition and Reconstruction

Vibration signals acquired under actual operating conditions are often contaminated by mechanical friction, oil-film deformation, and environmental disturbances, exhibiting significant nonstationarity and high noise levels. Consequently, direct application of time-domain or frequency-domain methods fails to extract fault features accurately. Furthermore, different fault types of manifests in distinct spectral components and energy distributions, making it difficult for conventional filtering to simultaneously preserve fault signatures and suppress noise. To address these challenges, this study applies Singular Spectrum Decomposition (SSD) to decompose the raw signal and reconstruct only those components whose cumulative singular value energy exceeds a threshold of 0.9. Figures 3 and 4 respectively show the original outer-race fault signal and its SSD-decomposed components.

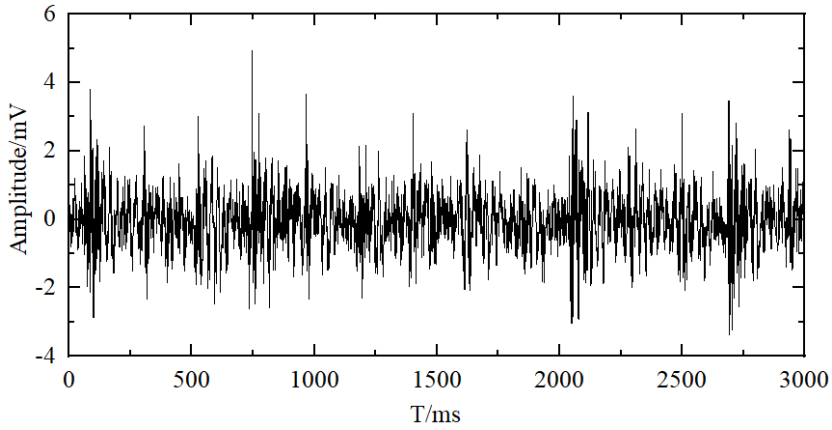


Figure 3. Original Outer-Race Fault Signal.

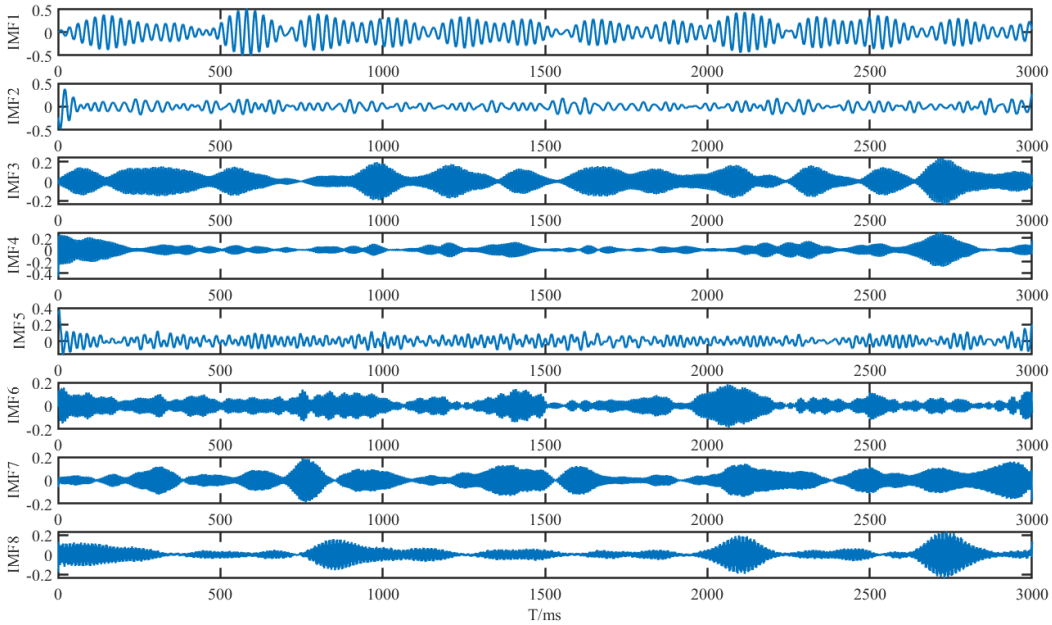


Figure 4. SSD Decomposed Components of Outer-Race Fault Signal.

