

Article

An Improved Incipient Fault Diagnosis Method of Bearing Damage Based on Hierarchical Multi-Scale Reverse Dispersion Entropy

Jiaqi Xing and Jinxue Xu * 

Marine Electrical Engineering College, Dalian Maritime University, Dalian 116026, China; 1120190185@dlnu.edu.cn

* Correspondence: xujx@dlnu.edu.cn

Abstract: The amplitudes of incipient fault signals are similar to health state signals, which increases the difficulty of incipient fault diagnosis. Multi-scale reverse dispersion entropy (MRDE) only considers difference information with low frequency range, which omits relatively obvious fault features with a higher frequency band. It decreases recognition accuracy. To defeat the shortcoming with MRDE and extract the obvious fault features of incipient faults simultaneously, an improved entropy named hierarchical multi-scale reverse dispersion entropy (HMRDE) is proposed to treat incipient fault data. Firstly, the signal is decomposed hierarchically by using the filter smoothing operator and average backward difference operator to obtain hierarchical nodes. The smoothing operator calculates the mean sample value and the average backward difference operator calculates the average deviation of sample values. The more layers, the higher the utilization rate of filter smoothing operator and average backward difference operator. Hierarchical nodes are obtained by these operators, and they can reflect the difference features in different frequency domains. Then, this difference feature is reflected with MRDE values of some hierarchical nodes more obviously. Finally, a variety of classifiers are selected to test the separability of incipient fault signals treated with HMRDE. Furthermore, the recognition accuracy of these classifiers illustrates that HMRDE can effectively deal with the problem that incipient fault signals cannot be easily recognized due to a similar amplitude dynamic.

Keywords: incipient fault; hierarchical multi-scale reverse dispersion entropy; feature extraction



Citation: Xing, J.; Xu, J. An Improved Incipient Fault Diagnosis Method of Bearing Damage Based on Hierarchical Multi-Scale Reverse Dispersion Entropy. *Entropy* **2022**, *24*, 770. <https://doi.org/10.3390/e24060770>

Academic Editors: Anne Humeau-Heurtier, Hamed Azami, Mostafa Rostaghi, Daniel Abasolo, Javier Escudero

Received: 27 March 2022

Accepted: 20 May 2022

Published: 30 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the actual industrial process, a slight degree of deviation is regarded as a minor symptom. The fault with minor symptoms is defined as the incipient fault [1]. This means that the fault amplitude of incipient faults is less obvious, which increases recognition difficulty for these fault signals in the time domain or frequency domain [2].

Incipient faults are similar to each other, which is characterized by a slight deviation from the normal health condition, but each fault and normal state belong to two classes of objective existence, respectively. Therefore, it is significant to select and improve signal treatment methods to reflect the great difference.

Different from feature extraction methods, such as deep learning, which enhances the learning ability by constructing various network structures [3–5], the signal treatment method decreases the distinguishing difficulty of incipient fault signals by restructuring new variables which can embody more obvious difference information. The amplitude of the signal changes with the passage of time [6]. Furthermore, amplitude deviations of fault are different from those of the normal state. Many methods with the measurement of the disorder of nonlinear time series have been proposed and applied to the field of fault diagnosis [7], such as approximate entropy (AE) [8], sample entropy (SE) [9], fuzzy entropy (FE) [10], and permutation entropy (PE) [9]. For example, approximate entropy has been used in bearing fault diagnosis;

different fault sizes of the bearing inner are measured by approximate entropy [11]. Compared with approximate entropy, data length does not influence the calculation of sample entropy. Sample entropy and empirical mode decomposition are combined as the battery fault detection method [12]. Fuzzy entropy introduces the idea of threshold segmentation. A Euclidean distance based multi-scale fuzzy entropy method has been proposed to diagnose bearing faults, which measures the similarity of two vectors with continuous values from zero to one based on the Euclidean distance of the two vectors [13]. An improved FE named refined composite multi-scale fuzzy entropy (RCMFE) has been applied to diagnose the significant bearing fault [14]. Being different from AE, SE, RCMFE, and FE, PE compares and analyzes the order of amplitude values to obtain the corresponding feature information rather than considering the value of the time series. Therefore, PE possesses the merit of fast computation. However, it ignores the difference between different amplitude values, which will cause the omission of important amplitude information. A method based on variational mode decomposition and permutation entropy has been used in wind turbine roller bearing fault diagnosis, and its feature extraction ability is superior to PE [15]. All the same, PE and its improved methods play an important role in fault diagnosis, such as weighted PE (WPE) [16], dispersion entropy (DE) [17], reverse permutation entropy (RPE) [18], and reverse dispersion entropy (RDE) [19]. For example, WPE has been combined with an improved support vector machine as a bearing fault classification method [20]. Both WPE and DE add amplitude information to PE, but DE is proposed to generate different fluctuation dispersion patterns by mapping each element of a measured series to different classes, which means that DE has faster calculation and the signals that are treated with DE have better separability [17,21]. To promote the feature extraction ability of DE, an improved refined composite multi-scale dispersion entropy (RCMDE) has been proposed to isolate bearing fault data provided by Case Western Reserve University [22]. The optimized method RPE is defined as the distance from white noise and it is better than PE in feature extraction [18]. The merits of DE and RPE are combined in RDE; therefore, RDE has better feature extraction ability than DE and RPE [19,23]. Based on RDE, multi-scale reverse dispersion entropy (MRDE) [24] has been proposed in 2022; it can describe the disorder of the signal from different scales, which solves the problem that RDE ignores useful information on other scales, and it obtains better performance on feature extraction of the ship-radiated noise.

So far, a lot of recent work has focused on regular fault data for testing the optimized diagnosis method and proving the promotion of recognition accuracy. Different from regular faults, once incipient fault occurs in a system, its amplitude difference from the normal state is more slight [1]. Regular methods, which only extract amplitude change information, may not be satisfied to deal with incipient fault data. Compared with normal state signals, incipient fault signals are in a higher frequency band. Because RDE and MRDE only extract amplitude difference features in the low frequency band, the obvious incipient fault feature with higher frequency range will be omitted, which will cause the lower incipient fault recognition accuracy. To overcome the defect of MRDE, an improved hierarchical multi-scale reserve dispersion entropy (HMRDE) method is proposed to enhance the separability of signals.

The contributions are summarized as follows:

- (1) A new fault extraction approach, named HMRDE, based on MRDE is proposed to extract obvious difference features with various frequency ranges. It introduces hierarchical thought to MRDE and uses hierarchical nodes to analyze the frequency difference features of incipient fault signals for the first time.
- (2) HMRDE enhances the disorder difference of each state by calculating the change deviation with a high-frequency operator and reflects this difference by entropy values of hierarchical nodes obviously, which helps classifiers greatly in recognizing incipient faults.

