

## Article

# Fault Recognition of Rolling Bearings Based on Parameter Optimized Multi-Scale Permutation Entropy and Gath-Geva

Haiming Wang<sup>1</sup>, Qiang Li<sup>1,\*</sup>, Shaopu Yang<sup>2</sup> and Yongqiang Liu<sup>2</sup>

<sup>1</sup> School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 010044, China; whm@stdu.edu.cn

<sup>2</sup> State Key Laboratory of Mechanical Behavior and System Safety of Traffic Engineering Structures, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; yangsp@stdu.edu.cn (S.Y.); liuyq@stdu.edu.cn (Y.L.)

\* Correspondence: qli3@bjtu.edu.cn

**Abstract:** To extract fault features of rolling bearing vibration signals precisely, a fault diagnosis method based on parameter optimized multi-scale permutation entropy (MPE) and Gath-Geva (GG) clustering is proposed. The method can select the important parameters of MPE method adaptively, overcome the disadvantages of fixed MPE parameters and greatly improve the accuracy of fault identification. Firstly, aiming at the problem of parameter determination and considering the interaction among parameters comprehensively of MPE, taking skewness of MPE as fitness function, the time series length and embedding dimension were optimized respectively by particle swarm optimization (PSO) algorithm. Then the fault features of rolling bearing were extracted by parameter optimized MPE and the standard clustering centers is obtained with GG clustering. Finally, the samples are clustered with the Euclid nearness degree to obtain recognition rate. The validity of the parameter optimization is proved by calculating the partition coefficient and average fuzzy entropy. Compared with unoptimized MPE, the propose method has a higher fault recognition rate.

**Keywords:** rolling bearing; fault recognition; parameter optimized multi-scale permutation entropy; skewness; GG fuzzy clustering



**Citation:** Wang, H.; Li, Q.; Yang, S.; Liu, Y. Fault Recognition of Rolling Bearings Based on Parameter Optimized Multi-Scale Permutation Entropy and Gath-Geva. *Entropy* **2021**, *23*, 1040. <https://doi.org/10.3390/e23081040>

Academic Editor:  
Anne Humeau-Heurtier

Received: 13 July 2021  
Accepted: 10 August 2021  
Published: 13 August 2021

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



**Copyright:** © 2021 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

As the core component of rotating machinery, the state of rolling bearing directly affects the use of the equipment [1]. Vibration signals collected by the sensor are often contaminated by noise and thus unusable for direct machine faults diagnosis [2]. How to identify the state of rolling bearing quickly and effectively has become a focus of current research. Fault feature extraction and pattern recognition are key links in the fault diagnosis of rolling bearing [3,4]. At present, for the non-stationary complex signal, the feature extraction method mainly applies traditional time-frequency analysis [5] and filtering. Its statistical characteristics in time and frequency domain change with time, such as root mean square (RMS) [6], kurtosis [7], and shape factor [8]; however, these indicators will change whether the fault location occurs in the bearing outer ring, bearing inner ring or rolling element when a bearing fails. Relying solely on these eigenvalues cannot effectively distinguish and identify the fault location. Fast Fourier transform (FFT) [9], Wavelet transform [10], and ensemble empirical mode decomposition (EEMD) [11] are commonly used to signal denoising in feature extraction of fault diagnosis. Fault types are determined by comparing current fault features with standard or existing fault features [12,13]. However, due to the factors such as friction, vibration, and load in the process of mechanical operation, the vibration signal of mechanical system often shows nonlinear behavior. Using the method of time-frequency analysis to decompose the signal into stable signal inevitably has some limitations and difficulties [14].

The method of nonlinear analysis can directly extract the fault information hidden in the vibration signal of mechanical system without decomposing the original signal [15]. For the nonlinear complex signal of equipment fault, its complexity is different in different states. The complexity analysis methods commonly used in fault diagnosis are approximate entropy [16], fuzzy entropy [17], sample entropy [18], and permutation entropy [19]. However, the above methods are based on the single scale analysis of time series. Multi scale analysis [20,21] has been applied in the field of mechanical fault diagnosis because it can reflect complex features and obtain more feature information of signals. The Approximate Entropy algorithm is used to detect cracks in a rotating shaft in the reference [22]. Zheng [23] applied the concept of multi-scale fuzzy entropy (MFE) to the fault diagnosis of bearing and achieved good results. However, the calculation amount of MFE is large, and the selection of the number of characteristic parameters lacks a certain standard. The multi-scale sample (MSE) analysis was applied to the turbulence data and successfully captured two important features of the turbulent soap films [24]. However, MSE method is slow in calculation of long data, poor in real-time performance, and is greatly affected by abrupt signal. Aziz et al. [25] put forward the concept of multi-scale permutation entropy (MPE), which is used to measure the complexity and randomness of time series at different scales, to enhance the robustness. It has advantages including simple calculation, strong anti-noise ability, short time required to obtain stable system eigenvalues, and suitable for online monitoring [14]. In references [26–30], multi-scale permutation entropy (MPE) is applied to fault diagnosis of rolling bearing respectively. However, the above research did not study the parameters of the MPE, Table A1 of Appendix A compares the differences between their research contents and that of this paper. Because the result of multi-scale permutation entropy value is affected by its own parameters, if the parameters are set unreasonably, the best processing effect will not be achieved. Reference [31] proposed a method to determine the permutation entropy parameters based on reconstructing the optimal phase space of time series, studied the methods to determine the embedding dimension and delay time, but ignored the length of time series. In reference [14], the length of time series was determined by observing the permutation entropy of Gaussian white noise with different lengths. Although the method can achieve a certain processing effect, the number of given data length values is fixed, and it is difficult to accurately reflect the characteristic information of original signal with Gaussian white noise. Through analysis, the setting delay time  $t$  of MPE has little influence but the embedding dimension  $m$  and the length  $L$  of time series have a great influence on the final processing result. If the parameter setting is unreasonable, the best processing effect will not be achieved, therefore, the parameter influence analysis and optimization of the MPE are studied in this paper.

The pattern recognition can make substantive discrimination to fault types [32]. The selection of recognizer can be based on classification [33] idea or clustering [34] idea. The typical representative of classification idea is SVM [35]. However, the calculation process of SVM needs the participation of relevant existing empirical parameters, mainly including kernel parameters and penalty factors. The selection of these two parameters which play an important role in SVM is usually based on the user's existing experience, which greatly reduces the universality of SVM model. Cluster analysis is also one of the important methods of pattern recognition. This kind of algorithm does not need difficult parameter selection process and is widely used. Fuzzy C-means (FCM) [36], Gustafson-Kessel (GK) [37], and Gath-Geva (GG) [38] are commonly used. Generally, The FCM is only applicable to the data sets with spherical distribution. GK algorithm introduces adaptive distance norm and covariance matrix, which can reflect the dispersion degree of data along any direction or subspace, but does not change the clustering state of clustering algorithm which is similar to sphere [39–42]. GG clustering algorithm is the improved result of FCM and GK clustering algorithm [38], because it measures the distance between samples by introducing fuzzy maximum likelihood estimation method and can reflect the data of different shapes and directions [43,44], which has stronger adaptability.

Based on the above reasons, this paper proposes a method which combines parameter optimized MPE and GG clustering algorithm to extract fault features and recognition pattern of rolling bearing. The effectiveness of the proposed method is verified by several rolling bearing fault experiments.