As illustrated in the figures, the SSD-decomposed components exhibit markedly different energy contributions, with only the high-energy modes containing pertinent fault signatures, while

 doi:10.20944/preprints202510.1308.v1

the low-energy modes predominantly consist of random noise or background interference. Consequently, it is necessary to select components judiciously to avoid the influence of irrelevant modes on subsequent analysis. In this study, the singular value energy ratio method is adopted: the sum of the first r singular values, ordered in descending magnitude, is divided by the total sum of all singular values, and a threshold of 0.9 is imposed. This criterion ensures that the reconstructed signal retains the principal structural components of the original signal while effectively eliminating low-energy noise modes, thus achieving a balance between noise suppression and feature preservation. The signal subspace is formed by the eigenvectors corresponding to the first r singular values, and diagonal averaging is applied to map this subspace back to the time domain, yielding a high-signal-to-noise-ratio reconstructed signal, as shown in Figure 5.

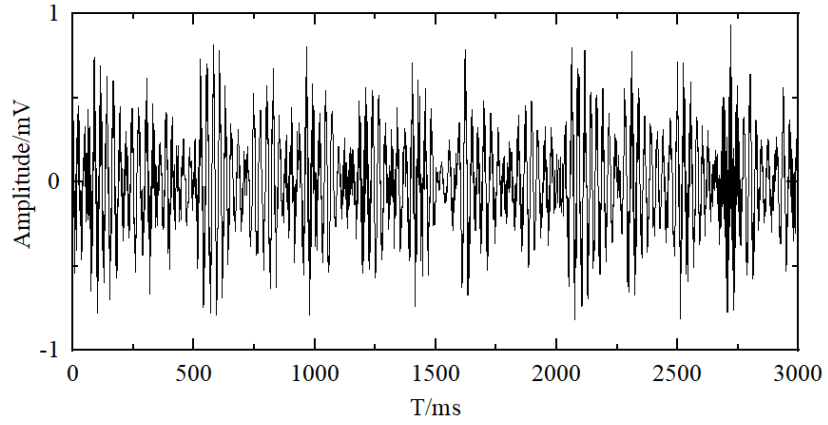


Figure 5. Reconstructed Outer-Race Fault Signal.

The reconstructed signal shows pronounced enhancement of fault impulses and clear periodicity in the time domain, while its signal-to-noise ratio (SNR) improves by approximately 5–8 dB over the raw signal, thereby boosting the accuracy and robustness of subsequent MPE feature extraction. Similar reconstruction results for inner race and cage faults are presented in Figures 6 and 7.

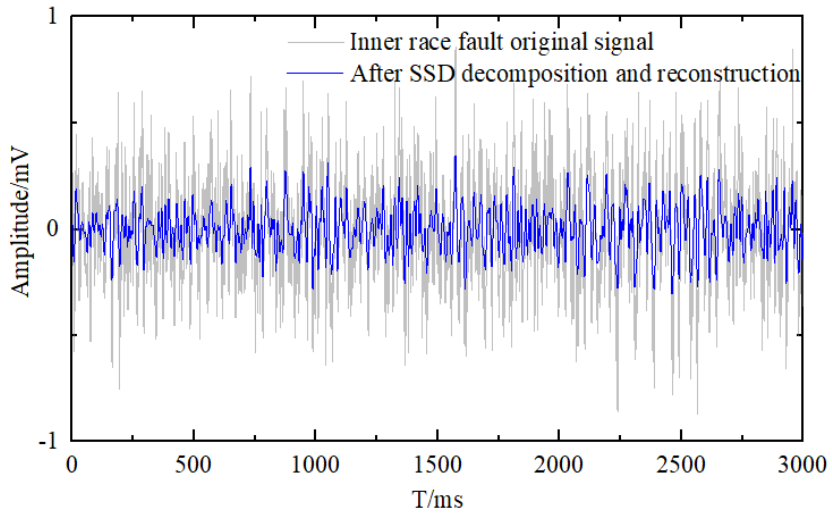


Figure 6. Inner Race Fault: Before and After SSD Reconstruction.