The remainder is organized as follows. Section 2 briefly describes the motivation of the proposed method for incipient faults and describes the proposed method, HMRDE. Section 3 gives a numerical example to test the feature extraction ability of HMRDE for similar signals and a real fault diagnosis experiment to test the effectiveness of HMRDE for real incipient faults. The findings and their implications are discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Aim Formulation and Methods

2.1. Aim Formulation

Fault signals and normal signals are two kinds of objective existence. Both fault signals and normal signals have inherent center frequency. Compared with normal signals, fault signals are in a higher frequency band. Different kinds of faults have different frequency domain characteristics. Normal signals and fault signals with fixed center frequency can be analyzed in the frequency domain. Each health status signal can be expressed in the form of a periodic $f(t)$ with a period T , (T can approach positive infinity). Fourier decomposition of $f(t)$ can be defined as

$$f(t) = d + \sum_{n=1}^{\infty} (a_n \cos(\frac{2\pi n}{T}t) + b_n \sin(\frac{2\pi n}{T}t)) \quad (1)$$

where d represents constant term, a_n and b_n denote amplitudes of periodic function $\sin(w_n t)$ and $\cos(w_n t)$ with frequency w_n , $w_n = \frac{2\pi n}{T}$. In the time domain, the amplitude difference of each fault is not obvious. For example, assume a normal state and one incipient fault, $f_0(t)$ and $f_1(t)$, respectively, is described through $\sin(wt)$ as

$$\begin{cases} f_0(t) = a_1 \sin(w_1 t) \\ f_1(t) = a_1 \sin(w_1 t) + \sigma \sin(w_2 t) \end{cases} \quad (2)$$

where $1 \leq w_1 < w_2$ and σ is set to be $0.02a_1$ [25]. Thus, their first order derivatives can be calculated as

$$\begin{cases} f'_0(t) = a_1 w_1 \cos(w_1 t) \\ f'_1(t) = a_1 w_1 \cos(w_1 t) + \sigma w_2 \cos(w_2 t) \end{cases} \quad (3)$$

It can be seen that $|(f'_1(t) - f'_0(t))| \geq |(f_1(t) - f_0(t))|$, which shows that the amplitude difference between f'_1 and f'_0 is greater than that between f_1 and f_0 . f'_1 is more obviously different from f'_0 .

This shows that the natural frequency characteristics of the incipient fault signal are obviously different from those of the normal signal, and the natural center frequency of the incipient fault is in a higher frequency band.

Thus, the motivation of the proposed method regarding the recognition of incipient faults is that the signal treatment method needs to consider the obvious differences of each incipient fault from others in higher frequency ranges, and it reflects them greatly.

2.2. Methods

Hierarchical multi-scale reverse dispersion entropy defeats the defect that multi-scale reverse dispersion entropy only analyzes the obvious differences of each incipient fault from others in low frequency ranges.

For a time series $\{x(1), x(2), \dots, x(n)\}$, we define the averaging operator Q_0 and high-frequency operator Q_1 as follows [26]

$$Q_0(x) = \frac{x(i) + x(i+1)}{2}, i = 1, 2, \dots, n \quad (4)$$

$$Q_1(x) = \frac{x(i) - x(i+1)}{2}, i = 1, 2, \dots, n \quad (5)$$

where $Q_0(x)$ and $Q_1(x)$ can be regarded as approximations of a filtering smooth operation and an average backward differential operation, and they can depict the low frequency and high frequency information of the time series respectively.

The matrix form of operators Q_j^k ($j = 0, 1$) at hierarchical layer k can be expressed as

$$Q_j^k = \begin{bmatrix} \frac{1}{2} & \underbrace{0 \cdots 0}_{2^{k-1}-1} & \frac{(-1)^j}{2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \underbrace{0 \cdots 0}_{2^{k-1}-1} & \frac{(-1)^j}{2} & \cdots & 0 & 0 & 0 \\ \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \underbrace{0 \cdots 0}_{2^{k-1}-1} & \frac{(-1)^j}{2} \end{bmatrix}_{a \times b} \tag{6}$$

where $a = n - 2^k + 1$ and $b = n - 2^{k-1} + 1$. The hierarchical decomposition structure is exposed in Figure 1.

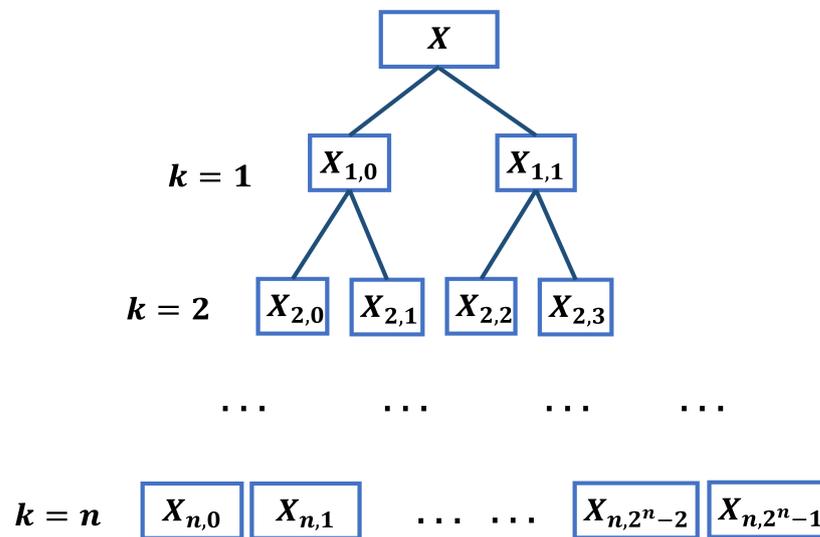


Figure 1. The hierarchical decomposition structure.

Furthermore, hierarchical nodes can be calculated by

$$X_{k,e} = Q_{r_k}^k \cdot Q_{r_{k-1}}^{k-1} \cdots \cdots Q_{r_1}^1 \cdot X \tag{7}$$

where $X = \{x(1), x(2), \cdots, x(n)\}$, and vector $[r_1, r_2, \cdots, r_k]$ is given by non-negative integer e

$$e = \sum_{m=1}^k 2^{k-m} r_m \tag{8}$$

where $e \in \{0, 1, \cdots, 2^k - 1\}$, r_m is 0 or 1, which denotes the average or difference operator at layer m . The partial calculation process is shown in Figure 2.