## 2. The Proposed Method

### 2.1. MPE Theory

The MPE is to calculate the permutation entropy of time series at different scales, that is to consider the characteristics of time series at multi scales. The calculation steps are as follows. For the time series  $X = [x_i, i = 1, 2, \dots, N]$ , the coarse-grained time series  $y_j^{(s)}$  are obtained by coarse-grained processing [25],

$$y_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_i, \quad (j = 1, 2, \dots, [N/s]) \quad (1)$$

where  $s$  is the scale factor of  $X$  and  $N$  is the length of  $X$ .

The phase space of each coarse-grained sequence is reconstructed, the  $l_{\text{th}}$  reconstruction component is

$$Y_l^{(s)} = \{y_l^{(s)}, y_{l+t}^{(s)}, \dots, y_{l+(m-1)t}^{(s)}\}, \quad (l = 1, 2, \dots, N - (m-1)t) \quad (2)$$

where  $m$  is the embedding dimension and  $t$  is the delay time.

By arranging the elements of each reconstruction component in ascending order, a group of corresponding symbol sequence  $symbol(\alpha)$  can be obtained

$$symbol(\alpha) = (j_1, j_2, \dots, j_m), \quad (\alpha = 1, 2, \dots, \gamma; \gamma \leq m!) \quad (3)$$

Calculating the PE of each coarse-grained sequence at different scales, we get the MPE of time series  $X$

$$H_D = - \sum_{\alpha=1}^{\gamma} \rho_{\alpha} \ln \rho_{\alpha} \quad (4)$$

where  $\rho_{\alpha}$  is the probability of each  $symbol(\alpha)$ .

### 2.2. Parameter Selection for MPE

In order to analyze the general trend of a group of data, the first step is to find the mean value. However, the mean value alone cannot fully represent the overall situation of a group of data, so the skewness of the data can be obtained [45]. The smaller the absolute value of skewness is, the more reliable the value is.

The MPE value of  $X$  ( $X = [x_i, i = 1, 2, \dots, N]$ ) from all scales constitutes the sequence  $H_D(X)$

$$H_D(X) = \{H_D(1), H_D(2), \dots, H_D(s)\} \quad (5)$$

The skewness of  $H_D(X)$  is *skew*

$$skew = E[H_D(X) - H_D^{ave}(X)]^3 / [H_D^{std}(X)]^3 \quad (6)$$

where  $H_D^{ave}(X)$  and  $H_D^{std}(X)$  are the average value and standard deviation of the  $H_D(X)$ ,  $E(*)$  stands for expectation.

Therefore, this paper selects the square function of skewness as the objective function [42] to calculate the minimum value and optimize the maximum value of the  $F(X)$

$$F(X) = \frac{1}{skew^2 + 1} \quad (7)$$

### 2.3. Particle Swarm Optimization

Particle swarm optimization (PSO) [46] regards the individuals in the population as particles without mass and volume in the multi-dimensional search space. Each particle has its own position and velocity, in the solution space, the fitness evaluation function is used to continuously aggregate to its personal best historical position  $p_{best}$  and the group best historical position  $g_{best}$  in the whole field to realize the evolution of candidate solutions.

The special memory function of PSO makes it possible to dynamically track the current search situation and adjust its search strategy. The evolution process of particle swarm optimization is as follows

$$\begin{cases} ve_i^{\sigma+1} = w \cdot ve_i^\sigma + c_1 r_1 (p_i^\sigma - po_i^\sigma) + c_2 r_2 (g_i^\sigma - po_i^\sigma) \\ po_i^{\sigma+1} = po_i^\sigma + ve_i^{\sigma+1} \end{cases} \tag{8}$$

where  $\sigma$  is an evolutionary algebra,  $ve_i^\sigma$  is the flight velocity of particle  $i$ ,  $po_i^\sigma$  is the position vector of particle  $i$ ,  $p_i^\sigma$  is the best position experienced by particle  $i$  and  $g_i^\sigma$  is the best position of the whole particle swarm to experience in the solution space.  $r_1$  and  $r_2$  are random numbers between  $[0, 1]$ ,  $c_1$  and  $c_2$  are learning factors,  $w$  is the inertia weight factor. While  $po_i$  and  $ve_i$  meet the following condition,

$$\begin{cases} po_i \in [po_{min}, po_{max}] \\ ve_i \in [ve_{min}, ve_{max}] \\ ve_{max} = \delta po_{max} \end{cases} \tag{9}$$

where  $\delta$  is the proportional coefficient between the maximum velocity  $ve_{max}$  and the maximum search space  $po_{max}$ .

When the position or velocity of a certain dimensional variable exceeds the boundary range, the boundary absorption strategy is adopted, that is, the particle falls on the boundary of the search space in the next iteration.

The parameters of PSO algorithm in this paper are set as follows: population size  $group = 20$ , maximum iterations  $T_{max} = 10$ , acceleration constant  $c_{1,2} = 1.5$ , and inertia weight  $w = 0.5$ . The process of MPE parameter optimization using PSO is shown in Figure 1.

### 2.4. GG Algorithm

The specific algorithm given in the reference [47] is as follows.

(1) Suppose a sample set  $\Omega = (\psi_1, \psi_2, \dots, \psi_k, \dots, \psi_n)$  has a  $z$  ( $2 \leq z \leq n$ ) common class,  $\psi_k = [\psi_{k1}, \psi_{k2}, \dots, \psi_{kd}]$  representing  $d$  features of the  $k_{th}$  ( $1 \leq k \leq n$ ) sample.

(2) Initialize membership matrix  $U = [u_{ik}]_{z \times n}$ ,  $u_{ik}$  is objective function, which indicates the degree of the  $k_{th}$  sample belonging to the  $i_{th}$  ( $1 \leq i \leq z$ ) category.  $V = [v_1, v_2, \dots, v_z]$  is cluster center vector,  $z$  is the number of clusters.

(3) Update cluster center  $v_i$

$$v_i^\lambda = \frac{\sum_{k=1}^n (u_{ik}^{\lambda-1})^\beta \psi_k}{\sum_{k=1}^n (u_{ik}^{\lambda-1})^\beta} \tag{10}$$

where  $\lambda$  is the iterations,  $\beta$  is fuzzy exponent and generally taken as 2.

(4) Calculate the distance measure  $DM_{ik}$

$$DM_{ik} = \frac{(\det(A_i))^{1/2}}{q_i} \exp\left(\frac{1}{2}(\psi_k - v_i^\lambda)^T A_i^{-1}(\psi_k - v_i^\lambda)\right) \tag{11}$$

$$q_i = \frac{1}{n} \sum_{k=1}^n u_{ik} \tag{12}$$

where  $DM_{ik}$  is the maximum likelihood estimation distance and  $A_i$  is the covariance matrix of the  $i$ th cluster center,  $q_i$  is the prior probability of the  $i$ th cluster being selected.

(5) Update membership matrix  $U$

$$u_{ik}^\lambda = \frac{\sum_{k=1}^n u_{ik}^{\lambda-1}}{\sum_{j=1}^z (DM_{ik}(\psi_k, v_i) / DM_{jk}(\psi_k, v_j))^{\frac{2}{\beta-1}}} \tag{13}$$

where if the condition  $\|U^\lambda - U^{\lambda-1}\| < \varepsilon$  ( $\varepsilon$  is the termination tolerance) is satisfied, the operation will be terminated, otherwise,  $\lambda = \lambda + 1$ , until the condition is satisfied.