 doi:10.20944/preprints202510.1308.v1

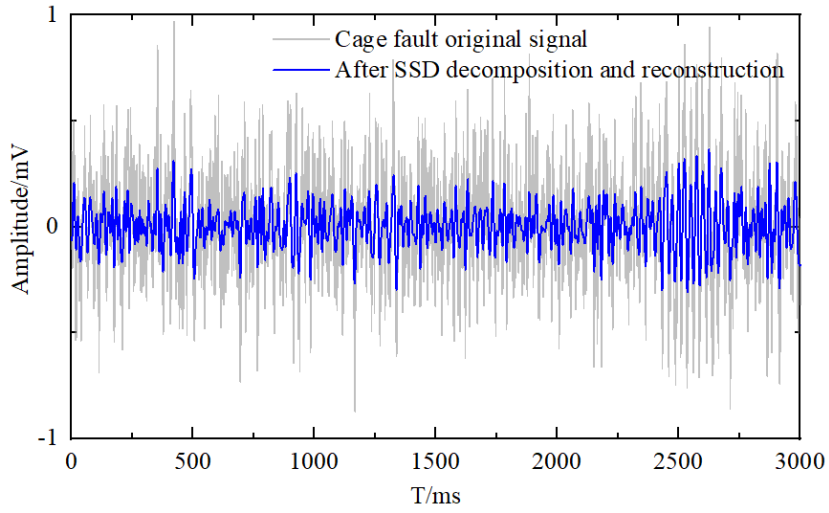


Figure 7. Cage Fault: Before and After SSD Reconstruction.

Overall, SSD reconstruction preserves key fault information while effectively suppressing random noise, providing high-quality input for the diagnostic algorithms.

5.3. Fault Diagnosis Comparative Experiments and Analysis

From the high-SNR reconstructed signals, Multiscale Permutation Entropy (MPE) is computed across multiple scale factors to form feature vectors that capture complexity at various time resolutions. For model training and testing, inner race faults, outer race faults, and cage faults are labeled as 1, 2, and 3, respectively. These feature vectors are then applied to three diagnostic methods—traditional ELM, CS-ELM (ELM optimized by conventional Cuckoo Search), and the proposed MS-CS-ELM—for comparative validation. The diagnostic results for each method are shown in Figures 8–10.

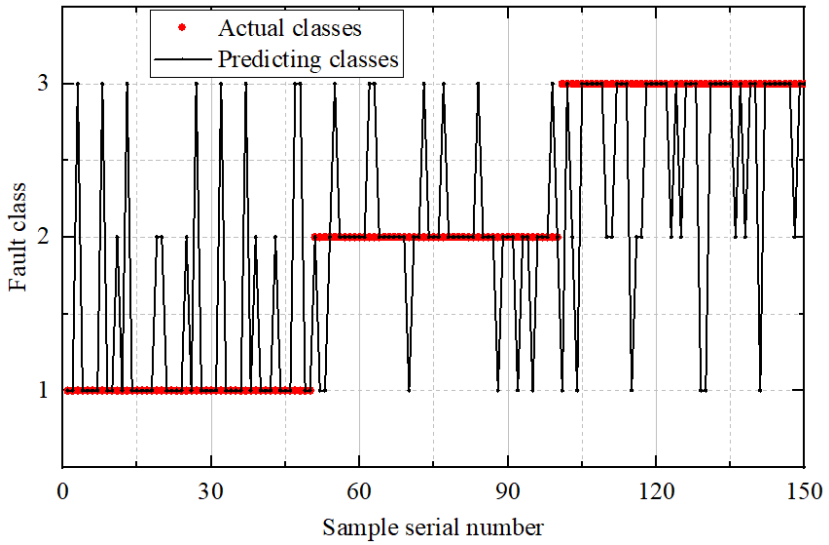


Figure 8. Diagnostic Results of the ELM Model.