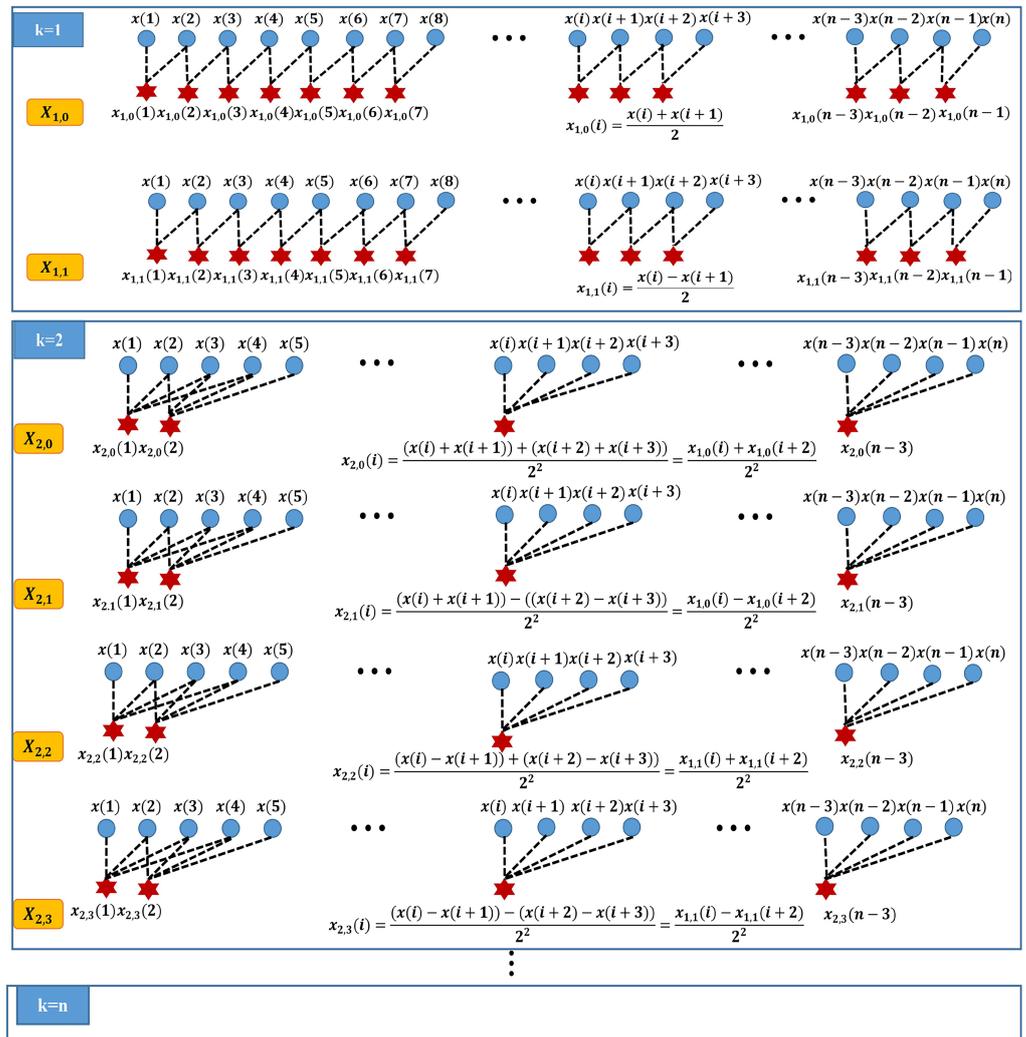


Figure 2. The partial calculation process regarding hierarchical decomposition structure.

In Figures 1 and 2, the larger the value of k , the higher the utilization rate of high-frequency operators. The higher frequency range of time series is analyzed by the node on the right side of HMRDE. In a certain unique frequency band, the change of one incipient fault signal must be obviously different from other faults. Therefore, the difference information from low frequency range to high frequency range can be analyzed by increasing layer k suitably.

Entropy is a reflection of signal disorder, so this signal difference can be measured by multi-scale reverse dispersion entropy. For a certain component with length $n - 2^k + 1$, $X_{k,e} = \{x_{k,e}(1), x_{k,e}(2), \dots, x_{k,e}(n - 2^k + 1)\}$, the coarse-grained result is as follows

$$x_{k,e}^s(j) = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_{k,e}(i) \tag{9}$$

where s is the scale factor of MRDE. Map $X_{k,e}^s = \{x_{k,e}^s(1), x_{k,e}^s(2), \dots, x_{k,e}^s((n - 2^k + 1)/s)\}$ to $Y_{k,e}^s$ using the normal cumulative distribution function, which is expressed as

$$y_{k,e}^s(j) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_{k,e}^s(j)} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \tag{10}$$

where μ and σ^2 denote expectation and variance, respectively, and $y_{k,e}^s(j)$ ranges from 0 to 1. Then, map each $y_{k,e}^s(j)$ to the sequence $\{1, 2, \dots, c\}$ by linear transformation as follows

$$z_{k,e}^{s,c}(j) = \text{round}(c * y_{k,e}^s(j) + 0.5) \tag{11}$$

where $\text{round}(\cdot)$ represents the integral function and c is the class number. This formula limits the magnitude of $y_{k,e}^s(j)$ to an integer range of $[1, c]$. The embedding vector of reconstructed matrix $Z_{k,e}^{s,c,m} = \{z_{k,e}^{s,c,m}(1), z_{k,e}^{s,c,m}(2), \dots, z_{k,e}^{s,c,m}((n - 2^k + 1)/s - (m - 1)\tau)\}$ with the embedding dimension m is defined by

$$z_{k,e}^{s,c,m}(j) = [z_{k,e}^{s,c}(j), z_{k,e}^{s,c}(j + \tau), \dots, z_{k,e}^{s,c}(j + (m - 1)\tau)] \tag{12}$$

where τ represents the time delay. Each $z_{k,e}^{s,c,m}(j)$ corresponds to a dispersion mode which can be described by $[\pi_{v_0, \dots, v_{m-1}}]$. Calculate the relative frequency of each dispersion mode by the following equation

$$p_j(\pi_{v_0, \dots, v_{m-1}}) = \frac{\text{Number}(\pi_{v_0, \dots, v_{m-1}})}{((n - 2^k + 1)/s - (m - 1)\tau)} \tag{13}$$

where $\text{Number}(\cdot)$ is the number of mappings from $z_{k,e}^{s,c,m}(j)$ to $\{\pi_{v_0, \dots, v_{m-1}}\}$. Reverse dispersion entropy (RDE) is used to calculate the entropy value of each node $X_{k,e}$ in the hierarchical layer. RDE is defined as the distance to white noise by combining distance information [19]. The entropy value of each node $X_{k,e}^s$ with scale factor s in the hierarchical layer can be expressed as [19]

$$RDE(X_{k,e}^s) = \sum_{j=1}^{c^m} (p_j - \frac{1}{c^m})^2 \tag{14}$$

when $p_j = \frac{1}{c^m}$, the value of $RDE(X_{k,e}^s)$ is 0 (minimum value) [19]. This means that the smaller the RDE value is, the more disorderly the signal is. The HMRDE of a given time series X is defined as

$$\text{HMRDE}(X) = [RDE(X_{k,0}^s), RDE(X_{k,1}^s), \dots, RDE(X_{k,2^k-1}^s)] \tag{15}$$

Notably, $X_{k,0}$ is generated by k operations of filtering smooth and $X_{k,2^k-1}$ is acquired through k calculations of mean change deviation of adjacent sample values. $X_{k,0}$ equals sample entropy at 2^k scale in multi-scale analysis. Based on HMRDE, the proposed fault diagnosis scheme for rolling bearings is given in Figure 3. The specific steps for the proposed scheme are given as follows.

Step 1: Collect vibration signals with l classes. Each type of data file has the same number of time series samples, and each series sample has the same number of consecutive non-overlapping points. Divide the signals randomly into two groups: one for the training samples, which can be used to optimize the parameters of the method, and the other for the testing samples.

Step 2: Determine the hyperparameter adjustment range and set the hyperparameter initialization. For example, the adjustment range of layer k is $\{n, n + 1, \dots, n_{max}\}$.

Step 3: Select optimal hyperparameters of HMRDE. In the training stage, the hyperparameters are adjusted, and the same classifier is used to test the effectiveness of different parameter setting methods. Select the HMRDE parameter setting with the best feature extraction effect. Under this parameter setting, the data processed by HMRDE has better distinguishability, and the same classifier can achieve higher classification accuracy. The flowchart of hyperparameter optimization of layer k is shown in Figure 4. Assume that the optimal hyperparameter layer k is m .

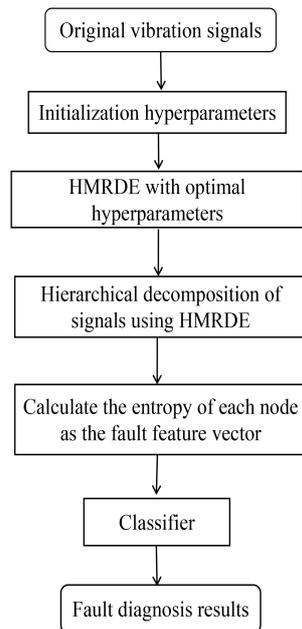


Figure 3. The proposed fault diagnosis scheme for rolling bearings.