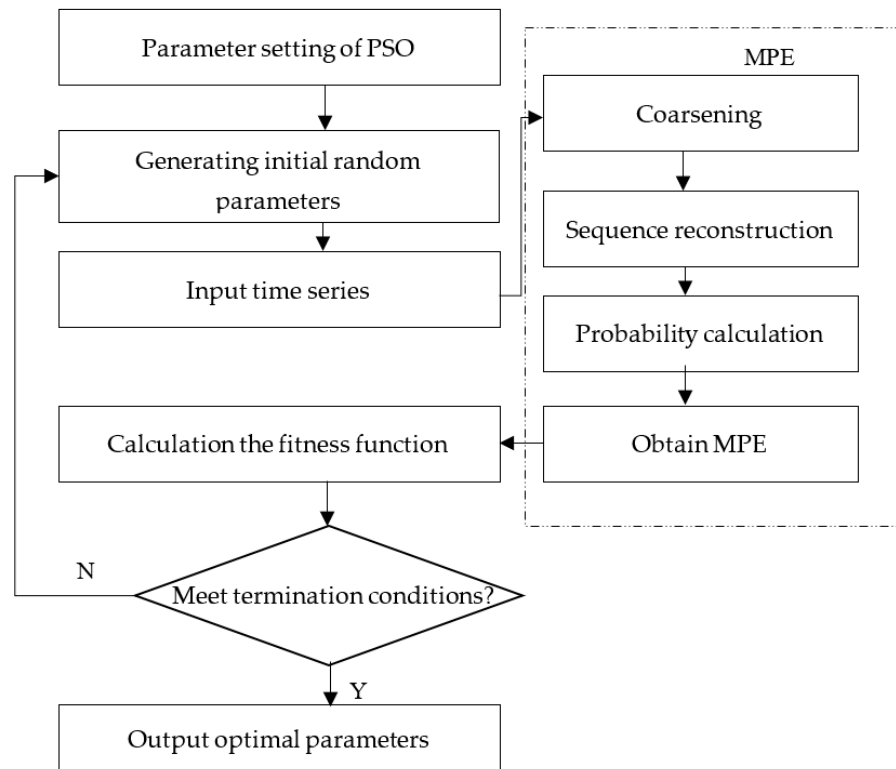


Figure 1. Parameter optimization procedure of MPE based on PSO.

2.5. Evaluation Index of Clustering Effect

The clustering effect of GG fuzzy clustering can be made quantitative assessment with partition coefficient (PAC) [48] and partition entropy (PAE), which are as follows

$$PAC = \frac{1}{n} \sum_{i=1}^z \sum_{k=1}^n u_{ik}^2 \tag{14}$$

$$PAE = - \sum_{i=1}^z \frac{\zeta_i}{n} \left( \sum_{\tau=1}^{\ell} \frac{\zeta_{i\tau}}{\zeta_i} \log_2 \frac{\zeta_{i\tau}}{\zeta_i} \right) \tag{15}$$

where  $\zeta_i, \zeta_{i\tau}$  are the number of all members in cluster  $i$  and the number of members belonging to class  $\tau$ , respectively.  $\ell$  is the number of category from cluster  $i$ .

Suppose the sample set  $\Omega = (\psi_1, \psi_2, \dots, \psi_k, \dots, \psi_n)$  is composed of the sample set called  $\theta$  and the set  $\varphi$ ,  $n$  is the number of samples in  $\Omega$ , Euclidean closeness [49] is used to fault recognition in this paper, then the Euclid closeness between  $\theta$  and  $\varphi$  is

$$Euclid(\theta, \varphi) = 1 - \frac{1}{\sqrt{n}} \sqrt{\sum_{k=1}^n [\theta(\psi_k) - \varphi(\psi_k)]^2} \quad (16)$$

where  $\theta(\psi_k)$  and  $\varphi(\psi_k)$  are membership functions of  $\theta$  and  $\varphi$ , respectively.

### 2.6. The Process of Bearing Fault Pattern Recognition

The framework of the proposed method is shown in Figure 2. The general implementation procedure is summarized as follows

- (1) Carry out the experiment and collect the vibration experiment data.
- (2) For the signal, the initial parameters of MPE are optimized by PSO algorithm. The optimal parameters ( $m, L$ ) of MPE is determined.
- (3) The optimized parameters are reset to MPE, the entropy of signal is calculated by PSO-MPE and the eigenvalue matrix is established.
- (4) Input eigenvalue matrix into GG clustering classifier to realize clustering.
- (5) The Euclidean distance between the samples to be identified and the clustering center is calculated to realize the classification and recognition.

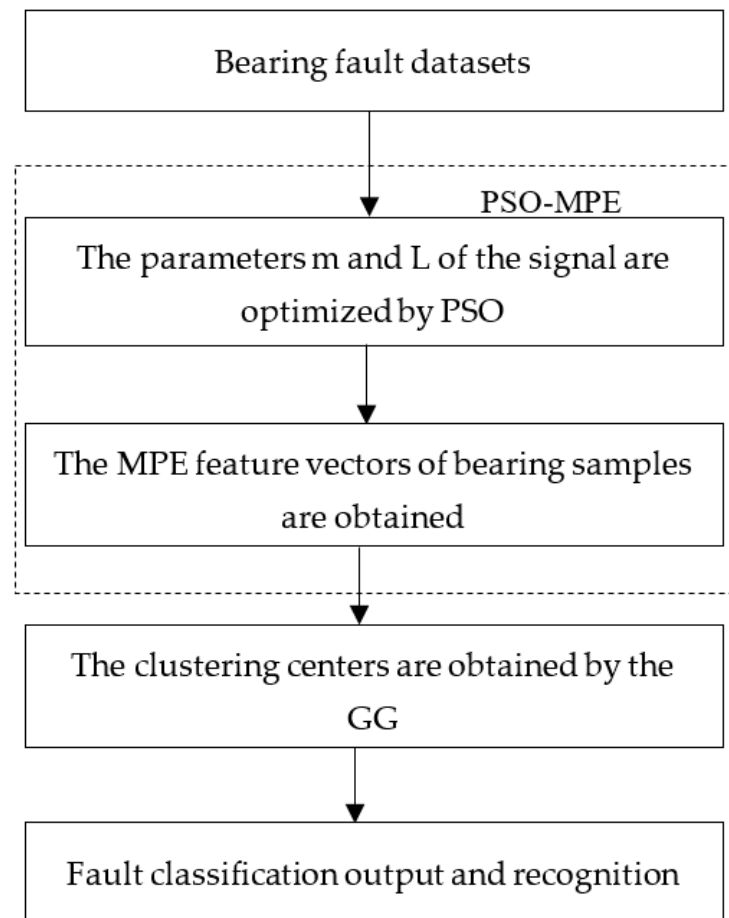
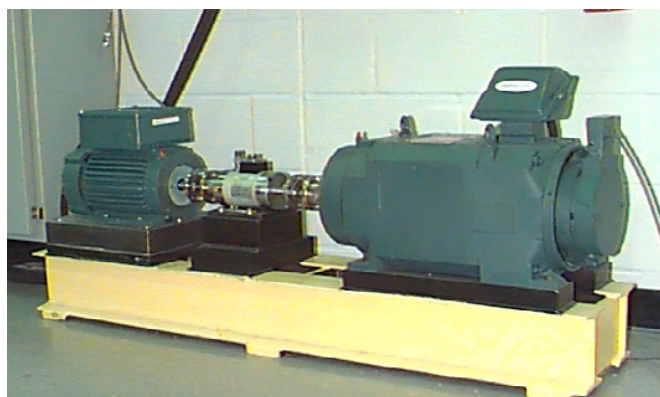


Figure 2. Implementation flow chart of proposed method.

