algorithms

MDPI

Article

Intelligent Fault Diagnosis for Rotating Mechanical Systems: An Improved Multiscale Fuzzy Entropy and Support Vector Machine Algorithm

Yuxin Pan [1], Yinsheng Chen [2], Xihong Fei [1], Kang Wang [1], Tian Fang [1] and Jing Wang [1,*]

[1] School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan 243002, China; yxpan0219@126.com (Y.P.); fxhong@mail.ustc.edu.cn (X.F.); ahutwk@ahut.edu.cn (K.W.); fangtian@ahut.edu.cn (T.F.)

[2] Huangshan Special Equipment Supervision and Inspection Center, Huangshan 245061, China; hstjcc@163.com

* Correspondence: jingwang08@126.com

Abstract: Rotating mechanical systems (RMSs) are widely applied in various industrial fields. Intelligent fault diagnosis technology plays a significant role in improving the reliability and safety of industrial equipment. A new algorithm based on improved multiscale fuzzy entropy and support vector machine (IMFE-SVM) is proposed for the automatic diagnosis of various fault types in elevator rotating mechanical systems. First, the empirical mode decomposition (EMD) method is utilized to construct a decomposition model of the vibration data for the extraction of relevant parameters related to the fault feature. Secondly, the improved multiscale fuzzy entropy (IMFE) model is employed, where the scale factor of the multiscale fuzzy entropy (MFE) is extended to multiple subsequences to resolve the problem of insufficient coarse granularity in the traditional MFE. Subsequently, linear discriminant analysis (LDA) is applied to reduce the dimensionality of the extracted features in order to overcome the problem of feature redundancy. Finally, a support vector machine (SVM) model is utilized to construct the optimal hyperplane for the diagnosis of fault types. Experimental results indicate that the proposed method outperforms other state-of-the-art methods in the fault diagnosis of elevator systems.

Keywords: rotating mechanical systems (RMS); intelligent fault diagnosis; empirical mode decomposition (EMD); multiscale fuzzy entropy (MFE); support vector machine (SVM)

Citation: Pan, Y.; Chen, Y.; Fei, X.; Wang, K.; Fang, T.; Wang, J. Intelligent Fault Diagnosis for Rotating Mechanical Systems: An Improved Multiscale Fuzzy Entropy and Support Vector Machine Algorithm. *Algorithms* **2024**, *17*, 588. https://doi.org/10.3390/a17120588

Academic Editors: Sumika Chauhan, Govind Vashishtha and Rajesh Kumar

Received: 26 November 2024
Revised: 11 December 2024
Accepted: 16 December 2024
Published: 20 December 2024

1. Introduction

Rotating machinery plays an increasingly important role in industries such as healthcare, manufacturing, and transportation [1,2]. Under long-term operating conditions, critical components of rotating machinery, such as bearings and gears, are prone to various types of faults. These faults not only lead to production stagnation but may also trigger safety incidents. Consequently, the introduction of intelligent diagnostic technologies is crucial for accurately diagnosing faults in rotating machinery, thereby ensuring equipment safety and enhancing production efficiency [3–5].

Traditional fault diagnosis methods for rotating machinery typically rely on experiential knowledge and physical mechanisms [6]. However, in changeable working conditions, such methods often struggle to meet the demands for real-time and efficient automated diagnosis due to their reliance on expert experience, insufficient timeliness, and high maintenance costs. Li et al.[7] highlighted the limitations of expert experience in complex fault scenarios, which can result in the overlooking of subtle faults in equipment. Badihi et al. [8] proposed that fault diagnosis based on physical mechanisms often fails under dynamic and nonlinear conditions, leading to suboptimal diagnostic outcomes. Furthermore, traditional fault diagnosis methods impose high requirements on the quality of input data, with noise and anomalous data significantly impacting diagnostic results. Currently,

with advancements in equipment diagnostic sensor technologies, researchers can set up fault conditions in laboratory settings based on actual production scenarios to obtain more comprehensive data sources. This development has allowed data-driven intelligent fault diagnosis technologies to gradually emerge in the field of fault diagnosis [9], particularly methods based on entropy for feature extraction.

Entropy is a tool for measuring uncertainty and complexity, and it is widely applied in the fault diagnosis of rotating machinery [10,11]. By analyzing the complexity of vibration signals, fault features can be extracted. Guan et al. [12] proposed a fault diagnosis method for rotating machinery based on sample entropy (SE). Through the analysis of fault signals using SE, effective differentiation between normal and fault states can be achieved. Results indicate that this method demonstrates high accuracy in identifying early faults. Ma et al. [13] utilized improved composite multiscale approximate entropy to analyze fault signals from rolling bearings. It was observed that this improved entropy exhibits high sensitivity to fault types, enabling the detection of minor changes in vibration signals and significantly improving fault classification accuracy. Chang et al. [14] introduced fuzzy entropy (FE) as a feature extraction method, combining it with support vector machine (SVM) for fault diagnosis, which yielded favorable results. Research has shown that FE possesses strong robustness in handling noisy signals. Ge et al. [15] employed multiscale entropy methods to analyze fault signals, and the results revealed that this method significantly improves fault detection sensitivity. Li et al. [16] integrated multiscale permutation entropy (MPE) with deep learning, utilizing MPE to extract signal features and performing fault classification through a convolutional neural network (CNN). This approach exhibited high diagnosis accuracy when addressing complex faults, particularly outperforming traditional diagnostic methods in noisy environments. However, challenges still exist in fault diagnosis techniques based on entropy algorithms:

1. Variable entropy algorithms exhibit the disadvantage of difficult hyperparameter adjustment, which undermines the generalization ability and adaptability of such algorithms [17]. Even when the tuning time for the optimization of hyperparameters is reduced, the time required for the model to implement responsive tuning strategies in response to different vibration signals remains unavoidable.
2. In order to conduct research across multiple time scales, commonly used coarse-graining techniques may result in the loss of local variations within the signals. When the scale factor is large, the original data length is significantly reduced. This can lead to issues such as insufficient smoothness in feature generation, bias in feature extraction, and the neglect of certain signal features.
3. The widely used multiscale entropy calculates the average entropy values of coarse-grained sequences at different scale factors, which can lead to overfitting [18]. Due to the presence of many similar patterns within vibration signals, averaging the entropy values of various subsequences may introduce redundant information. This causes the model to become dependent on these redundant features, hindering its ability to accurately capture subtle changes and trends in the signals, thereby affecting the algorithm's generalization capability.

To address the aforementioned challenges, the main contributions of this article are outlined as follows:

1. FE is extended to multiple time scales through the method of shifted coarse-graining to handle data uncertainty and fuzziness. The advantage of this shifted coarse-graining approach is that the vibration signals are utilized multiple times. Consequently, each window can capture more subtle changes and trends, thereby reducing abrupt variations between different windows and accurately reflecting the local features of the original signal.
2. To avoid overfitting caused by redundant features, a non-parametric feature extraction method named improved multiscale fuzzy entropy (IMFE) is proposed. This method utilizes the nonlinear dynamic model of IMFE to quantify temporal complexity, cir-

cumventing the issue of hyperparameter tuning time. Comparative experimental results demonstrate that IMFE exhibits superior robustness.

3. An intelligent fault diagnosis method for rotating machinery is proposed, consisting of four modules: data decomposition, feature extraction, feature dimensionality reduction, and fault diagnosis. First, the empirical mode decomposition (EMD) method is used to perform three-layer decomposition on the data, yielding three intrinsic mode functions. Then, the signal features extracted using IMFE are input into linear discriminant analysis (LDA) for dimensionality reduction. Finally, the support vector machine (SVM) algorithm is employed as the classifier to achieve fault diagnosis, with *K*-fold cross-validation methods utilized to evaluate the diagnostic model's performance [19].
4. The proposed method is validated using open-source data from rotating mechanical rolling bearings. The results indicate that the proposed method outperforms other comparative methods.