Step 4: Hierarchical decomposition of testing signals using HMRDE with optimal hyperparameters, which generates hierarchical nodes of layer k ($k = m$). Then, calculate entropy values of these nodes as the fault feature vectors.

Step 5: Use the classifier to classify the test dataset processed by HMRDE.

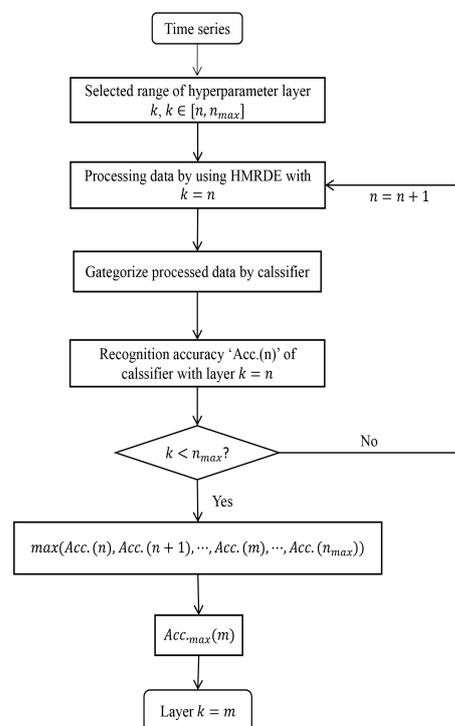


Figure 4. The flowchart of hyperparameter optimization of layer k .

3. Results

3.1. Case 1: Numeral Example

Assume the normal condition $f_0(t)$ and incipient fault signals $f_1(t)$ are described as

$$\begin{cases} f_0(t) = \sin(50t) \\ f_1(t) = \sin(50t) + 0.02 \sin(100t) \end{cases} \quad (16)$$

The amplitude of incipient faults is similar to that of the health condition from Equation (16). Figure 5 shows that there must be relatively obvious fault features in the higher frequency range when the amplitude difference information with the low frequency range is very hidden. The first order derivative of the time series under two health conditions is calculated as

$$\begin{cases} f_0'(t) = 50 * \sin(50t) \\ f_1'(t) = 50 * \sin(50t) + 0.02 * 100 * \sin(100t) \end{cases} \quad (17)$$

and it is depicted in Figure 6. The difference of $f_1(t)$ from $f_0(t)$ shown in Figure 6 is more obvious than that depicted in Equation (16), which indicates that relatively obvious difference information exists in the higher frequency range rather than in the low frequency band. Figure 6 shows that the standardized derivative values of f_1' are more obviously different from those of f_0' , which illustrates that difference information with a higher frequency band can be reflected by a derivative operation. At the same time, the relative obvious fault features with higher frequency range also can be reflected by a high-frequency operator, as shown in Figure 7.

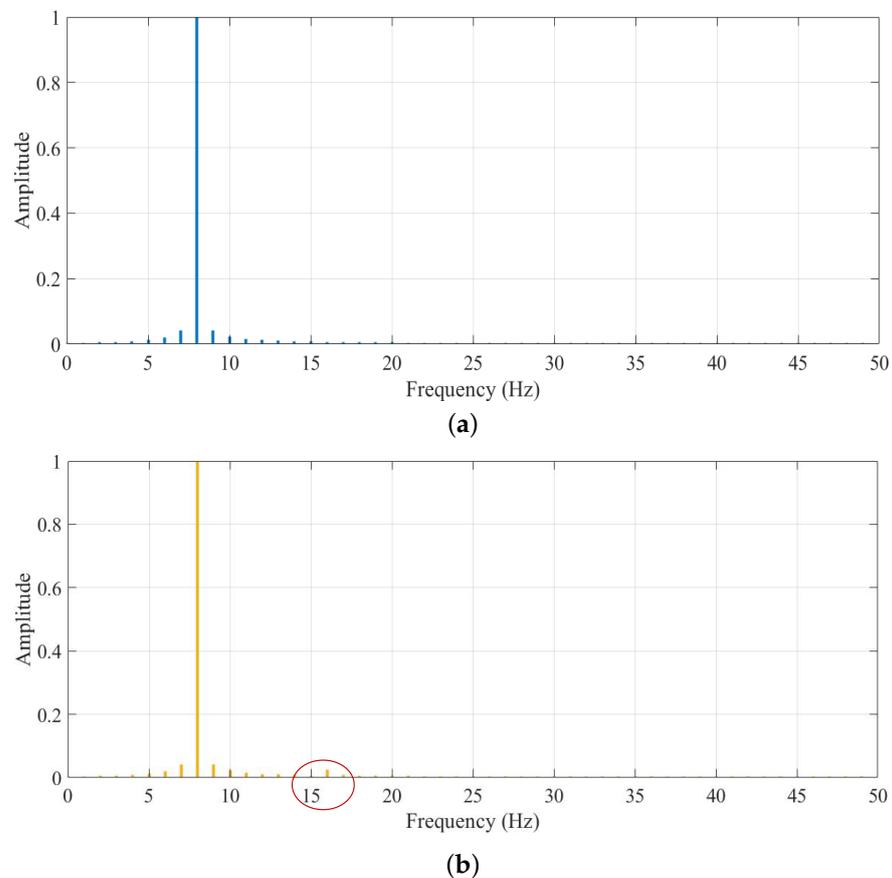


Figure 5. The frequency spectrum of time series under two health conditions. (a) Frequency spectrum of f_0 ; (b) Frequency spectrum of f_1 .

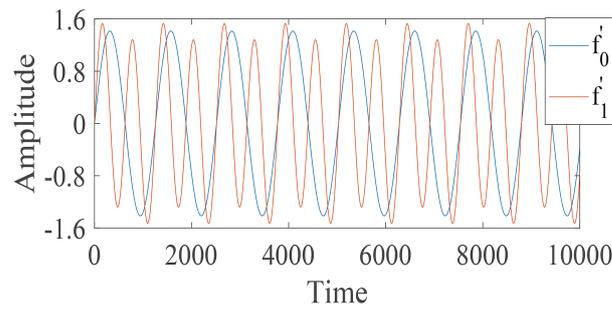


Figure 6. The waveform of standardized first order derivative values of time series under two health conditions.

The node entropy values of time series under two health conditions are depicted in Figure 7, and these node entropy values are calculated by HMRDE, where layer k is 2, embedding dimension m is 3, time delay τ is 1, scale factor s is 1, and class number c is 5. In Figure 7, the entropy value of $X_{2,3}$ of f_1 is lower than that of f_0 , which is easily distinguished. The disorder of the signal treated with a high-frequency operator can be effectively reflected by the entropy values of the high frequency node.

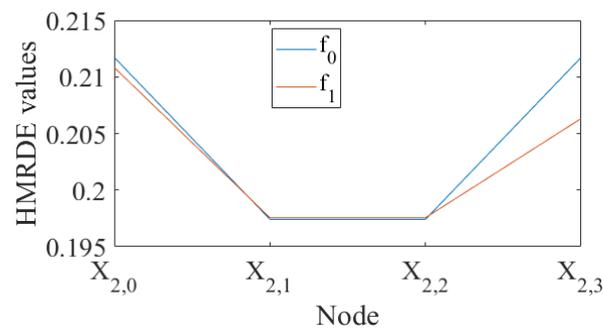


Figure 7. The node entropy values of time series under two health conditions.

3.2. Case 2: Dataset Provided by Padborn University in Germany

In order to verify the practicability of HMRDE, the dataset provided by Padborn University in Germany [25,27] is used to carry out the real incipient fault diagnosis experiment. Specifically, the fault data generated by the accelerated lifetime test was used in the recognition of incipient fault in 2020 [28].