### 3. Parameter Influence Analysis of MPE

In order to study the influence of different parameters on MPE, the experimental data of rolling bearing in Case Western Reserve University [50] is used for analysis. The

test stand is shown in Figure 3, which is composed of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The test bearings support the motor shaft. Single point faults were introduced to the test bearings using electro-discharge machine. Vibration data was collected using accelerometers, which were attached to the housing with magnetic bases. The rolling bearing near the drive end is tested in the experiment. Its type is 6205-2 RSJEMSKF.



**Figure 3.** The test stand of CWRU.

Taking the normal vibration signal of the drive end bearing as an example when the motor speed is 1797 r/min, the sampling frequency is 12 kHz. The values of data length  $L$  are 128–4096, respectively. The values of embedding dimension  $m$  are 3–8, delay time  $t$  is 1–6 and scale factor  $s$  is set from 1 to 12. Figure 4 shows the amplitude variation of MPE from samples in each state under different lengths, different embedding dimensions, and different delay time.

It can be seen from Figure 4 that for the normal vibration signal of the bearing, when  $m = 6$ ,  $t = 1$ , the value of  $L$  changes from small to large, the entropy increases obviously. So different  $L$  values have a greater impact on the entropy, it is necessary to select the appropriate value of  $L$ . Fixed  $L = 1024$ ,  $t = 1$ ,  $m$  values from small to large change, with the increase of  $m$ , the entropy decreases obviously, different  $m$  value has a different entropy, so it is necessary to select the appropriate value of  $m$ .

Fixed  $m = 6$ ,  $L = 1024$ , as can be seen from Figure 4c, with the increase of delay time  $t$ , the entropy value does not increase or decrease obviously at different scales, which indicates that it has little effect on the entropy value, so the fixed value of  $t$  is 1 in this paper. When  $m$  value is too small, the ability of the algorithm to detect signal mutation is low, but the larger  $m$  value is, the larger the amount of calculation is, and the longer the running time of the algorithm is. In summary, selecting the appropriate data length  $L$  and embedding dimension  $m$  is necessary.