The remainder of this article is organized as follows. A brief introduction to the foundational theories is provided in Section 2. The proposed method is presented in Section 3. Two case studies are discussed in Section 4, in which self-constructed signal data and the Case Western Reserve University (CWRU) bearing dataset are utilized for validation. The conclusions and future work are summarized in Section 5.

2. Multiscale Entropy Theory

Multiscale entropy (ME) is a method utilized for analyzing the complexity of time-series data [20]. The complexity of signals is measured at different time scales, providing more comprehensive information compared to traditional single-scale methods. The ME theory was initially proposed by Costa et al. in 2002 [21], aimed at addressing the limitations associated with measuring complexity using a single scale. The computational process involves the following steps: (1) signal data are constructed at different time scales. Initially, a series of new signals is generated by averaging the samples of the original signal inside windows. These generated signal series represent different time scales. (2) The sample entropy is calculated at each time scale for each downsampled signal to obtain entropy values at various time scales. A smaller sample entropy value indicates stronger regularity in the signal and better diagnostic capability, while a larger value suggests greater randomness and poorer diagnostic capability in the signal. The specific computation steps are outlined as follows:

1. A downsampled signal of the vibration data is constructed. Given a signal of $X = \{x_1, x_2, \ldots, x_N\}$, the downsampled sequence ($X^{(\tau)}$) at a time scale of τ is constructed as follows:
$$X^{(\tau)} = \{y_1, y_2, \ldots, y_{N/\tau}\} \tag{1}$$
where $y_j = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i$ and τ is the size of the time window.
2. The constructed signal is used to calculate the sample entropy. For each downsampled signal ($X^{(\tau)}$), the sample entropy is calculated using two parameters: m (the embedding dimension) and r (the similarity tolerance). A vector ($X^{(\tau)}$) of length m of the following sequence is constructed: $\mathbf{u}_i = \{y_i, y_{i+1}, \ldots, y_{i+m-1}\}, \quad i = 1, 2, \ldots, N/\tau - m$. Then, the distance between the elements of the vector is determined ($d(\mathbf{u}_i, \mathbf{u}_j) = \max_{k=0,\ldots,m-1} |y_{i+k} - y_{j+k}|$). The similarity ratio ($C^m(r)$) is defined, which is calculated as follows:
$$C^m(r) = \frac{1}{N/\tau - m} \sum_{j=1}^{N/\tau - m} \theta(r - d(\mathbf{u}_i, \mathbf{u}_j)) \tag{2}$$
where $\theta(x)$ is the indicator function. The value is 1 when x exceeds 0 and 0 otherwise.

The approximation ratio ($B^m(r)$) is calculated, and the likelihood ratio is the ratio of the approximate number to the total number. It is calculated as follows:

$$B^m(r) = \frac{1}{N/\tau - m} \sum_{i=1}^{N/\tau - m} C_i^m(r) \tag{3}$$

Then, the dimension is increased by 1 to become $m + 1$, and the above steps are repeated to obtain $B^{m+1}(r)$. It is calculated as follows:

$$B^{m+1}(r) = \frac{1}{N/\tau - m - 1} \sum_{i=1}^{N/\tau - m - 1} C_i^{m+1}(r) \tag{4}$$

Finally, the sample entropy is defined as follows:

$$\text{SampEn}(m, r, N) = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right) \tag{5}$$

3. Finally, the ME is calculated. For each value of τ, the corresponding SampEn is calculated to obtain the ME. It is calculated as follows:

$$\text{MSE} = \{\text{SampEn}(X^{(1)}), \text{SampEn}(X^{(2)}), \ldots, \text{SampEn}(X^{(\tau_N)})\} \tag{6}$$

3. Proposed Method

3.1. EMD Construction

The EMD, due to its adaptive nature, is able to effectively decompose non-stationary and nonlinear signals. Unlike methods that rely on pre-defined basis functions, EMD performs stepwise decomposition based on the extreme points of the local signal, allowing it to better capture the intrinsic modes within the signal. Compared to wavelet transform and VMD, EMD offers greater flexibility and adaptability, making it particularly suitable for complex dynamic vibration signals. The basis of EMD lies in adaptively decomposing nonlinear and non-stationary signals in vibration data into several intrinsic mode functions (IMFs) [22,23]. Each IMF reflects the dynamic characteristics of the signal at different time scales. For the selection of IMFs, a criterion based on the energy contribution of the decomposed modes was employed in this study. Specifically, modes with significant energy contributions were retained, while those with minimal energy contributions were discarded. Additionally, when processing the residual components after decomposition, optimization was performed based on the frequency characteristics of the residuals. Typically, the low-frequency component is considered noise, and it was filtered out to improve the signal quality and enhance the stability of the model.

This method emphasizes time-frequency analysis by extracting local features through an iterative process, effectively capturing instantaneous frequency and amplitude variations within the vibration signals. EMD is employed as a data processing method for vibration signals due to its ability to decompose based on the inherent characteristics of the signal, in contrast to other decomposition methods. It effectively extracts dynamic features across various time scales and possesses time-frequency analysis capabilities, allowing for a more comprehensive characterization of the signal. Specifically, the EMD method can decompose a non-stationary signal into a trend component and several meaningful IMFs with instantaneous frequencies [24,25].

The processing of vibration signal data is conducted based on three assumptions: (1) The signal must have at least two extreme points, including one maximum and one minimum. (2) The time scale of signal is defined by the time between the two extreme points. (3) If the signal lacks extreme points but contains inflection points, extreme points may be obtained by differentiating the data once or multiple times, followed by integration to achieve the decomposition result. The EMD algorithm, as illustrated in Figure 1, sepa-

rates different frequency components from the original signal, thereby enhancing feature extraction. The main steps are detailed as follows:

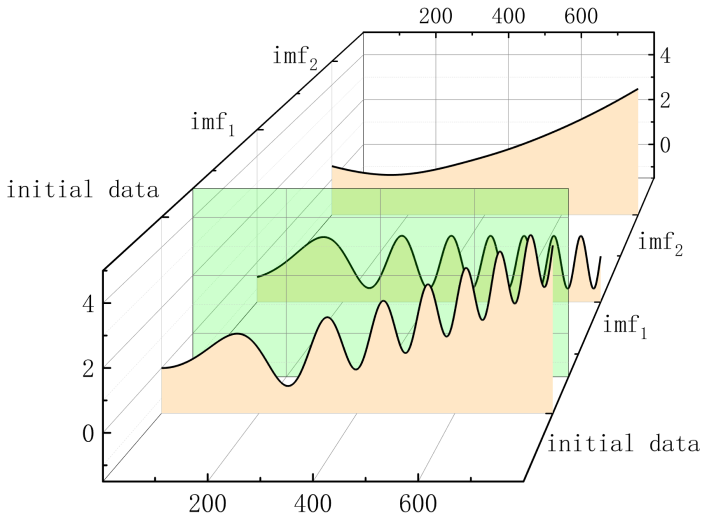

Figure 1. EMD of the signal.

Step 1: For the input signal ($x(t)$), the maximum and minimum points of the original signal are fitted using spline interpolation to construct the upper and lower envelope curves. The mean envelope ($x_0(t)$) is then calculated, and the original signal is subtracted from this mean to obtain a new signal ($h_1(t)$). This process can be expressed as follows:

$$h_1(t) = x(t) - x_0(t) \tag{7}$$

Step 2: The new signal ($h_1(t)$) is evaluated to determine whether it satisfies the conditions of an IMF. If the conditions are satisfied, $h_1(t)$ is designated as the first IMF component. If $h_1(t)$ does not meet the conditions, the process continues in a loop using *Step 1* until $h_1(t)$ fulfills the feature conditions of an IMF.