The basic setup of operation parameters is that $N = 1500$ rpm, $M = 0.7$ Nm, and $F = 1000$ N [25]. Then, fault data are assigned to five levels according to Table 1.

Table 1. Damage levels to determine the extent of damage.

Damage Level	Assigned Percentage	Limits for Bearing
1	0–2%	≤ 2 mm
2	2–5%	> 2 mm
3	5–15%	> 4.5 mm
4	15–35%	> 13.5 mm
5	$> 35\%$	> 31.5 mm

The dataset consists of three kinds of health conditions: normal, inner ring (IR) fault, and outer ring (OR) fault; the types of these faults are: single point (S) fault, repetitive (R) fault, and multiple (M) fault. All the incipient faults of rolling bearing belong to level 1 (extent of damage: 0–2%). A detailed description of the datasets is illustrated in Table 2.

Table 2. Detailed description of datasets.

Code <i>n</i>	Component <i>m</i>	Combination	Characteristic	Level
N01	–	–	–	–
A01	OR	R	distributed	1
B01	OR+IR	M	distributed	1
I01	IR	M	single point	1
I02	IR	R	single point	1

The number of sample values is 256,000 for each fault and the length of each time series input is 3000. There are 85 time series inputs. Then, 60% of the time series inputs is randomly chosen for training and the remaining 40% is chosen for testing.

The waveform of the two random time series samples under five bearing conditions is sketched in Figure 8. It indicates that the amplitudes of five health condition signals are similar.

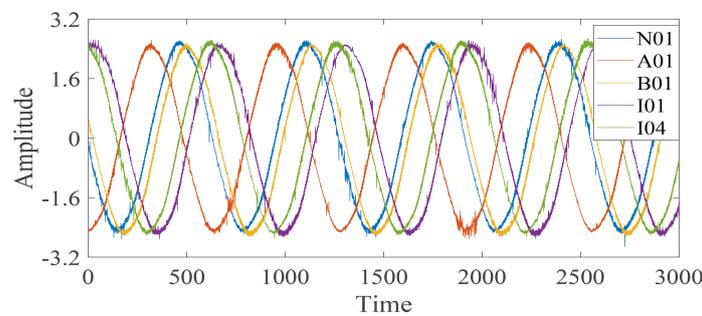


Figure 8. Waveform of time series under five bearing health conditions.

Then, conclude the feature frequency spectrum using FFT transform, as shown in Figure 9. The sample frequency is 64 kHz and the sample length is 256,000. In Figure 9, it is difficult to distinguish the five health conditions through amplitudes; although, the amplitudes with a low frequency range are high. However, the frequency features of five health conditions are obviously different from each other in the frequency band marked by the red, five pointed star, and the frequency spectrum in this frequency range is sketched in Figure 10. It illustrates that there must be obvious frequency difference information of the five health condition signals in a certain unique frequency range; although, their amplitudes are very low and similar. The frequency bands with distinct fault characteristics for A01, B01, I01, I02, and N01 are described as *Fre.(A01)*, *Fre.(B01)*, *Fre.(I01)*, *Fre.(I02)*, and *Fre.(N01)*.

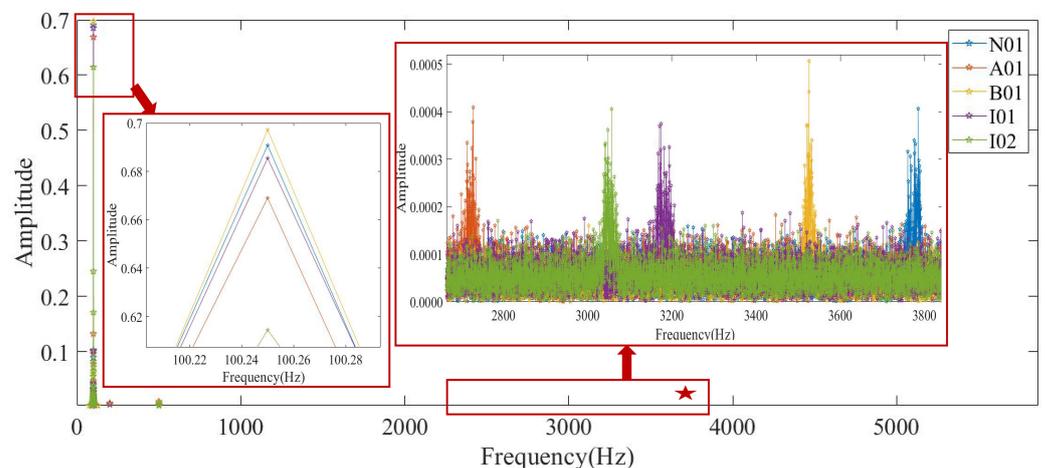


Figure 9. The frequency spectrum of bearing samples.

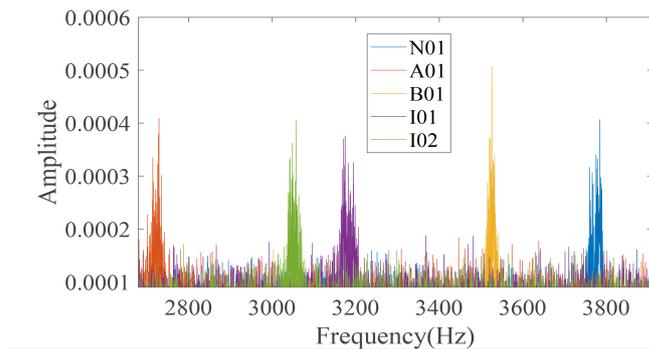


Figure 10. The frequency spectrum part marked by red, five pointed star.

In the real application of HMRDE, there are five parameters that need to be determined. Because the low-frequency smoothing operation in HMRDE can be regarded as an MRDE calculation with scale, for example, entropy values of $X_{k,0}$ of HMRDE are equal to MRDE values of X with scale 2^k , the scale of HMRDE should not be larger. Because the obvious difference information exists in 2690 ~ 3820 Hz from these two figures, layer k cannot be selected too large; usually, it is set as 2–6. The embedding dimension m and the number of classes c can be 3 and 5, respectively, and the time delay τ is 1. For more information about the parameters m , c , and τ , please refer to the literature [22,24]. Assume the embedding dimension m is 3, the time delay τ is 1, the scale factor s is 1, the layer k is 3, and the class number c is 5. Then, calculate the HMRDE of the five health condition data. The node entropy values of a time series under five bearing conditions are shown in Figure 11. In Figure 11, node $X_{3,5}$ and $X_{3,7}$ of five health conditions are more easily distinguished than node $X_{3,0}$, $X_{3,2}$, $X_{3,4}$, and $X_{3,6}$, which explains that the obvious difference information of the dataset exists in some unique higher frequency ranges rather than low frequency ranges.

A01 with lower frequency range $Fre.(A01)$ is more ordered than other health conditions, and the entropy values of nodes of A01 are larger than others, as shown in Figure 11. The difference of the disorder of each fault is more easily separated through calculations of mean change deviation with high-frequency operators.