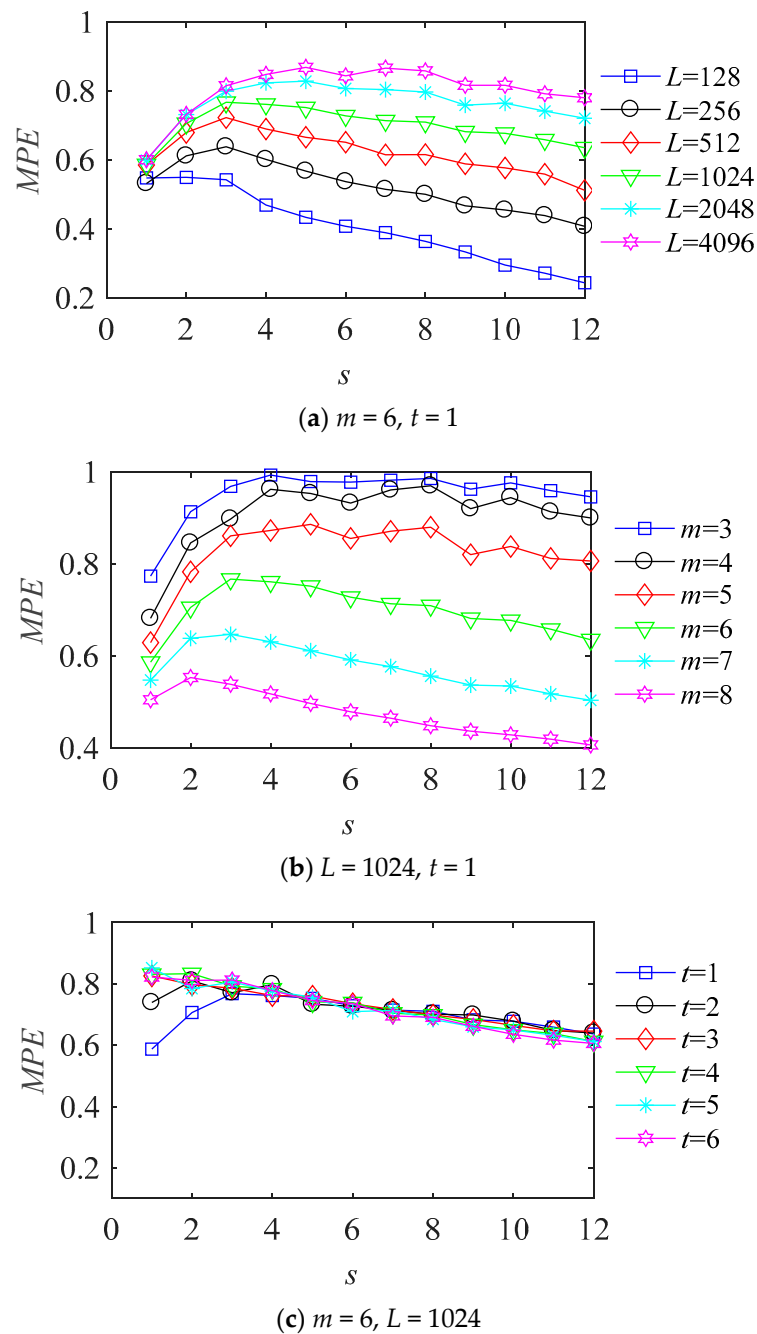


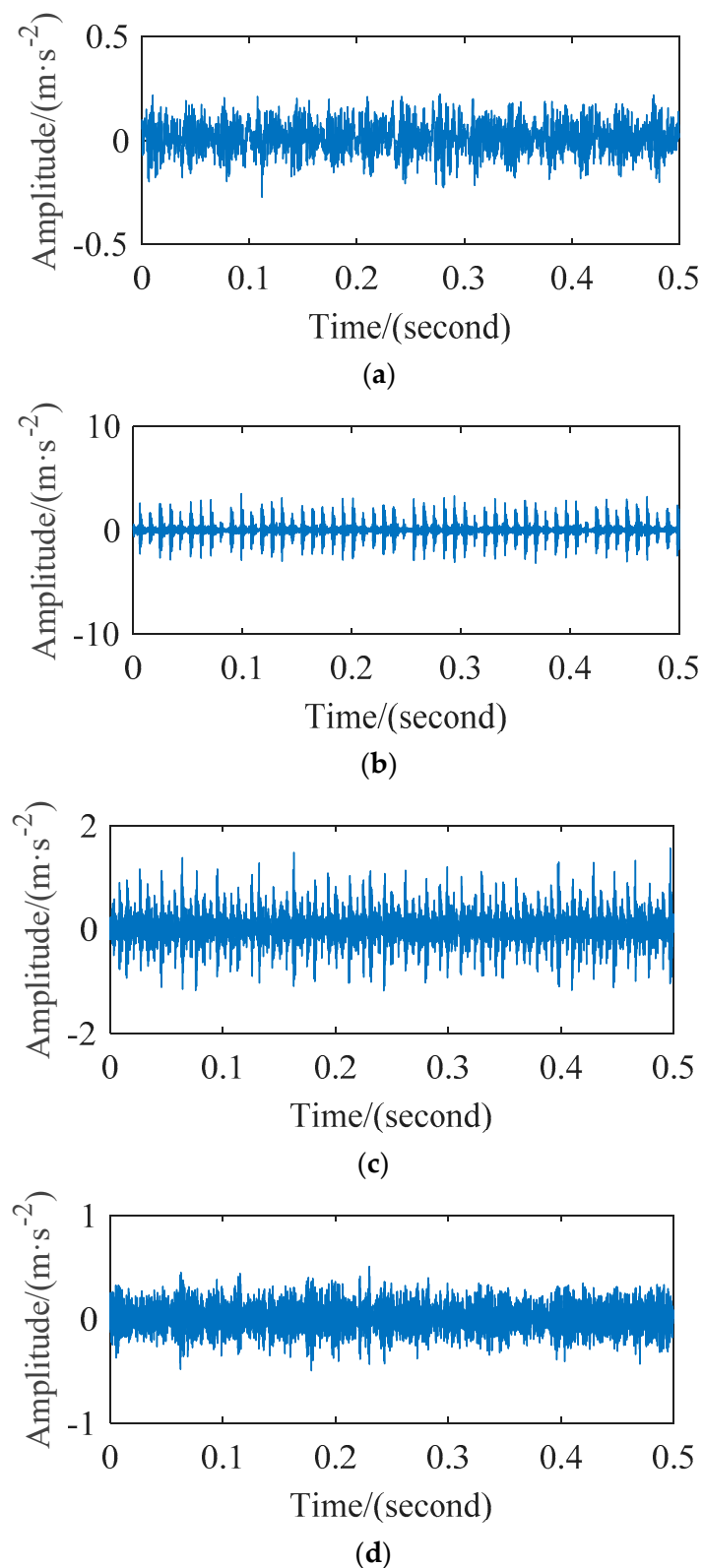
Figure 4. The multi-scale permutation entropy of different parameters.

#### 4. Experimental and Comparative Analysis

##### 4.1. Case 1: CWRU Data Analysis

When the motor speed is 1797 r/min, four types of vibration signals are analyzed, including normal (NR) bearings, outer ring fault (ORF) bearings, inner ring fault (IRF) bearings and ball fault (BF) bearings. Figure 5 shows a part of time waveform of the vibration signal collected by sensors in four states, the horizontal axis is the time, the vertical axis is the acceleration amplitude of the vibration signals, their units are second and  $m \cdot s^{-2}$  respectively. Intercept each state signals from the original signal according to different lengths to obtain four state samples. The number of samples is 30 for each state, a total of 120 feature vectors can be obtained, and each feature vector has 12 dimensions.





**Figure 5.** The waveforms of the vibration signal: (a) normal; (b) outer race fault; (c) inner race fault; (d) ball fault.

Firstly, every sample is analyzed by MPE to extract the features, the effectiveness of parameter optimization of PSO algorithm is verified compared with the parameters in

reference [32]. The vibration signals of four states of the bearing are analyzed, and the change of fitness value in the optimization process is shown in Figure 6.

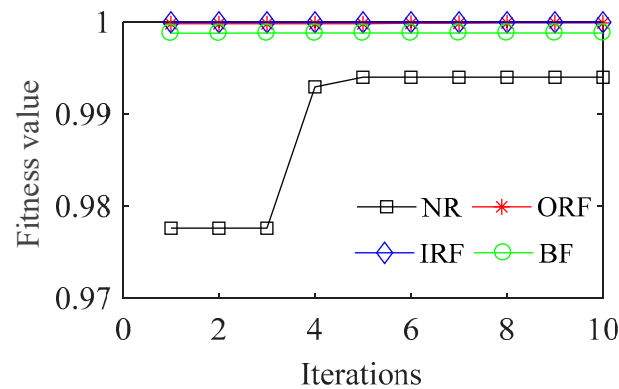
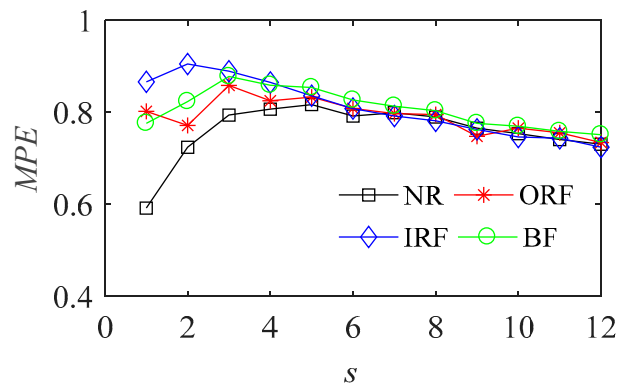
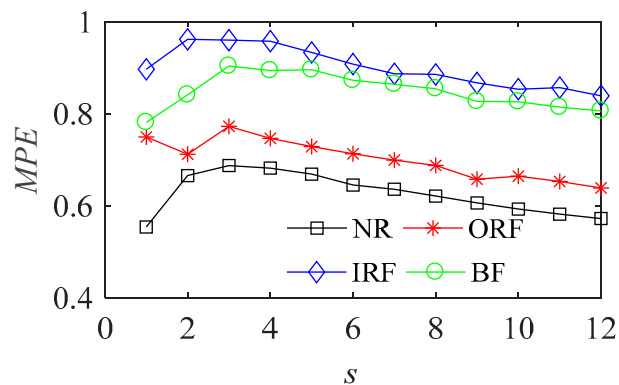


Figure 6. The iterative process.

The optimized parameters of MPE for various state samples are shown in Table 1. The MPE values and cluster results before and after the optimization of vibration signal in four states are shown in the Figures 7 and 8 respectively.

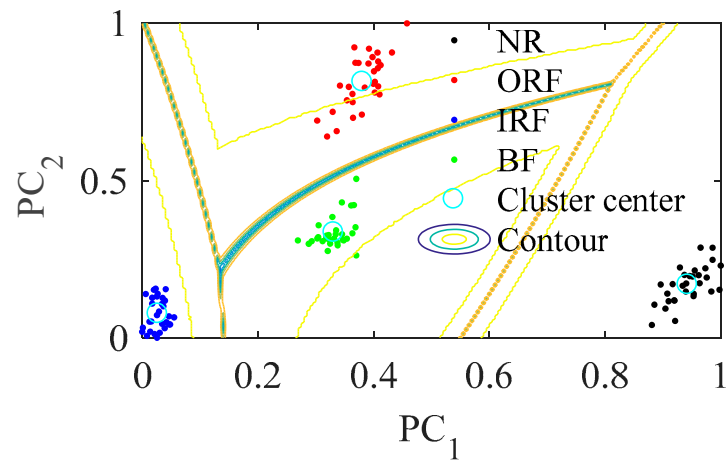


(a) Parameters unoptimized MPE

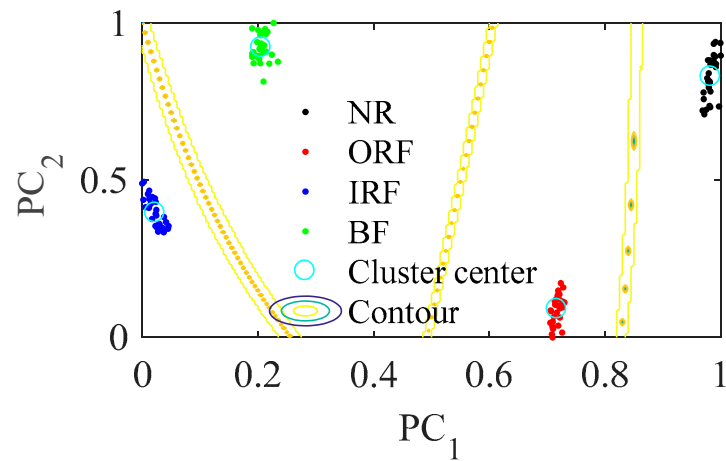


(b) Parameters optimized MPE

Figure 7. The MPE of bearing vibration signals in four states.