Step 3: The first IMF (imf_1) is separated from the original vibration signal to obtain ($r_1(t)$). The new signal can be expressed as $r_1(t) = x(t) - imf_1$.

Step 4: The process is repeated until the *K*-th IMF is obtained. The final result of the EMD can be expressed as follows:

$$x(t) = \sum_{j=1}^{n} imf_j + r_n(t) \tag{8}$$

where $r_n(t)$ denotes the residue, which represents the central tendency of the vibration signal.

IMFE Design

By evaluating the fuzziness and uncertainty of vibration signals, fuzzy entropy effectively extracts fault-related features. Compared to traditional fuzzy entropy algorithms, the multiscale fuzzy entropy (MFE) algorithm extracts multi-level signal features by calculating entropy values across different time scales, allowing potential fault patterns to be identified [26,27]. However, as the scale factor of entropy increases, the length of the coarse-grained time series decreases significantly. This sudden reduction in sequence length causes sharp fluctuations in entropy values, complicating the extraction of key features.

To address this issue, a novel time segmentation method is proposed in this article to replace traditional segmentation approaches, aiming to address the problem of high volatility in entropy values at large scale factors. Figure 2 illustrates the improved coarse-graining approach, unlike traditional coarse-graining that produces only a single subsequence of length 7, which preserves more local information essential for feature

construction. The vibration signal sequence is denoted as $x(t)$, and sequences $\mu_1(t)$, $\mu_2(t)$, and $\mu_3(t)$ are defined based on a sliding window approach. Specifically, $\mu_1(t)$ is obtained by averaging the first three elements of $x(t)$, then shifting the sliding window one step to the right and averaging the subsequent three elements. Similarly, $\mu_2(t)$ is derived by averaging the first three elements, followed by a two-step rightward shift of the window and averaging of the next three elements. $\mu_3(t)$ is calculated by averaging the first three elements, shifting the window three steps to the right, and averaging the subsequent three elements. Assuming $x(t) = [x_1, x_2, x_3, \ldots, x_n]$, the following results are obtained: $\mu_1(t) = \left[\frac{x_1+x_2+x_3}{3}, \frac{x_2+x_3+x_4}{3}, \frac{x_3+x_4+x_5}{3}, \ldots\right]$, $\mu_2(t) = \left[\frac{x_1+x_2+x_3}{3}, \frac{x_3+x_4+x_5}{3}, \frac{x_5+x_6+x_7}{3}, \ldots\right]$, and $\mu_2(t) = \left[\frac{x_1+x_2+x_3}{3}, \frac{x_4+x_5+x_6}{3}, \frac{x_7+x_8+x_9}{3}, \ldots\right]$. For instance, when the sequence length is 21 and the scale factor is 3, the proposed offset coarse-graining method generates three subsequences with lengths of 19, 10, and 7, respectively, thereby reducing sudden entropy fluctuations and computational bias. This approach, combined with fuzzy entropy, forms an IMFE algorithm, with the algorithmic design shown in Figure 3. The main steps are described as follows:

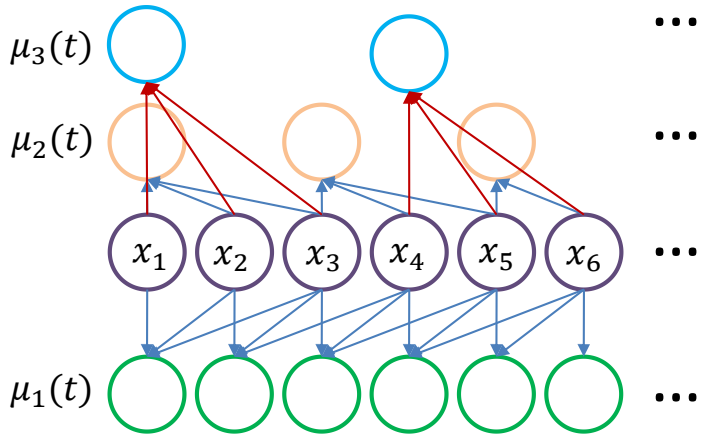

Figure 2. An improved coarse-graining method for a scale factor of 3.

Step 1: The signal samples are input from the equipment, and set the coarse-graining scale factor sequence is set to $scales = (scale_1, scale_2, \ldots, scale_N)$, where $1 < scale_i < N$.

Step 2: The offset factor is set to $k = 1$. When the scale factor is $scale_1$, it increments from 1 until it equals $scale_1$. At each distinct k, coarse-graining is performed to obtain a signal, the fuzzy entropy of which is calculated. This process is repeated for each subsequent scale factor.

Step 3: The fuzzy entropy is calculated [28]. Given an N-dimensional signal ($[u(1), u(2), \ldots, u(N)]$), the phase space dimension ($m(m \leq N-2)$) and similarity tolerance (r) are defined. Then, the phase space is reconstructed as

$$X(i) = [u(i), u(i+1), \ldots, u(i+m-1)] - u_0(i), i = 1, 2, \ldots, N-m+1 \quad (9)$$

where $u_0(i) = \frac{1}{m}\sum_{j=0}^{m-1} u(i+j)$. The fuzzy membership function is calculated by

$$A_{ij}^m = \exp[-\ln(2)\left(d_{ij}^m / r\right)^2] \quad (10)$$

where $j = 1, 2, \ldots, N-m+1$ and $j \neq i$. The maximum absolute distance between window vectors $X(i)$ and $X(j)$ is obtained as

$$d_{ij}^m = d[X(i), X(j)] = \max_{p=1,2,\ldots,m}(|u(i+p-1) - u_0(i)| - |u(j+p-1) - u_0(j)|) \quad (11)$$

Taking the average for each i yields $C_i^m(r) = \frac{1}{N-m}\sum_{j=1,j\neq i}^{N-m+1} A_{ij}^m$, and $\Phi^m(r)$ is defined by

$$\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} C_i^m(r) \tag{12}$$

The fuzzy entropy of the original signal is denoted by

$$FuzzyEn(m,r) = \lim_{N\to\infty}[\ln\Phi^m(r) - \ln\Phi^{m+1}(r)] \tag{13}$$

In the case of a finite dataset, the fuzzy entropy can be estimated as $FuzzyEn(m,r,N) = \ln\Phi^m(r) - \ln\Phi^{m+1}(r)$. The combination of all fuzzy entropy is the improved multiscale fuzzy entropy, and the output entropy matrix is $entropy = [\ldots]_{sum(scale_1, scale_{2,\ldots}, scale_N)}$.

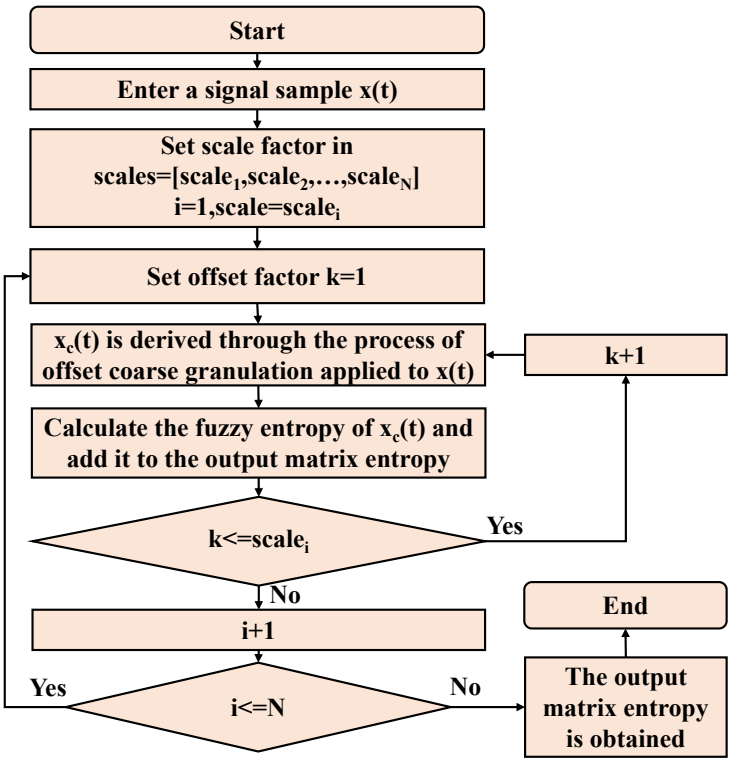

Figure 3. The design of the IMFE algorithm.