 doi:10.20944/preprints202510.1308.v1

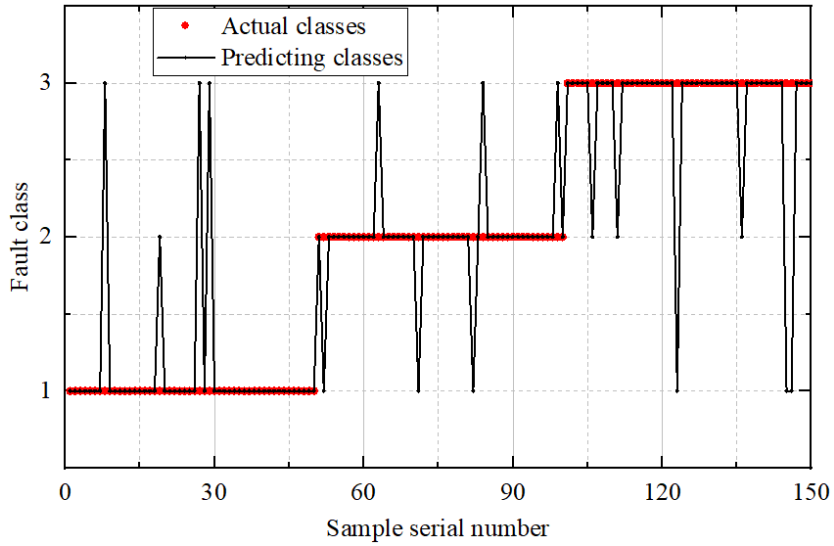


Figure 9. Diagnostic Results of the CS-ELM Mode.

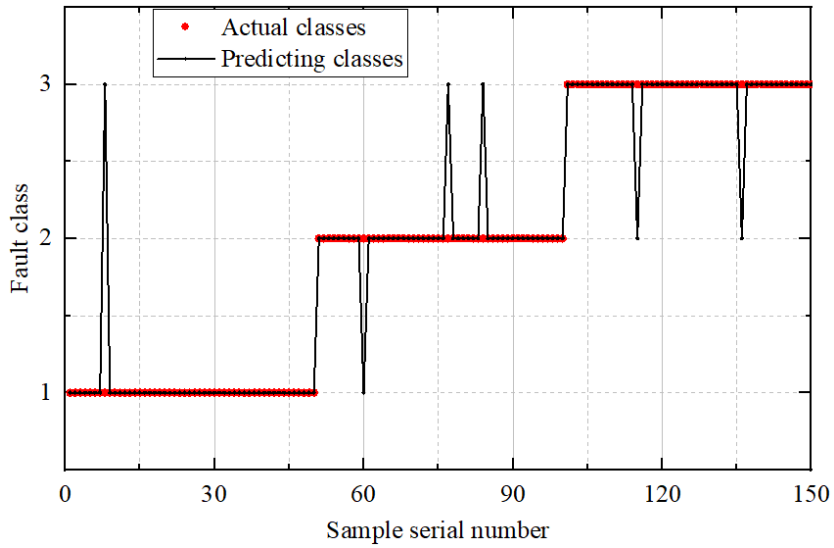


Figure 10. Diagnostic Results of the MS-CS-ELM Model.

The comparative results indicate that the proposed MS-CS -ELM framework substantially outperforms both the unoptimized ELM and the CS-ELM across all fault categories. Specifically, the baseline ELM yields relatively low classification accuracies of 0.72 for inner race faults and 0.74 for outer race faults, although it performs better on cage faults (0.88). When the conventional Cuckoo Search algorithm is employed to optimize the ELM parameters, the CS-ELM improves recognition rates markedly for inner and outer faults—achieving 0.92 and 0.88, respectively—while maintaining the same accuracy of 0.88 for cage faults. However, the multi-strategy enhancements introduced in MS-CS -ELM drive further gains: the inner race fault accuracy rises to 0.98, outer race recognition to 0.94, and cage fault detection to 0.96. These results demonstrate that the adaptive Levy flight and Cauchy mutation strategies in MS-CS effectively guide the search toward superior weight and bias configurations, thereby enhancing the model's ability to capture subtle fault signatures and improving robustness against noise. Overall, MS-CS -ELM achieves relative improvements of 36% on inner race faults, 27% on outer race faults, and 9% on cage faults compared to the baseline ELM, confirming the efficacy of the proposed multi-strategy optimization in achieving high-precision bearing fault diagnosis.