(a) GG clustering results of the MPE



(b) GG clustering results of the PSO-MPE

Figure 8. The clustering results of bearing vibration signals in four states.

Table 1. The parameters of MPE.

Fault Types	Parameters of MPE from Reference [32]		Parameters of MPE with PSO	
	<i>m</i>	<i>L</i>	<i>m</i>	<i>L</i>
NR	6	2048	7	1732
ORF	6	2048	7	3152
IRF	6	2048	5	1559
BF	6	2048	6	3251

It can be seen from Figure 7b that the MPE with optimized parameters can better distinguish the four different states of bearings, and is better than the effect of fixed parameters in Figure 7a. The parameter optimized MPE can distinguish the four states of the bearing more obviously, it can be used as the feature vector to further classify and identify the bearing fault modes.

In Figure 8,  $PC_1$  and  $PC_2$  are two vectors in two-dimensional space after data visualization, they have the same meaning in the following similar figures. As can be seen from Figure 8b, the samples are distributed around four clustering centers according to fault types after processed by the proposed method, the distance between different classes becomes larger and the distance within classes becomes smaller compared with Figure 8a.

In order to further illustrate the effectiveness of this research method, the *PAC*, *PAE*, and fault recognition rate are used to evaluate quantitatively. Corresponding to Figure 8, the performance comparison of the two recognition methods is shown in Table 2. It can be seen that

- (1) The closer the *PAC* is to 0, the better the clustering effect. Although the *PAC* value of MPE and PSO-MPE are all 1, the *PAE* decreases gradually. The closer the *PAE* is to 0, the better the clustering effect.
- (2) The fault recognition rate of PSO-MPE with GG clustering reaches 100%, which is consistent with its clustering performance.
- (3) It can be seen that the PSO-MPE method proposed by the author can effectively extract the fault feature information of rolling bearing and accurately identify different fault types of rolling bearings.

**Table 2.** Performance comparison of two recognition methods.

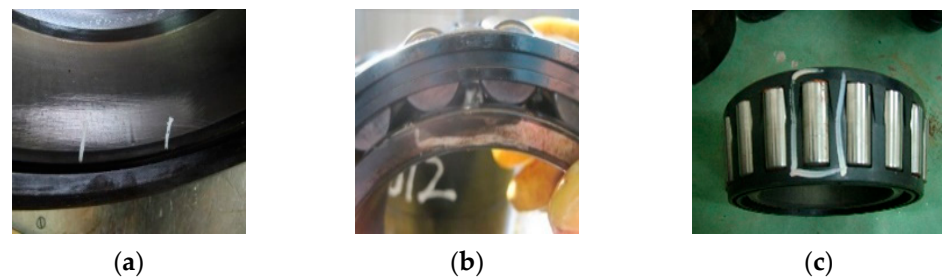
Classifier	Evaluating Indicators	MPE	PSO-MPE
GG	<i>PAC</i>	1	1
	<i>PAE</i>	0.2624	0
	Recognition rate	84.17%	100%

#### 4.2. Case 2: A Freight Locomotive Wheelset Bearing Signal

To further demonstrate the performance of the proposed method, a fault experiment is carried out in this section. The experimental setup and the tested wheelset bearing are shown in Figure 9. RD2 wheel set and 197,726 double row tapered roller bearing are installed on the test bench. The fault bearings are shown in Figure 10. The wheelset bearing deflections are natural damages generated during the operation of the railway freight vehicles, which are located in the outer raceway, inner raceway and ball, respectively. The experimental device includes three DASP data processing software of CA-YD-188 piezoelectric accelerometer, signal amplifier and INV36DF signal acquisition instrument. The sensors are installed on the test bench in turn, and the position is shown in Figure 9. The sampling frequency is 25.6 kHz.

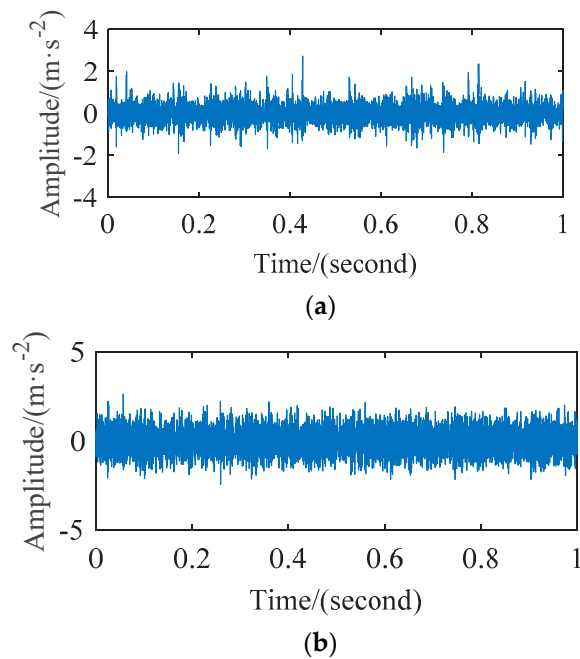


**Figure 9.** Experimental setup.



**Figure 10.** The fault locations: (a) The slight dent fault on the outer raceway; (b) a fatigue spall fault on the inner raceway; (c) the scratch fault on the rolling ball.

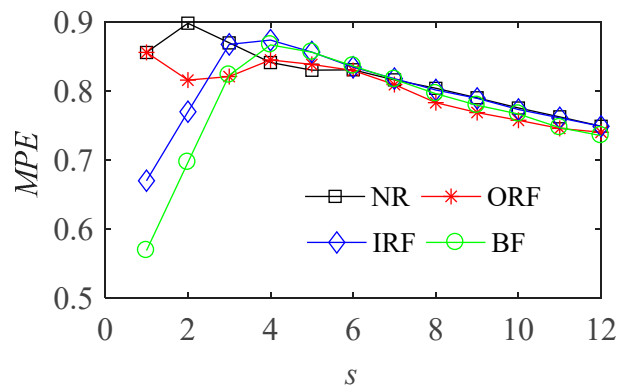
In order to observe the time domain characteristics and save the paper space, Figure 11 shows the time domain waveform of the bearing inner ring and rolling ball. We can see the noise component of the collected signal from this experiment in the Figure 11 is more than bearing experiment of CWRU in the Figure 5, which increases difficulty of the method verification.



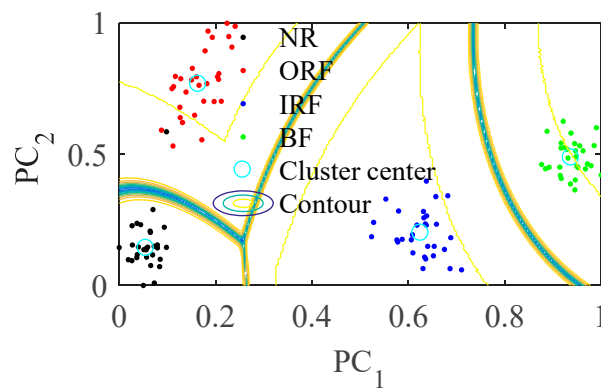
**Figure 11.** The time-domain waveforms: (a) Inner raceway fault; (b) ball fault.