3.2. LDA Applied

LDA is a supervised learning algorithm. By maximizing between-class scatter and minimizing within-class scatter, LDA effectively reduces high-dimensional data to lower dimensions, simplifying the feature space and improving computational efficiency. In binary classification problems, it is also referred to as Fisher discriminant analysis (FDA) [29]. However, there are some distinctions between LDA and FDA. LDA assumes that the data of each class follows a Gaussian distribution, with identical covariance matrices that are full-rank [30]. Additionally, when handling fault modes with clear class distinctions, LDA can provide high classification accuracy. The main steps of LDA are outlined as follows:

Step 1: The signal data are input ($X = [(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)]$), where any sample ($x_i$) is an n-dimensional vector and y_i represents the fault-class label.

Step 2: The mean vector (μ_i) for each class in the dataset and the overall mean vector (μ) are calculated. The within-class scatter matrix (S_ω) and the total scatter matrix (S_t) are computed, and the between-class scatter matrix ($S_b = S_t - S_\omega$) is computed.

Step 3: Eigenvalue decomposition is performed on matrix $S_\omega^{-1}S_b$, and the eigenvalues are sorted in descending order. The feature vectors corresponding to the top k eigenval-

ues are selected, and the n-dimensional samples are reduced to k dimensions using the feature matrix ($x' = \omega^T x$).

3.3. SVM Adopted

SVM is a commonly used machine learning algorithm capable of handling high-dimensional vibration signal data [31]. By constructing an optimal hyperplane for fault classification, SVM achieves high diagnosis accuracy, performing particularly well when the sample size is limited. Furthermore, SVM exhibits strong robustness, being insensitive to noise and outliers, which enhances its reliability in practical applications. Its primary concept involves finding a hyperplane in the feature space that separates data from different fault categories while maximizing the distance between the hyperplane and the closest data points from each fault class [32]. The main steps of SVM are outlined as follows:

Step 1: The vibration signal data are input ($X = [(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)]$), where any sample ($x_i$) is an n-dimensional vector, with $x_i \in \mathbb{R}^n$. Let y_i represent the fault class label, where $i = 1, 2, \ldots, N$.

Step 2: A penalty parameter ($C > 0$) is selected, and a convex quadratic programming minimization problem is constructed, denoted as $\min \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i$, subject to the constraints of $\sum_{i=1}^{N} \alpha_i y_i = 0$ and $0 < \alpha_i < C, i = 1, 2, \ldots, N$. This problem is solved to obtain the optimal solution ($\alpha^* = \left(\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*,\right)^T$).

Step 3: $\omega^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$ is computed; then, a component (α^*) that satisfies the condition of $0 < \alpha_j < C$ is selected, and $b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$ is calculated.

Step 4: The separating hyperplane defined by $\omega^* \cdot x + b^* = 0$ is obtained, and the classification decision function is set as $f(x) = sign(\omega^* \cdot x + b^*)$.

3.4. Overall Framework of the Proposed Method

The overall framework of the proposed method is illustrated in Figure 4. Initially, the vibration signal data are decomposed using EMD, and the first three IMFs are extracted. These IMFs capture the vibration information within specific frequency ranges while preserving the nonlinear and non-stationary characteristics of the signal. Subsequently, IMFE is employed to extract features from each of the three IMFs, forming an initial feature vector. LDA is then applied for dimensionality reduction, followed by diagnosis using an SVM. Finally, the superiority of the proposed method is validated through K-fold cross-validation. The following are the main procedures.

Step 1: Vibration signal data are collected from the rotating machinery system. First, the signals are decomposed using the EMD method to extract fault features. This process decomposes complex nonlinear and non-stationary signals into a series of IMFs, with each IMF representing a specific frequency component in the signal, thereby obtaining IMFs that reflect the system's condition.

Step 2: The IMFs obtained from EMD are applied to the IMFE algorithm. IMFE is a nonlinear dynamics technique, extending the fuzzy entropy approach to multiple time scales. Through the IMFE algorithm, finer vibration data features can be captured across different time scales, addressing the limitations of traditional MFE in handling vibration data with insufficient granularity.

Step 3: After multiscale feature extraction, LDA is applied to reduce the dimensionality of the features. This produces a set of features with enhanced classification performance, reducing feature redundancy and improving diagnostic efficiency.

Step 4: Using the extracted and dimensionally reduced features, an SVM model is trained. This model is designed to accurately recognize various fault types in the elevator rotating machinery system. Finally, K-fold cross-validation is applied to assess the effectiveness of the proposed algorithm.