In addition to the comparative analysis of ELM, CS-ELM and MS-CS -ELM, two further experiments were conducted to deepen the evaluation of the proposed framework. The first additional study assessed the diagnostic prowess of ELM by benchmarking it against two

 doi:10.20944/preprints202510.1308.v1

contemporary deep learning models: a one-dimensional convolutional neural network (1D-CNN) designed for end-to-end feature extraction and classification, and a deep belief network (DBN) that applies unsupervised pretraining on MPE features followed by supervised fine-tuning. The second additional study examined the generalizability of the MS-CS optimization strategy by comparing MS-CS -ELM with two other population based optimization methods, namely Particle Swarm Optimization ELM (PSO-ELM) and Grey Wolf Optimizer ELM (GWO-ELM), under the same diagnostic conditions. The results of these experiments are summarized in Table 2.

Table 2. Diagnostic accuracy results of different models.

	Inner race fault	Outer race fault	Cage fault	Average precision
ELM	0.72	0.74	0.88	0.78
CS-ELM	0.92	0.88	0.88	0.8933
MS-CS -ELM	0.98	0.94	0.96	0.96
ID-CNN	0.9	0.92	0.86	0.89
DBN	0.84	0.8	0.86	0.8333
PSO-ELM	0.9	0.9	0.84	0.88
GWO-ELM	0.86	0.88	0.92	0.8867

When benchmarked against deep learning models, the MS-CS -ELM also outperforms the one-dimensional convolutional neural network (1D-CNN), which attains an average accuracy of 0.89, and the deep belief network (DBN), which achieves 0.8333 despite its unsupervised pretraining on MPE features. Furthermore, MS-CS -ELM surpasses other population-based optimization techniques: Particle Swarm Optimization-ELM yields 0.88 on average, while Grey Wolf Optimizer-ELM records 0.8867. These comparisons underscore that the proposed multi-strategy enhanced Cuckoo Search not only improves ELM's parameter optimization but also delivers a more robust and accurate diagnostic model than both traditional deep learning classifiers and alternative swarm-intelligence-based ELM optimizations under identical experimental conditions.

6 Conclusions and Prospect

In this work, the challenges of strong noise interference, non-stationarity and multi-fault feature extraction in practical rolling bearing vibration signals, as well as the instability and suboptimal convergence of ELM due to random initialization, have been systematically addressed by integrating SSD based signal reconstruction with MPE feature extraction and a MS-CS optimization of ELM parameters. SSD effectively suppresses noise and preserves fault-related components by selecting singular components whose cumulative energy exceeds 90%, while MPE captures the dynamic complexity of the reconstructed signals across multiple time scales. The subsequent MS-CS algorithm, incorporating Cauchy mutation and adaptive Levy flight, further refines ELM input weights and hidden biases. Comprehensive validation on the Xi'an Jiaotong University bearing dataset demonstrates that MS-CS -ELM achieves an average diagnostic accuracy of 0.96, significantly outperforming traditional ELM, CS-ELM, 1D-CNN, DBN, PSO-ELM and GWO-ELM, thereby confirming the synergistic benefits of the proposed signal-level and algorithmic-level enhancements.

Looking ahead, the SSD, MPE and MS-CS -ELM framework will be extended to an online, real-time monitoring system where incremental learning and adaptive thresholding can facilitate rapid adaptation to evolving operational conditions and emerging compound fault modes. In addition, to address data imbalance and long-term degradation in field applications, future research will explore the integration of transfer learning and deep domain adaptation techniques to further enhance model generalization and robustness in complex industrial environments.

Author Contributions: Conceptualization, C.T. and Y.R.; methodology, C.T.; software, Y.R.; validation, C.T., Y.R. and T.O.; formal analysis, C.T.; investigation, C.T.; resources, Y.R.; data curation, Y.R.; writing—original draft preparation, C.T.; writing—review and editing, T.O.; visualization, C.T.; supervision, T.O.; project administration, T.O.; funding acquisition, Y.R. All authors have read and agreed to the published version of the manuscript.

 doi:10.20944/preprints202510.1308.v1

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available in the article and its supplementary materials.