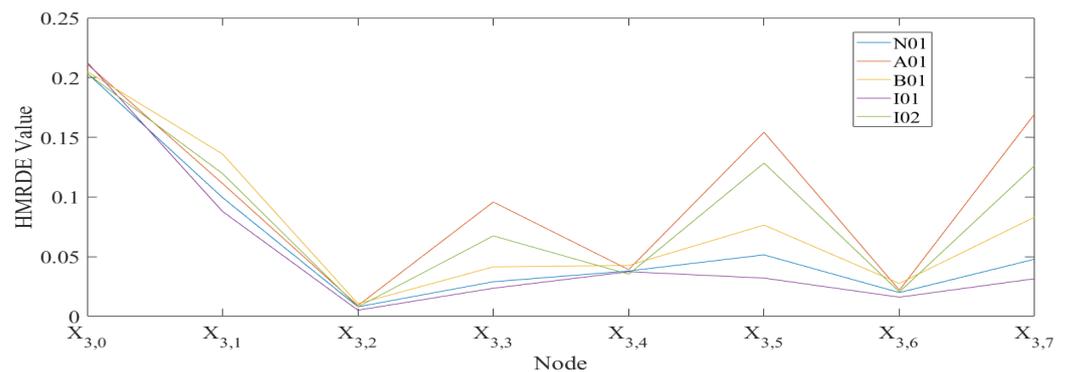


Figure 11. The node entropy values of a time series under five bearing conditions.

At the same time, in the test of incipient fault recognition, the setting of the parameters of the proposed method is important. To test the advantage of the proposed method, the different classifiers are selected to recognize the incipient faults. To guarantee the reliability of the experiment, eight classifiers are selected to test the effectiveness of HMRDE. The selected classifiers are linear discriminant (LD), linear support vector machine (SVM), medium Gaussian support vector machine (MGSVM), quadratic support vector machine (QSVM), coarse K nearest neighbors (CKNN), bagged trees (BT), medium tree (MT), and boosted trees (BoT).

Experiments for each setting are repeated five times. The influence of the selected layer k on recognition accuracy of classifiers is shown in Figure 12, which depicts the

highest recognition accuracy of eight classifiers in the training phase. In the training phase, when $k = 2, 3, 4, 5, 6$, the incipient fault recognition accuracy is on average $78.3 \pm 0.69\%$, $93.9 \pm 0.36\%$, $94.1 \pm 1.17\%$, $96.5 \pm 0.99\%$, and $94.6 \pm 1.11\%$. Therefore, the layer k of HMRDE can be 5 for these five health condition signals.

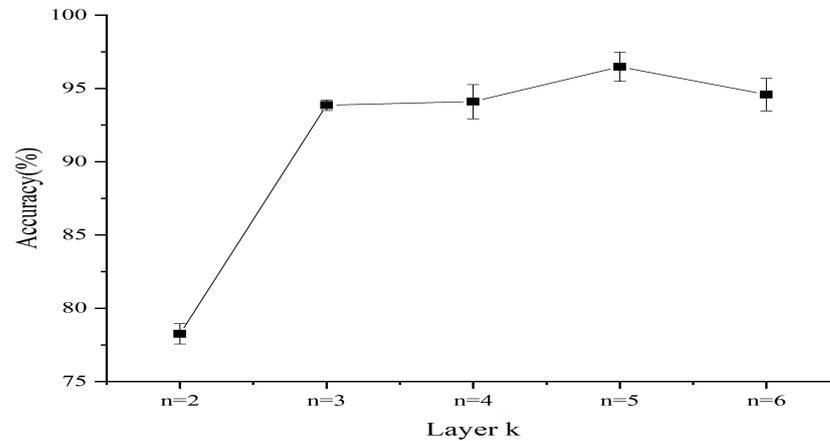


Figure 12. Recognition accuracy with different layers in training phase.

Figure 13 shows the recognition accuracy with different layers in the testing phase; it can be seen that the highest accuracy is on average $97.7 \pm 0.83\%$ when $k = 5$. Furthermore, Figure 14 displays confusion matrix results of SVM for the inputs treated with the proposed method. It shows that the data distinguished by simple classifier SVM is more easily recognized after being treated with HMRDE, and the difference information of incipient fault inputs is reflected greatly with HMRDE. Bearing data with different conditions are described in Table 3 [25]. To test the effectiveness of HMRDE for data under different conditions, the settings of HMRDE are the same in these tests, and experiments for data under each condition are repeated five times. In these experiments, the settings of HMRDE are $m = 3$, $\tau = 1$, $s = 1$, $k = 5$, and $c = 5$. Effectiveness test results of HMRDE for data with different conditions are shown in Figure 15, which depicts the highest recognition accuracy of eight classifiers. In Figure 15, the highest recognition accuracies of eight classifiers with data treated with HMRDE are $97.65 \pm 0.83\%$, $91.38 \pm 1.12\%$, $95.8 \pm 1.28\%$, and $91.2 \pm 0.65\%$, which are all higher than 90%. This illustrates that HMRDE can effectively extract incipient fault features from incipient data under different conditions.

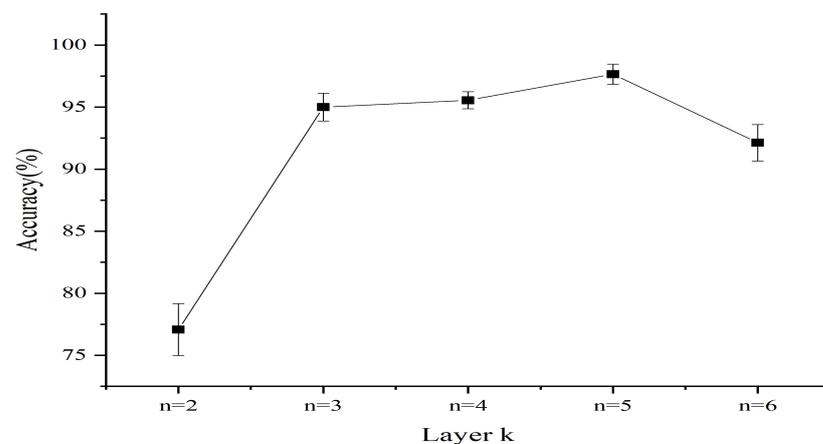


Figure 13. Recognition accuracy with different layers in testing phase.

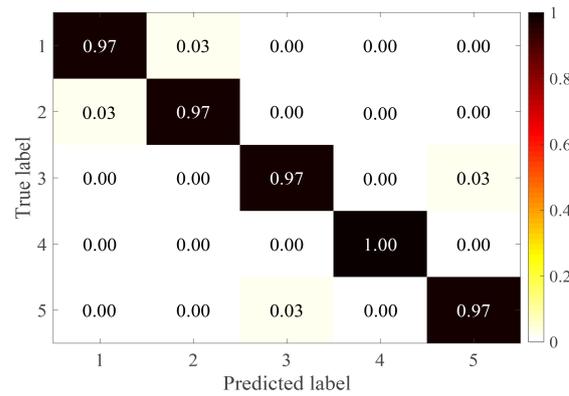


Figure 14. Classification results of SVM for the inputs treated with the proposed method.

Table 3. Bearing data with different conditions.