There are 30 group samples collected in each state. It can be seen from Figure 12 that without optimizing the parameters of the MPE, the entropy values of the four states of rolling bearing are intertwined, they are not effectively distinguished, which cannot effectively distinguish the four states, it is not suitable to use them as the quantitative features of rolling bearing fault.

GG with parameters unoptimized MPE is directly used for the signal. As shown in Figure 12a, the entropy value of the four states is not effectively distinguished. The sample distance of the same class is too large, and the distance between different classes is small in Figure 12b. Although we can see about four gathering teams, the distinction between NR and ORF is not obvious, some NR samples are wrongly classified into ORF, when the signal contains more noise components, it is easy to misjudge.



(a) Parameters unoptimized MPE of four state signals



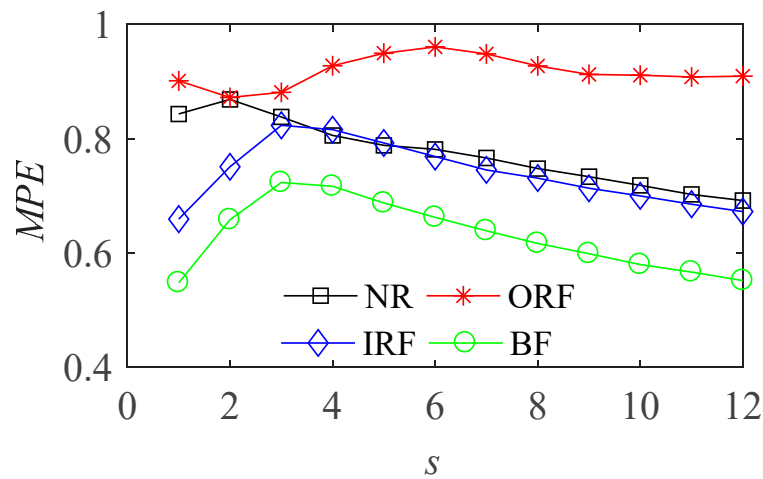
(b) GG clustering results of the parameters unoptimized MPE

Figure 12. The results with MPE.

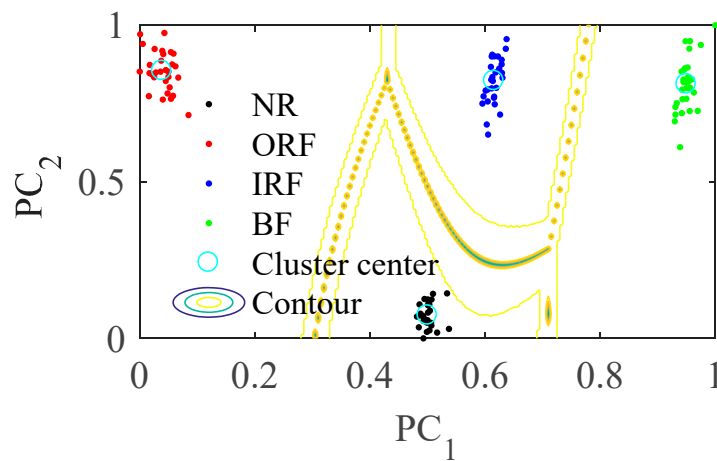
The Table 3 are the parameters of MPE in various states obtained by PSO. Figure 13a shows the PSO-MPE of four state signals, it can be seen that distance between the entropy curves of different operation states is significantly increased and entropy curves of different operation states are obviously separated completely. This is because when the rolling bearing has faults, the randomness of vibration signal changes, which changes the entropy values in different scales. In the same state, with the increase of scale, the randomness and complexity of coarse-grained sequence decrease, and the change range of entropy decreases.

Table 3. The PSO-MPE parameters.

Fault Types	<i>m</i>	<i>L</i>
NR	6	1339
ORF	5	3180
IRF	6	1168
BF	6	541



(a) Parameters optimized MPE of four state signals



(b) GG clustering results of the PSO-MPE

Figure 13. The results with PSO-MPE.

As can be seen from Figure 13b, after samples are processed by PSO-MPE and GG clustering algorithm, they are distributed around four clustering centers according to fault types, the distance between different classes becomes larger and the distance within class becomes smaller than Figure 13a.

According to Table 4, The fault recognition rate of rolling bearing based on PSO-MPE and GG clustering is 99.17%, which is higher than the recognition rate of MPE. Moreover, the PAC and PAE are better than those of parameters unoptimized MPE. It shows that the proposed method is still effective under relatively difficult experimental conditions.

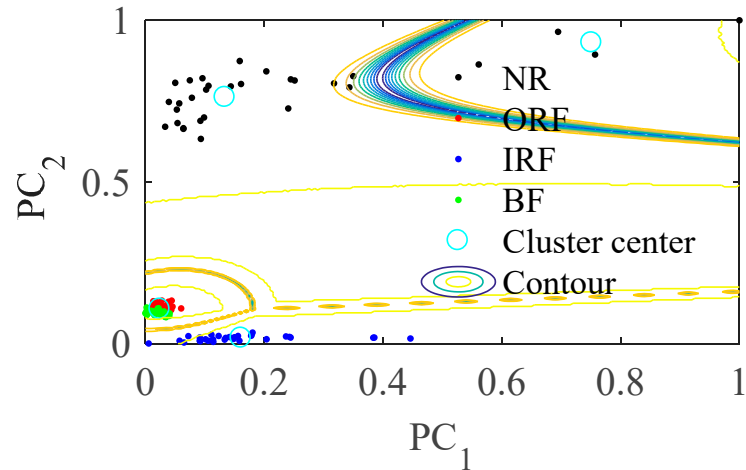
Table 4. Performance comparison of two recognition methods.

Classifier	Evaluating Indicators	MPE	PSO-MPE
GG	PAC	0.9999	1
	PAE	0.3658	0.0409
	Recognition rate	81.67%	99.17%

In order to prove the superiority of parameter optimized MPE as signal feature extraction index, compare it with the feature vector composed of kurtosis and root mean square. Figure 14 shows the effect of clustering with kurtosis and root mean square (RMS) as feature vector. Compared with the Figure 13b, it is obvious that the four types of samples are not effectively distinguished, because these indexes will change no matter which part



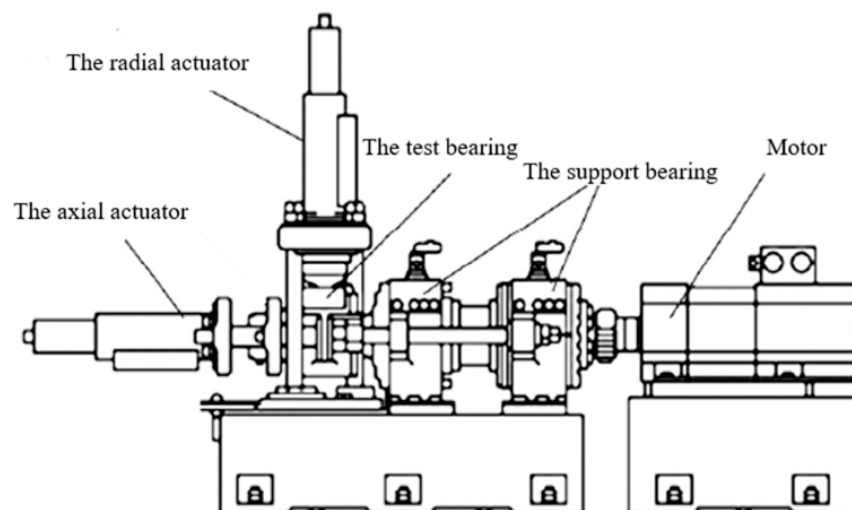
of the bearing fails, it cannot effectively distinguish the fault location only through kurtosis or root mean square. While the research method in this paper can effectively distinguish different types of fault samples.