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

References

1. Peng B, Bi Y, Xue B, et al. A survey on fault diagnosis of rolling bearings[J]. Algorithms, 2022, 15(10): 347.
2. Peng H, Zhang H, Fan Y, et al. A review of research on wind turbine bearings' failure analysis and fault diagnosis[J]. Lubricants, 2022, 11(1): 14.
3. Keshun Y, Puzhou W, Peng H, et al. A sound-vibration physical-information fusion constraint-guided deep learning method for rolling bearing fault diagnosis[J]. Reliability Engineering & System Safety, 2025, 253: 110556.
4. Peng D, Yazdanianasr M, Mauricio A, et al. Physics-driven cross domain digital twin framework for bearing fault diagnosis in non-stationary conditions[J]. Mechanical Systems and Signal Processing, 2025, 228: 112266.
5. Samavatian M, Behzad M, Mehdigholi H. Nonlinear modeling for bearing fault diagnosis in non-stationary operating conditions[J]. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2024, 46(5): 323.
6. Li Y, Zhou J, Li H, et al. A fast and adaptive empirical mode decomposition method and its application in rolling bearing fault diagnosis[J]. IEEE Sensors Journal, 2022, 23(1): 567-576.
7. Gu J, Peng Y. An improved complementary ensemble empirical mode decomposition method and its application in rolling bearing fault diagnosis[J]. Digital Signal Processing, 2021, 113: 103050.
8. Habbouche H, Amirat Y, Benkedjouh T, et al. Bearing fault event-triggered diagnosis using a variational mode decomposition-based machine learning approach[J]. IEEE Transactions on Energy Conversion, 2021, 37(1): 466-474.
9. Zhang K, Deng Y, Chen P, et al. Quaternion empirical wavelet transform and its applications in rolling bearing fault diagnosis[J]. Measurement, 2022, 195: 111179.
10. Wang S, Lian G, Cheng C, et al. A novel method of rolling bearings fault diagnosis based on singular spectrum decomposition and optimized stochastic configuration network[J]. Neurocomputing, 2024, 574: 127278.
11. He D, Liu C, Jin Z, et al. Fault diagnosis of flywheel bearing based on parameter optimization variational mode decomposition energy entropy and deep learning[J]. Energy, 2022, 239: 122108.
12. Yi C, Wang H, Ran L, et al. Power spectral density-guided variational mode decomposition for the compound fault diagnosis of rolling bearings[J]. Measurement, 2022, 199: 111494.
13. Minhas A S, Singh S. A new bearing fault diagnosis approach combining sensitive statistical features with improved multiscale permutation entropy method[J]. Knowledge-based systems, 2021, 218: 106883.
14. Suthar V, Vakharia V, Patel V K, et al. Detection of compound faults in ball bearings using multiscale-SinGAN, heat transfer search optimization, and extreme learning machine[J]. Machines, 2022, 11(1): 29.
15. Wei H, Zhang Q, Shang M, et al. Extreme learning Machine-based classifier for fault diagnosis of rotating Machinery using a residual network and continuous wavelet transform[J]. Measurement, 2021, 183: 109864.
16. Nezamivand Chegini S, Amini P, Ahmadi B, et al. Intelligent bearing fault diagnosis using swarm decomposition method and new hybrid particle swarm optimization algorithm[J]. Soft Computing, 2022, 26(3): 1475-1497.
17. Xiao M, Liao Y, Bartos P, et al. Fault diagnosis of rolling bearing based on back propagation neural network optimized by cuckoo search algorithm[J]. Multimedia tools and applications, 2022, 81(2): 1567-1587.

 doi:10.20944/preprints202510.1308.v1

18. Zhu R, Wang M, Xu S, et al. Fault diagnosis of rolling bearing based on singular spectrum analysis and wide convolution kernel neural network[J]. Journal of Low Frequency Noise, Vibration and Active Control, 2022, 41(4): 1307-1321.
19. Wang Y, Wang H, Bai R, et al. Enhanced Rolling Bearing Fault Diagnosis Using Multimodal Deep Learning and Singular Spectrum Analysis[J]. Applied Sciences, 2025, 15(9): 4828.
20. Jiang Q, Dai J, Shao F, et al. Bearing early fault diagnosis based on an improved multiscale permutation entropy and SVM[J]. Shock and Vibration, 2022, 2022(1): 2227148.
21. Wang L, Ai Q, Yan H, et al. Advanced Bearing Fault Diagnosis Using Cuckoo Optimization and KAN Algorithms[C]//2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS). IEEE, 2024: 98-102.
22. Xu Y, Chen N, Shen X, et al. Proposal and experimental case study on building ventilating fan fault diagnosis based on cuckoo search algorithm optimized extreme learning machine[J]. Sustainable Energy Technologies and Assessments, 2021, 45: 100975.