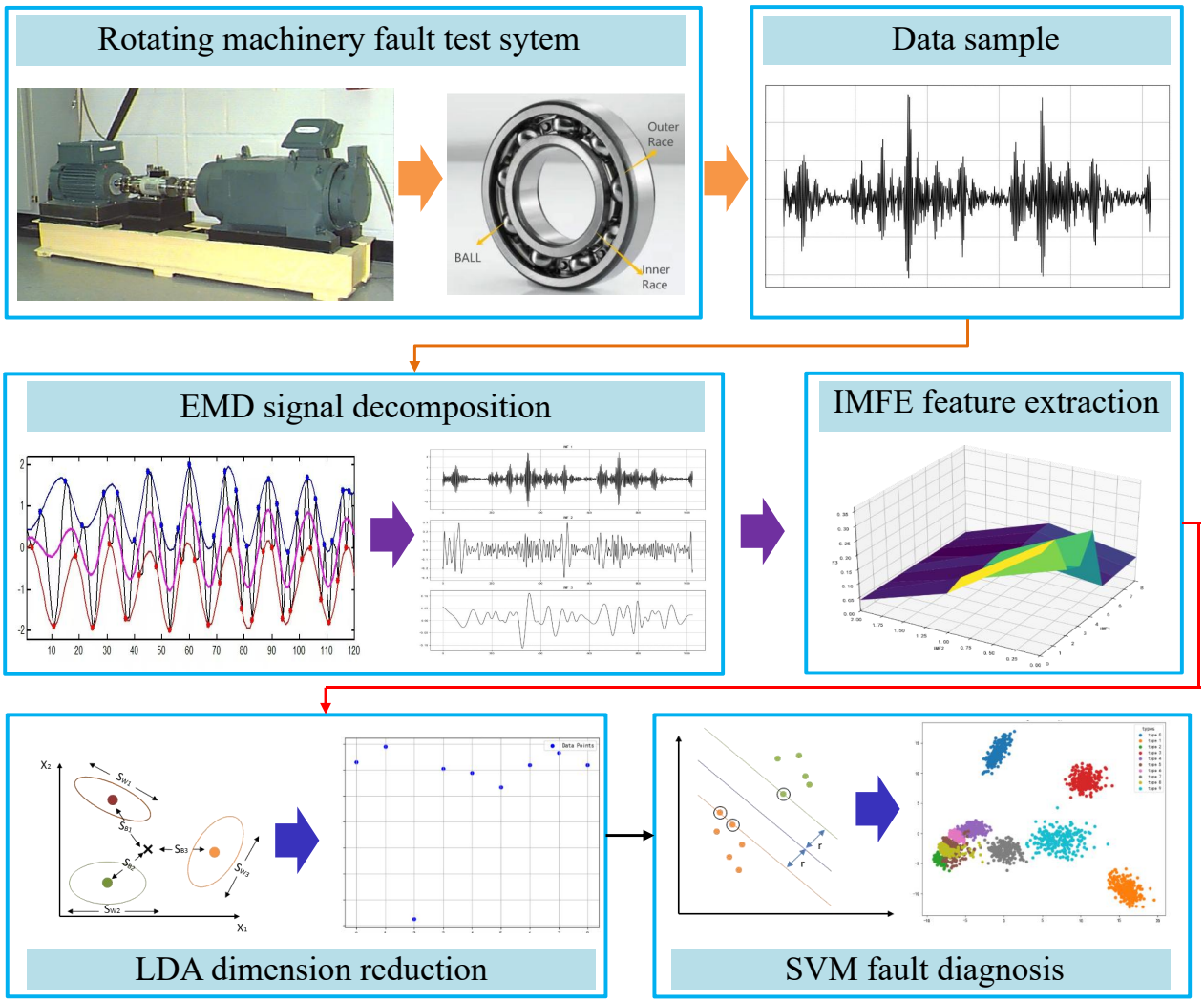

Figure 4. Overall framework of the proposed method.

4. Experimental Validation

4.1. CASE 1: Performance Verification of IMFE Algorithm

4.1.1. The Selection of a Vibration Signal

The length of a vibration signal impacts the performance of feature extraction algorithms. A signal that is too short can lead to significant feature fluctuations, while an overly long signal results in substantial time costs [33]. Therefore, this section examines the entropy value distributions of fuzzy entropy (FE), MFE, and IMFE across vibrations signal of varying lengths. Four specific signals are used in this case. A segment of white noise with a duration of 1 s is generated, with x_1 sampled at 10,000 Hz. This white-noise signal is produced by drawing random samples from a standard normal distribution characterized by a mean of 0 and a standard deviation of 1. Uniformly distributed in the frequency domain, this white noise is suitable for a range of signal processing and system testing applications. $\mathbf{x}_2(t) = 5(1+\cos(4\pi t))\cos(20\pi t) + 10(1+\cos(4\pi t))\cos(40\pi t) + 15(1+\cos(4\pi t))\cos(60\pi t) + n$, where n is Gaussian white noise with a value of -20 dB and x_2 exhibits periodic characteristics resembling a sinusoidal waveform. x_3 is generated by drawing random samples from a uniform distribution, with values ranging between -1 and 1. This uniformly distributed white noise maintains equal intensity across all frequency components in the frequency domain, making it equally suitable for signal processing, system testing, and noise analysis. x_4 is a sine wave signal. Inspired by [34], entropy distribution within the range of 1024 to 6656 signal lengths is analyzed, with a step size of 512, generating 12 different subsequences. A correlation metric is then introduced to comprehensively assess the algorithm's

adaptability to various vibration signal lengths. The specific forms of the four signals are shown in Figure 5. The primary differences among the four signals lie in amplitude, frequency, and signal trend.

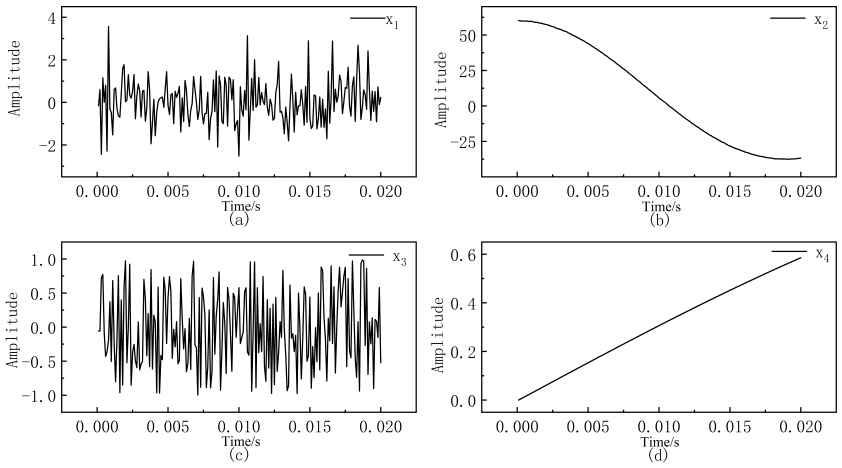

Figure 5. The specific forms of the four signals. (**a**) x_1; (**b**) x_2; (**c**) x_3; (**d**) x_4.

4.1.2. Performance Experiment

As illustrated in Figure 6, IMFE exhibits smaller entropy fluctuations across varying subsequence lengths (1024 to 6656), demonstrating greater stability. As the signal length increases, the FE values show a general upward trend with substantial fluctuations. For sequence 1, the FE value rises from 1.741 to 1.923, an increase of approximately 10.4%. In contrast, the IMFE for sequence 1 increases only from 1.240 to 1.293, a mere 4.3% increase, which is significantly smaller than the fluctuation range seen in FE, confirming that IMFE displays notably reduced variability across different lengths.

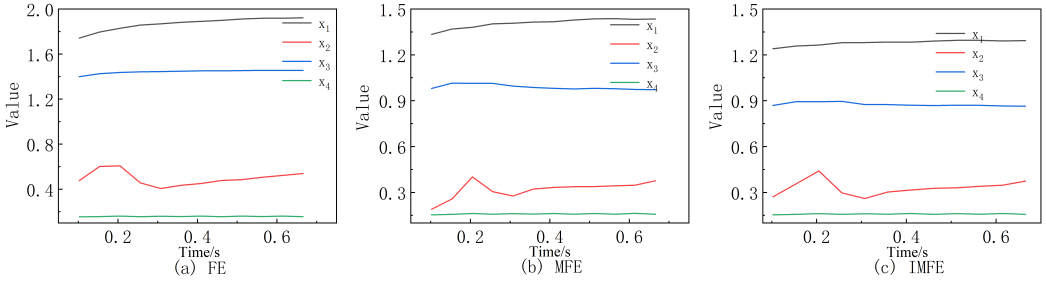

Figure 6. The results using FE, MFE, and IMFE methods for different lengths of four experimental signals.

Additionally, for sequence 2, MFE rises from 0.189 at a length of 1024 to 0.376 at 6656, an increase of 98.9%. However, the IMFE for the same sequence increases only from 0.271 to 0.374, an increase of 37.9%. This indicates that MFE fluctuates significantly with increasing signals length, while IMFE demonstrates lower sensitivity to scale variations, allowing it to capture sequence features with greater stability.

The results across all sequence data indicate that IMFE has a smaller fluctuation range, with an average increase kept within 5%. In contrast, FE and MFE exhibit increases of over 10% as length grows. Table 1 shows that IMFE demonstrates superior uncorrelated performance in sequences x_1 and x_3, while the values for x_2 and x_4 are also close to those of FE, showing that IMFE maintains significant robustness under varying signal lengths. This robustness makes it particularly advantageous in complex signal analysis. Tables 2 and 3 and Figure 7 present the standard deviation and variance results for IMFE, both of which are consistently low across the four signals. These results indicate that IMFE possesses higher stability and interference resistance across various signal lengths, making it more

reliable for reflecting sequence features and suitable for analyzing multiscale, variable-length signals.