Condition Number	Rotational Speed (rpm)	Load Torque (Nm)	Radial Force (N)	Name of Setting
1	1500	0.7	1000	N15M07F10
2	900	0.7	1000	N09M07F10
3	1500	0.1	1000	N15M01F10
4	1500	0.7	400	N15M07F04

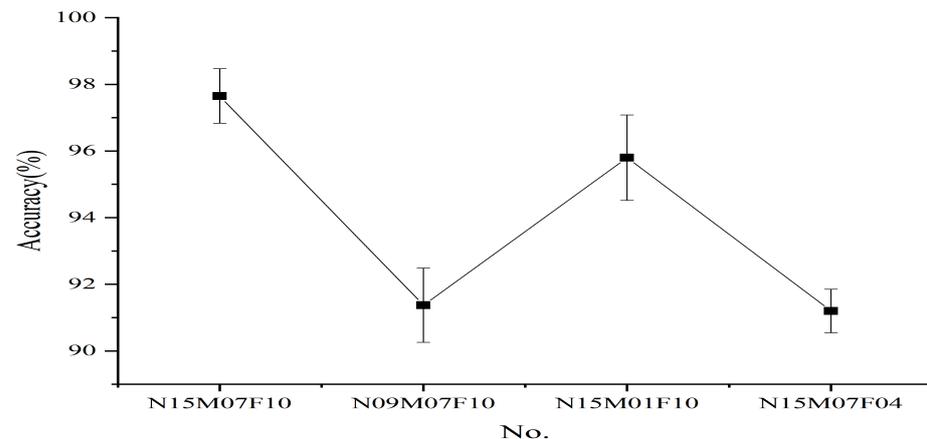


Figure 15. Effectiveness test results of HMRDE for data under different conditions.

Assume the inputs treated with HMRDE, MRDE, and the standardization method are named 'HMRDE data', 'MRDE data', and 'Stand. data', respectively. Here, standardization method refers to the zero-mean normalization method. The classification accuracy of these classifiers for 'HMRDE data', 'MRDE data', and 'Stand. data' is summarized in Figure 16. Experiments for data with different treatments are repeated five times. In Figure 16, the best average accuracy of the selected classifier for 'Stand. data' is 85.9% and that for 'HMRDE data' is 97.7%. Compared with 'Stand. data', the accuracy of all classifiers for data treated with HMRDE is increased by 11.8%, 17.6%, 77.7%, 63.5%, 76.6%, 64.1%, 65.3%, and 48.9%, and it is increased by 79.5%, 74.1%, 69.5%, 72.9%, 70.0%, 69.5%, and 53% compared to 'MRDE data', whose classifier classification accuracy is 21.1%.

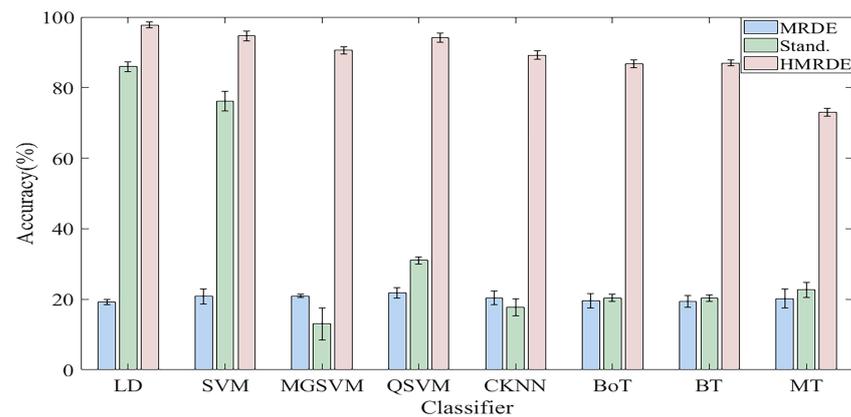


Figure 16. Classification accuracy of different classifiers for the data treated with HMRDE and standardization method.

4. Discussion

Fault state and normal state are two kinds of objective existence, but MRDE cannot extract relatively obvious difference successfully, which decreases the recognition accuracy, as depicted in Figure 16. Once incipient fault occurs in a system, the values of samples slightly fluctuate in the time domain, but amplitude dynamic deviation speed and frequency change may be obvious, as shown in Equation (17) and Figure 6. The obvious difference information might exist in a higher frequency band, as shown in Figures 5 and 10, and the difference information with the high frequency band can be extracted by calculating the mean change deviation of sample values, such as derivative operation and high-frequency operator filtering, which is manifested in Equation (17), Figures 6, 7 and 11. Furthermore, the difference in terms of disorder of each health condition is obviously reflected by MRDE values of hierarchical nodes, which enhances the separability of fault features and promotes the recognition accuracy of various classifiers, as depicted in Figures 7, 11, and 16. It takes between 2.5 and 3.0 s to compute HMRDE at five layers for a time series with 3000 points. HMRDE increases the calculation complexity compared to MRDE, and it has the same shortcoming that its hyperparameter selection requires expert experience. However, HMRDE defeats the drawback that MRDE omits frequency change features, and HMRDE compares favorably with deep learning approaches which require more hyperparameter adjustments and a more complex learning process.

In the practical application of the fault identification method of rolling bearing, the longer the transmission pathway of the fault signal, the greater the interference of the signal, and the less obvious the periodic pulse under the influence of noise. In order to filter out the noise of higher frequency band and effectively identify the fault signal with the long pathway, the generated hierarchical nodes can be low-pass filtered by increasing the scale value of HMRDE, and the higher frequency noise existing in the nodes can be filtered out, so that HMRDE can handle the fault signal with a longer pathway. Other denoising methods, such as wavelet denoising and empirical mode decomposition, etc., can also be used for signal pre-denoising. However, it may increase the computational complexity and the time of diagnostic methods. Furthermore, in the practical application of the method, after the fault diagnosis model is trained with data from a single condition, the optimized diagnosis model needs to be extended to other operating conditions. Figure 15 shows that HMRDE with the same hyperparameter setting has good feature extraction ability for data in different environments.

So far, a lot of work has focused on regular fault feature extraction through various entropy methods [29–31] but how to optimize entropy methods to extract incipient fault features is still in the early phase. Therefore, the improvement of entropy methods to extract incipient fault features can be regarded as the future research direction; this research

direction needs to consider characteristics of incipient fault signals to overcome problems of entropy methods in incipient fault sample processing.

5. Conclusions

To solve the problem that it is difficult to extract fault features from incipient fault signals, an improved HMRD method is proposed based on MRDE. The filter smoothing operator and average backward deviation operator are used to extract the relatively obvious difference information between incipient fault signals in different frequency ranges and normal signals. By selecting the appropriate number of layers, the samples are smoothed and backwardly differentiated in different degrees, and the hierarchical nodes which can reflect the difference features in different frequency domains are obtained. Entropy values of hierarchical nodes are calculated by MRDE, and these entropy values are taken as new characteristic variables. It enhanced the disorder difference of each state signal and the distinguishing ability of classifier inputs, which solves the problem that MRDE omits obvious fault features in a higher frequency range and gives classifiers a higher classification accuracy. The use of HMRDE features for incipient fault classification has been introduced and its effectiveness is verified with the use of a numeral example and a dataset generated by accelerated lifetime tests. The incipient fault recognition accuracy of LD, SVM, MGSVM, QSVM, CKNN, BT, MT, and BoT for the input treated with HMRDE is much higher than that for the data treated with MRDE and normalization processing, and HMRDE does not need to consume lots of time. Furthermore, for incipient data under different conditions, effectiveness test results of HMRDE with the same hyperparameter settings are excellent. These depict the effectiveness of HMRDE in incipient fault feature extraction.

Author Contributions: Conceptualization, J.X. (Jiaqi Xing) and J.X. (Jinxue Xu); methodology, J.X. (Jiaqi Xing); software, J.X. (Jiaqi Xing); validation, J.X. (Jiaqi Xing) and J.X. (Jinxue Xu); formal analysis, J.X. (Jiaqi Xing); investigation, J.X. (Jiaqi Xing); resources, J.X. (Jiaqi Xing); data curation, J.X. (Jiaqi Xing); writing—original draft preparation, J.X. (Jiaqi Xing); writing—review and editing, J.X. (Jinxue Xu); visualization, J.X. (Jiaqi Xing); supervision, J.X. (Jinxue Xu); project administration, J.X. (Jinxue Xu); All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available on the KAT-DataCenter website of the Chair of Design and Drive Technology, Paderborn University, Germany: <http://mb.uni-paderborn.de/kat/datacenter> (accessed on 12 December 2021).