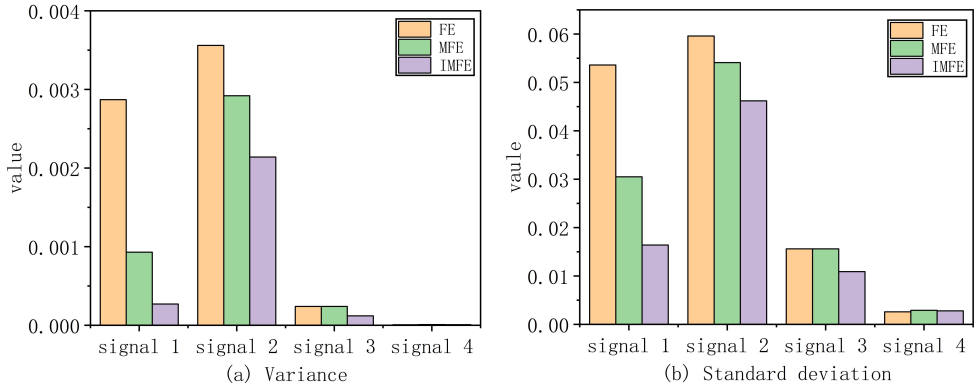

Figure 7. The variance and standard deviation results of entropy values for four signals at different lengths.

Table 1. Correlation coefficient between entropy and sequence length for four signals at different lengths.

Correlation Coefficient	FE	MPE	IMFE
x_1	0.9219	0.9091	0.8911
x_2	0.0709	0.6306	0.1901
x_3	0.8380	0.7203	0.6845
x_4	0.2189	0.3634	0.3069

Table 2. Standard deviation of entropy values for four signals at different lengths.

Standard Deviation	FE	MPE	IMFE
x_1	0.0536	0.0305	0.0164
x_2	0.0596	0.0541	0.0462
x_3	0.0156	0.0156	0.0109
x_4	0.0026	0.0029	0.0028

Table 3. Variance of entropy values of four signals at different lengths.

Variation	FE	MPE	IMFE
x_1	0.00287	0.00093	0.00027
x_2	0.00356	0.00292	0.00214
x_3	0.00024	0.00024	0.00012
x_4	0.0000069	0.0000087	0.0000077

The four signals were processed by combining EMD with IMFE to extract features from IMFs. Due to the fixed frequency of the sine wave function, EMD cannot effectively extract valid intrinsic functions from it. Therefore, only x_1, x_2, and x_3 were analyzed. The sample length was set to 1024, with a step size of 512. Then, 10 samples were taken from each signal, totaling 30 samples for the experiment. The clustering results are shown in Figure 8, where "feature 1" represents the IMFE of imf_1 after EMD and "feature 2" represents the IMFE of imf_2. As shown in Figure 8, IMFE exhibits effective feature extraction following EMD, enabling clear separation of samples from different signal classes.

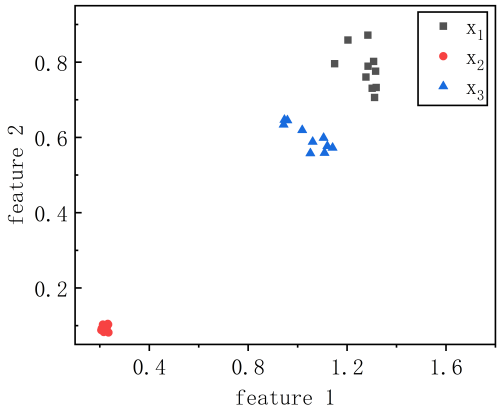

Figure 8. Visualization of clustering effects of different signal features based on EMD and IMFE.

4.2. CASE 2: Fault Diagnosis of a Roller Bearing Unit

4.2.1. Fault Data Description of Roller Bearing

The Case Western Reserve University (CWRU) bearing fault simulation platform consists of bearings, a motor, a torque sensor, an encoder, a power tester, and other electronic control equipment [35]. Its platform is shown in Figure 4. Single-point faults were introduced on the bearings using electric discharge machining, with fault sizes ranging from 0.007 inches to 0.040 inches in diameter. The fault types include inner ring faults, rolling element faults, and outer ring faults. Faulty bearings were mounted onto the test motor. Accelerometers were placed on the drive end, fan end, and base of the motor casing to record vibration data under four load conditions (from 0 to 3 horsepower) at motor speeds ranging from 1797 to 1720 RPM.

The vibration data consist of four datasets. The normal baseline data dataset contains baseline vibration signals from the base under normal operating conditions. The 12 *k* drive-end bearing fault data contain drive-end fault data collected at a 12 kHz sampling rate. The 12k fan-end bearing fault data comprise fan-end fault data collected at a 12 kHz sampling rate, and the 48 *k* drive-end bearing fault data comprise drive-end fault data collected at a 48 kHz sampling rate.

4.2.2. Comparative Analysis of Experimental Results

For this experiment, the dataset used is the 12 *k* drive-end bearing fault data collected at a 12 kHz sampling rate. Under conditions of a 2-horsepower motor load and a motor speed of 1750 rpm, a dataset comprising 10 classes was prepared, including 9 fault types (involving inner and outer race faults at different locations and rotor faults), as well as normal operational data. From the initial data, subsequences of a length of 1024 were extracted as samples, with a step size of 512 used to select subsequent samples. A total of 2330 of the first samples were used for the experiment.

Table 4 presents the ten working conditions of a rolling bearing. Among them, Class 1 represents data collected under normal operating conditions, while Classes 2 to 10 correspond to data recorded during motor faults. The faults are primarily located in three areas: the inner ring, outer ring, and rolling element. Each fault type is further divided based on defect diameter, including 0.07 inches, 0.14 inches, and 0.21 inches, yielding a total of nine fault categories. Figure 9 illustrates sample waveforms for each of the nine fault types: (a) inner ring fault with a 0.07-inch defect, (b) tolling element fault with a 0.07-inch defect, (c) outer ring fault with a 0.07-inch defect, (d) inner ring fault with a 0.14-inch defect, (e) rolling element fault with a 0.14-inch defect, (f) outer ring fault with a 0.14-inch defect, (g) inner ring fault with a 0.21-inch defect, (h) rolling element fault with a 0.21-inch defect, and (i) outer ring fault with a 0.21-inch defect. These fault categorizations allow for a comprehensive dataset to effectively assess the fault diagnosis model's performance across varying fault types.

Table 4. Ten working conditions of rolling bearings.

Number	Fault (Yes/No)	Fault Location	Fault Diameter (Inches)	Abbreviation
Class 1	No	/	/	Nor
Class 2	Yes	Inner ring	0.007	IR7
Class 3	Yes	Rolling element	0.007	RE7
Class 4	Yes	Outer ring	0.007	OR7
Class 5	Yes	Inner ring	0.014	IR14
Class 6	Yes	Rolling element	0.014	RE14
Class 7	Yes	Outer ring	0.014	OR14
Class 8	Yes	Inner ring	0.021	IR21
Class 9	Yes	Rolling element	0.021	RE21
Class 10	Yes	Outer ring	0.021	OR21

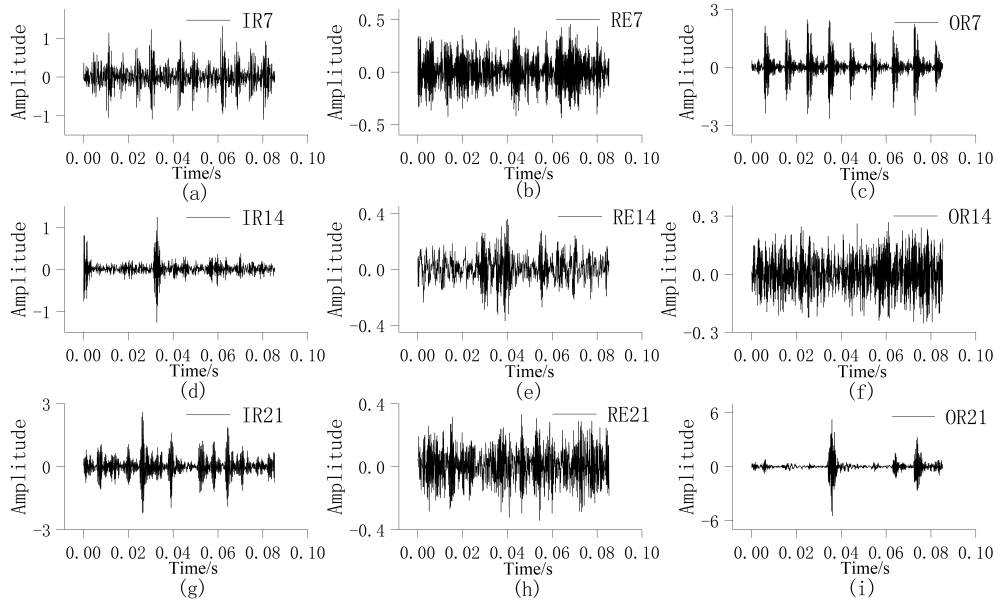

Figure 9. Time-domain waveforms of the nine types of working fault.

This section compares the diagnostic results of different entropy-based algorithms, including FE, MFE, FE combined with waveform factor, and the proposed method. The scale factors for MFE and IMFE are set to $scales = [2, 3, 4]$, with an embedding dimension of 2 and a time delay of 1. The embedding dimension (m) and similarity tolerance r parameters were optimized through multiple experiments to achieve the best performance of the proposed method, with $m = 2$ and $r = 0.2$. The features extracted using IMFE are reduced to three dimensions by LDA. Given that the maximum dimensionality reduction achievable by LDA corresponds to the total number of fault categories minus one, which is 9 in the case of 10 categories, the dimensionality was reduced to 9 in this article. This choice ensures that as much feature information as possible is retained while maintaining the ability to effectively distinguish between different categories. All methods are classified using an SVM classifier and validated through K-fold cross-validation, where K is set to 10 (i.e., the dataset is split into 10 folds, each containing 233 samples). In all comparative experiments, the parameters of the SVM were kept consistent, with the penalty parameter (C) set to 1.0. As shown in Figure 10, the proposed method achieves a consistently high classification accuracy across all folds. Table 5 shows a comparison of diagnostic results, showing the average classification accuracy over ten runs, as well as the highest and lowest accuracy achieved across these runs. The proposed IMFE method demonstrates superior performance compared to the baseline entropy methods, with a notable improvement in diagnostic accuracy across various scales and validation folds.

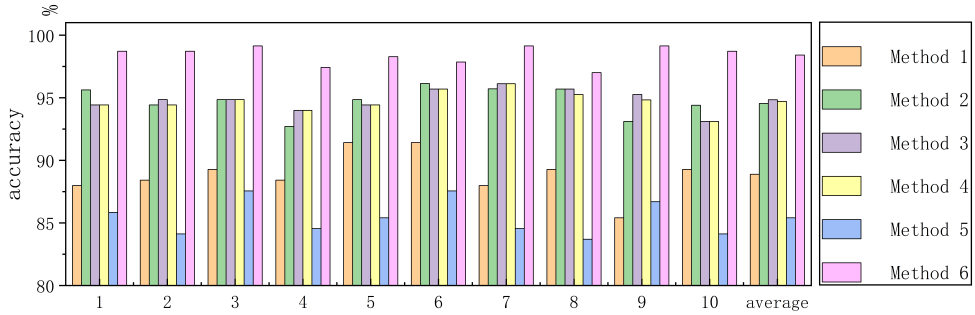

Figure 10. Detailed diagnosis results of different methods.

As shown in Table 5, under the operating conditions of a motor load of 2 horsepower and a motor speed of 1750 rpm, the proposed method achieves the highest accuracy across all metrics, demonstrating its superior diagnostic performance. From the results of Method 1 and Method 2, it is evident that the number of EMD layers has a significant impact on diagnostic accuracy, with diagnostic performance improving as the number of decomposition layers increases. Notably, when the decomposition layers are insufficient, traditional feature extraction methods yield suboptimal diagnostic results. For instance, with only 3 decomposition layers, the average classification accuracy of Method 1 and Method 4 fails to reach 90%, making these methods impractical for real-world applications. The results of Method 2, Method 3, and Method 4 also show that when extracting detailed features using fuzzy entropy, an excessive number of EMD layers can lead to overfitting. For Method 3 and Method 4, the average accuracy drops by 0.12%, with the highest and lowest accuracy remaining largely unchanged. This suggests that attempting further feature extraction on entropy features at high decomposition levels increases the risk of overfitting. To address this, the proposed approach introduces IMFE as the feature extraction method, combined with LDA for dimensionality reduction, under low decomposition-layer conditions. Experimental results show that with three EMD layers, the proposed method achieves high classification accuracy, confirming its feasibility and effectiveness in practical applications.

Table 5. Statistical diagnosis results of the six methods in CASE 2.

Model	Average Accuracy	Max Accuracy	Min Accuracy
Method 1 (EMD(3) & FE & SVM)	88.88%	91.42%	85.41%
Method 2 (EMD(7) & FE & SVM)	94.54%	96.14%	92.70%
Method 3 (EMD(7) & FE + WF(FE) & SVM)	94.83%	96.12%	93.10%
Method 4 (EMD(7) & FE + WF(FE) & CV(FE) & SVM	94.71%	96.12%	93.10%
Method 5 (EMD(3) & MFE & SVM)	85.40%	87.55%	83.69%
Method 6 (EMD(3) & IMFE & LDA & SVM, proposed method)	98.41%	99.14%	97.42%

5. Conclusions

Intelligent fault diagnosis plays a critical role in working condition maintenance for rotating machinery, significantly enhancing the operational efficiency and reliability of industrial equipment. To improve the diagnostic performance of intelligent fault diagnosis methods for various types of faults in rotating machinery, this article proposes a novel algorithm that combines EMD, IMFE, LDA, and SVM. The proposed IMFE offers advantages of few adjustable parameters, strong algorithmic robustness, and effective discrimination across different signal types. Experimental results were shown to assess the robustness of IMFE across varying time lengths and its recognition performance for different signals, verifying the stability of IMFE in feature extraction. The proposed method was compared with several other diagnostic methods for rotating machinery. The diagnostic results demonstrate a competitive edge over traditional diagnostic models, validating the strong diagnostic capabilities of this algorithm.

This article proposes an efficient and convenient fault diagnosis method for rotating machinery, which enhances equipment safety and stability while reducing maintenance costs. However, as the experiments were based solely on single-sensor signals, challenges remain concerning information loss and uncertainty. Future work will focus on improving the model to enable diagnostic applications with multi-source signals, further enhancing diagnostic accuracy and reliability.

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

Funding: This work was supported by the Key Natural Science Foundation of Higher Education Institutions of Anhui Province under grant 2024AH050154; the Open Project of Anhui Province Key Laboratory of Special and Heavy Load Robot under grant TZJQR005-2024; the Project of the Anhui International Joint Research Center for Metallurgical Process and System Science, Anhui University of Technology, under grant 2023002; and the Open Project of the Key Laboratory of Multi-disciplinary Management and Control of Complex Systems of Anhui Higher Education Institutes, Anhui University of Technology, under grant CS2023-ZD01.

Data Availability Statement: The data will be made available by the authors upon request.