Conflicts of Interest: The authors declare no competing interests.

References

1. Zhao, Q.; Zhou, D. Incipient fault detection and isolation in closed-loop systems. In Proceedings of the 39th Chinese Process Control Conference, Shenyang, China, 27–29 July 2020; pp. 636–641.
2. Russell, E.; Chiang, L.H.; Braatz, R.D. *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*; Springer: London, UK, 2000; p. 101.
3. Zhang, M.; Li, X.; Wang, R. Incipient fault diagnosis of batch process based on deep time series feature extraction. *Arab. J. Sci. Eng.* **2021**, *46*, 10125–10136. [[CrossRef](#)]
4. Zhao, H.; Sun, S.; Jin, B. Sequential fault diagnosis based on LSTM neural network. *IEEE Access* **2018**, *6*, 12929–12939. [[CrossRef](#)]
5. Yang, J.; Guo, Y.; Zhao, W. Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators. *Neurocomputing* **2019**, *360*, 85–96. [[CrossRef](#)]
6. Ivar, Z.; Aimar, K.; Ergo, R.; Anni, M.; Madis, J.; Toomas, T.; Oleg, A.; dC Sergio, R.; Taavo, T. Enhanced efficiency of nitritating-anammox sequencing batch reactor achieved at low decrease rates of oxidation–reduction potential. *Environ. Eng. Sci.* **2019**, *36*, 350–360.
7. Yun, K.; Chong, Y.; Enzhe, S.; Liping, Y.; Quan, D. Fault diagnosis method of diesel engine injector based on hierarchical weighted permutation entropy. In Proceedings of the 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Glasgow, UK, 17–20 May 2021; pp. 1–6.

8. Xingwei, Y.; Dawei, L.; Jun, Z.; Jianwei, W. Radar jamming detection based on approximate entropy and moving-cut approximate entropy. In Proceedings of the IET International Conference on Information Science and Control Engineering 2012 (ICISCE 2012), Shenzhen, China, 7–9 December 2012; pp. 1–6.
9. Zhang, H.; He, S. Analysis and comparison of permutation entropy, approximate entropy and sample Entropy. In Proceedings of the 2018 International Symposium on Computer, Consumer and Control (IS3C), Taichung, Taiwan, 6–8 December 2018; pp. 209–212.
10. Liu, H.; Xie, H.; He, W.; Wang, Z. Characterization and classification of EEG sleep stage based on fuzzy entropy detection and isolation for electro-mechanical actuators. *J. Data Acquis. Process.* **2010**, *25*, 484–489.
11. Zhao, J.; Liu, Y. Approximate entropy based on hilbert transform and its application in bearing fault diagnosis. In Proceedings of the 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Xi'an, China, 15–17 August 2018; pp. 41–44.
12. Li, X.; Dai, K.; Wang, Z.; Han, W. Lithium-ion batteries fault diagnostic for electric vehicles using sample entropy analysis method. *J. Energy Storage* **2020**, *27*, 101121. [[CrossRef](#)]
13. Zhou, R.; Wang, X.; Wan, J.; Xiong, N. Edm-Fuzzy: An Euclidean Distance Based Multiscale Fuzzy Entropy Technology For Diagnosing Faults Of Industrial Systems. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4046–4054. [[CrossRef](#)]
14. Gituku, E.W.; Kimotho, J.K.; Njiri, J.G. Cross-domain bearing fault diagnosis with refined composite multiscale fuzzy entropy and the self organizing fuzzy classifier. *Eng. Rep.* **2020**, *3*, 12307. [[CrossRef](#)]
15. An, X.; Pan, L. Bearing fault diagnosis of a wind turbine based on variational mode decomposition and permutation entropy. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2017**, *231*, 200–206. [[CrossRef](#)]
16. Fadlallah, B.; Chen, B.; Keil, A.; Príncipe, J. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Phys. Rev. E* **2013**, *23*, 022911. [[CrossRef](#)]
17. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [[CrossRef](#)]
18. Bandt, C. A new kind of permutation entropy used to classify sleep stages from invisible EEG microstructure. *Entropy* **2017**, *19*, 197. [[CrossRef](#)]
19. Li, Y.; Gao, X.; Wang, L. Reverse dispersion entropy: A new complexity measure for sensor signal. *Sensors* **2019**, *19*, 5203. [[CrossRef](#)] [[PubMed](#)]
20. Shenghan, Z.; Silin, Q.; Chang, W.; Yiyong, X.; Yang, C. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. *Sensors* **2018**, *18*, 1934.
21. Li, R.; Ran, C.; Luo, J.; Feng, S.; Zhang, B. Rolling bearing fault diagnosis method based on dispersion entropy and SVM. In Proceedings of the 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Beijing, China, 15–17 August 2019; pp. 596–600.
22. Songrong, L.; Wenxian, Y.; Youxin, L. Fault diagnosis of a rolling bearing based on adaptive Sparsest narrow-band decomposition and refined composite multi-scale dispersion entropy. *Entropy* **2020**, *22*, 375.
23. Jiao, S.; Geng, B.; Li, Y.; Zhang, Q.; Wang, Q. Fluctuation-based reverse dispersion entropy and its applications to signal classification. *Appl. Acoust.* **2021**, *175*, 107857. [[CrossRef](#)]
24. Wang, H.; Sun, W.; He, L.; Zhou, J. Intelligent fault diagnosis method for gear transmission systems based on improved multi-scale reverse dispersion entropy and swarm decomposition. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–13. [[CrossRef](#)]
25. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In Proceedings of the European Conference of the Prognostics and Health Management Society, Bilbao, Spain, 5–8 July 2016; pp. 1–18.
26. Li, Y.; Li, G.; Yang, Y.; Liang, X.; Xu, M. A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy. *Mech. Syst. Signal Process.* **2018**, *105*, 319–337. [[CrossRef](#)]
27. KAT-DataCenter Website of the Chair of Design and Drive Technology, Paderborn University, Germany. 2016. Available online: <http://mb.uni-paderborn.de/kat/datacenter> (accessed on 12 December 2021).
28. Yang, J.; Yang, Y.; Xie, G. Diagnosis of incipient fault based on sliding-scale resampling strategy and improved deep autoencoder. *IEEE Sens. J.* **2020**, *20*, 8336–8348. [[CrossRef](#)]
29. Liu, A.; Yang, Z.; Li, H.; Wang, C.; Liu, X. Intelligent diagnosis of rolling element bearing based on refined composite multiscale reverse dispersion entropy and random forest. *Sensors* **2022**, *22*, 2046. [[CrossRef](#)]
30. Li, Z.; Cui, Y.; Li, L.; Chen, R.; Dong, L.; Du, J. Hierarchical amplitude-aware permutation entropy-based fault feature extraction method for rolling bearings. *Entropy* **2022**, *24*, 310. [[CrossRef](#)] [[PubMed](#)]
31. Ying, W.; Tong, J.; Dong, Z.; Pan, H.; Liu, Q.; Zheng, J. Composite multivariate multi-Scale permutation entropy and laplacian score based fault diagnosis of rolling bearing. *Entropy* **2022**, *24*, 160. [[CrossRef](#)] [[PubMed](#)]