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

Abbreviations

The following abbreviations are used in this article:

RMS	Rotating mechanical system
IMFE-SVM	Improved multiscale fuzzy entropy and support vector machine
EMD	Empirical mode decomposition
IMFE	Improved multiscale fuzzy entropy
LDA	Linear discriminant analysis
SVM	Support vector machine
SE	Sample entropy
FE	Fuzzy entropy
MPE	Multiscale permutation entropy
CNN	Convolutional neural network
ME	Multiscale entropy
IMFs	Intrinsic mode functions
MFE	Multiscale fuzzy entropy
FDA	Fisher discriminant analysis
CWRU	Case Western Reserve University

References

1. Shao, H.; Jiang, H.; Zhang, H.; Liang, T. Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. *IEEE Trans. Ind. Electron.* **2017**, *65*, 2727–2736. [CrossRef]
2. Zhu, Z.; Lei, Y.; Qi, G.; Chai, Y.; Mazur, N.; An, Y.; Huang, X. A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. *Measurement* **2023**, *206*, 112346. [CrossRef]
3. Heng, A.; Zhang, S.; Tan, A.C.; Mathew, J. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mech. Syst. Signal Process.* **2009**, *23*, 724–739. [CrossRef]
4. Maione, F.; Lino, P.; Maione, G.; Giannino, G. A Machine Learning Framework for Condition-Based Maintenance of Marine Diesel Engines: A Case Study. *Algorithms* **2024**, *17*, 411. [CrossRef]
5. Li, Q.; Liang, S.Y. Degradation trend prediction for rotating machinery using long-range dependence and particle filter approach. *Algorithms* **2018**, *11*, 89. [CrossRef]
6. Shen, S.; Lu, H.; Sadoughi, M.; Hu, C.; Nemani, V.; Thelen, A.; Webster, K.; Darr, M.; Sidon, J.; Kenny, S. A physics-informed deep learning approach for bearing fault detection. *Eng. Appl. Artif. Intell.* **2021**, *103*, 104295. [CrossRef]
7. Li, C.; Li, S.; Zhang, A.; He, Q.; Liao, Z.; Hu, J. Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing* **2021**, *439*, 197–211. [CrossRef]

8. Badihi, H.; Zhang, Y.; Jiang, B.; Pillay, P.; Rakheja, S. A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis. *Proc. IEEE* **2022**, *110*, 754–806. [CrossRef]
9. Md Nor, N.; Che Hassan, C.R.; Hussain, M.A. A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems. *Rev. Chem. Eng.* **2020**, *36*, 513–553. [CrossRef]
10. Huo, Z.; Martínez-García, M.; Zhang, Y.; Shu, L. A multisensor information fusion method for high-reliability fault diagnosis of rotating machinery. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 3500412. [CrossRef]
11. Zhang, L.; Zhang, Y.; Li, G. Fault-diagnosis method for rotating machinery based on SVMD entropy and machine learning. *Algorithms* **2023**, *16*, 304. [CrossRef]
12. Guan, Z.; Liao, Z.; Li, K.; Chen, P. A precise diagnosis method of structural faults of rotating machinery based on combination of empirical mode decomposition, sample entropy, and deep belief network. *Sensors* **2019**, *19*, 591. [CrossRef]
13. Ma, J.; Li, Z.; Li, C.; Zhan, L.; Zhang, G.Z. Rolling bearing fault diagnosis based on refined composite multi-scale approximate entropy and optimized probabilistic neural network. *Entropy* **2021**, *23*, 259. [CrossRef]
14. Chang, C.K.; Boyanapalli, B.K.; Wu, R.N. Application of fuzzy entropy to improve feature selection for defect recognition using support vector machine in high voltage cable joints. *IEEE Trans. Dielectr. Electr. Insul.* **2020**, *27*, 2147–2155. [CrossRef]
15. Ge, J.; Niu, T.; Xu, D.; Yin, G.; Wang, Y. A rolling bearing fault diagnosis method based on EEMD-WSST signal reconstruction and multi-scale entropy. *Entropy* **2020**, *22*, 290. [CrossRef] [PubMed]
16. Li, H.; Huang, J.; Yang, X.; Luo, J.; Zhang, L.; Pang, Y. Fault diagnosis for rotating machinery using multiscale permutation entropy and convolutional neural networks. *Entropy* **2020**, *22*, 851. [CrossRef]
17. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
18. Zhang, L.; Dong, H.; Jia, L.; Wang, B. Multi-scale Siamese Network for Few-Shot Fault Diagnosis of Bogie Component. In Proceedings of the International Conference on Electrical and Information Technologies for Rail Transportation, Beijing, China, 19–21 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 638–646.
19. Dutschmann, T.M.; Kinzel, L.; Ter Laak, A.; Baumann, K. Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation. *J. Cheminform.* **2023**, *15*, 49. [CrossRef] [PubMed]
20. Humeau-Heurtier, A. The multiscale entropy algorithm and its variants: A review. *Entropy* **2015**, *17*, 3110–3123. [CrossRef]
21. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [CrossRef] [PubMed]
22. Boudraa, A.O.; Cexus, J.C. EMD-based signal filtering. *IEEE Trans. Instrum. Meas.* **2007**, *56*, 2196–2202. [CrossRef]
23. Wang, K.; Xu, Z.J.; Gong, Y.; Du, K.L. Mechanical Fault Prognosis through Spectral Analysis of Vibration Signals. *Algorithms* **2022**, *15*, 94. [CrossRef]
24. Zheng, J.; Su, M.; Ying, W.; Tong, J.; Pan, Z. Improved uniform phase empirical mode decomposition and its application in machinery fault diagnosis. *Measurement* **2021**, *179*, 109425. [CrossRef]
25. Chen, W.; Xiao, Y. An improved ABC algorithm and its application in bearing fault diagnosis with EEMD. *Algorithms* **2019**, *12*, 72. [CrossRef]
26. Azami, H.; Fernández, A.; Escudero, J. Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis. *Med. Biol. Eng. Comput.* **2017**, *55*, 2037–2052. [CrossRef] [PubMed]
27. Wang, Z.; Chen, H.; Yuan, Z.; Wan, J.; Li, T. Multiscale Fuzzy Entropy-Based Feature Selection. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 3248–3262. [CrossRef]
28. Versaci, M.; Morabito, F.C. Image edge detection: A new approach based on fuzzy entropy and fuzzy divergence. *Int. J. Fuzzy Syst.* **2021**, *23*, 918–936. [CrossRef]
29. Xu, Y.; Yang, J.Y.; Jin, Z. A novel method for Fisher discriminant analysis. *Pattern Recognit.* **2004**, *37*, 381–384. [CrossRef]
30. Xanthopoulos, P.; Pardalos, P.M.; Trafalis, T.B.; Xanthopoulos, P.; Pardalos, P.M.; Trafalis, T.B. Linear discriminant analysis. In *Robust Data Mining*; Springer: New York, NY, USA, 2013; pp. 27–33.
31. Vishwanathan, S.; Murty, M.N. SSVM: A simple SVM algorithm. In Proceedings of the Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN'02 (Cat. No. 02CH37290), Honolulu, HI, USA, 12–17 May 2002; IEEE: Piscataway, NJ, USA, 2002; Volume 3, pp. 2393–2398.
32. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215. [CrossRef]
33. Wang, X.; Si, S.; Li, Y. Variational embedding multiscale diversity entropy for fault diagnosis of large-scale machinery. *IEEE Trans. Ind. Electron.* **2021**, *69*, 3109–3119. [CrossRef]
34. Chen, F.; Zhang, L.; Liu, W.; Zhang, T.; Zhao, Z.; Wang, W.; Chen, D.; Wang, B. A fault diagnosis method of rotating machinery based on improved multiscale attention entropy and random forests. *Nonlinear Dyn.* **2024**, *112*, 1191–1220. [CrossRef]
35. Li, Y.; Wang, X.; Si, S.; Huang, S. Entropy based fault classification using the Case Western Reserve University data: A benchmark study. *IEEE Trans. Reliab.* **2019**, *69*, 754–767. [CrossRef]