**Figure 14.** The GG clustering results with Kurtosis and RMS.

#### 4.3. Case 3: A High-Speed Locomotive Wheelset Bearing Fault Signal

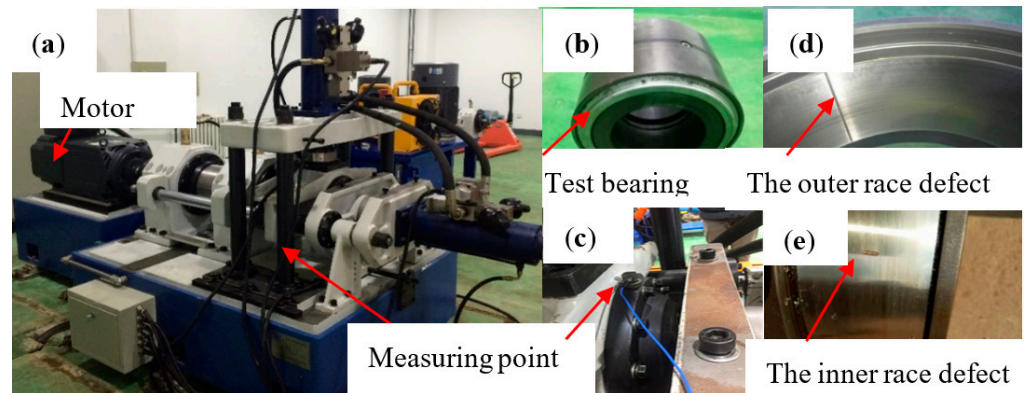
In order to verify whether the method is still effective in more complex working conditions with more noise components, the practical test data from the self-made experiment platform is selected for subsequent analysis. In this case, the vibration signal has been collected from a high-speed locomotive wheelset bearing. The test rig structure [51] is depicted in Figure 15. In order to simulate the load change of wheel set bearing during operation, apply a random force with a frequency of 0.2~20 Hz and an average value of about 10 kN in the radial direction, a simple harmonic force with a frequency of 1 Hz and a maximum value of 10 kN is applied axially on the test rig.



**Figure 15.** The structure of the test rig.

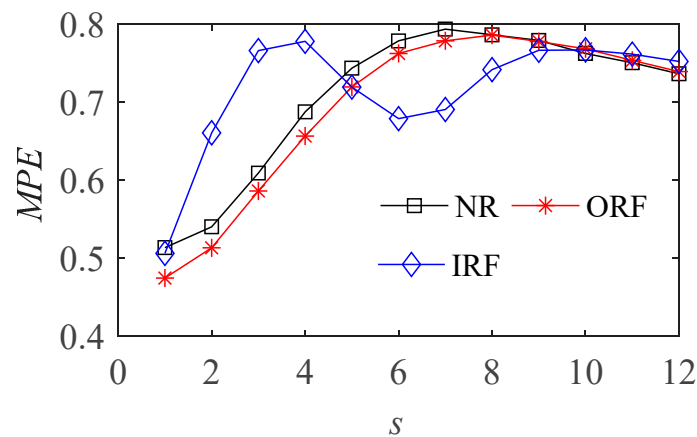
The field diagram of the test rig and the test bearings are depicted in Figure 16. The sensor is located at the top of the end-shield of the test bearing in Figure 16c and the vibration signal is collected by a PCB356A25 accelerometer. The dynamic loads can be obtained by the radial and axial actuators. There is an artificial local defect in the outer race of test bearing as plotted in Figure 16d, of which the width is 1 mm and length is

5 mm. It can be noted that the artificial defect is relatively slight in comparison with its geometries. The sampling frequency is set as 51.2 kHz and the set speed is 2000 r/min.

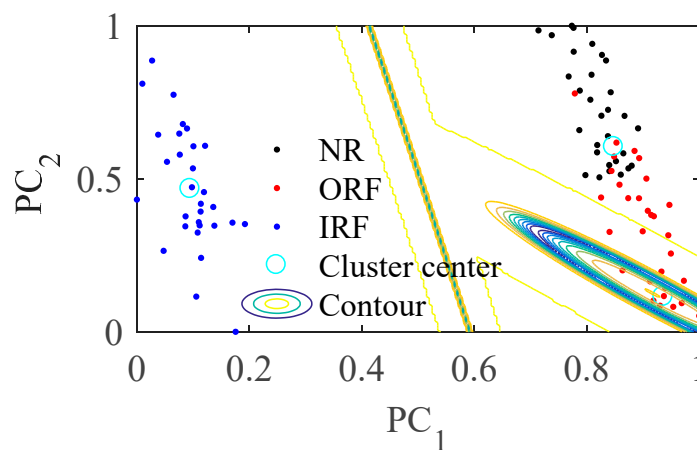


**Figure 16.** Railway bearing test bench and bearings: (a) Railway bearing test bench overview; (b) test bearing; (c) local enlarged test point; (d) partial enlarged view of outer race fault; (e) partial enlarged view of inner race fault.

Each state collected 30 samples. There were total three kinds of normal signals and vibration signals at different fault positions. The results solved by the MPE are shown in the Figure 17 when the  $L = 2048, m = 6$ .



(a) Parameters unoptimized MPE of four state signals



(b) GG clustering results based on MPE

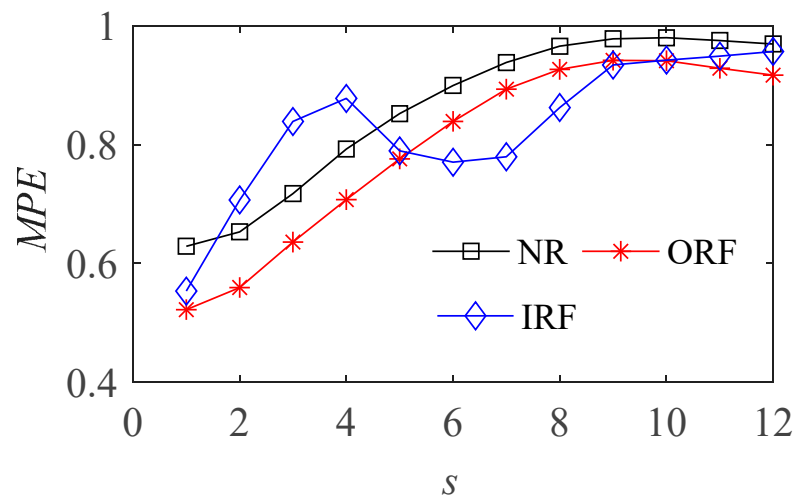
**Figure 17.** The results with MPE.

As can be seen from Figure 17, GG clustering cannot effectively cluster the fault feature samples constructed by MPE. It is difficult for the entropy to represent the different running states of bearings so further treatment is necessary. The steps are the same as last section, it will not be repeated here.

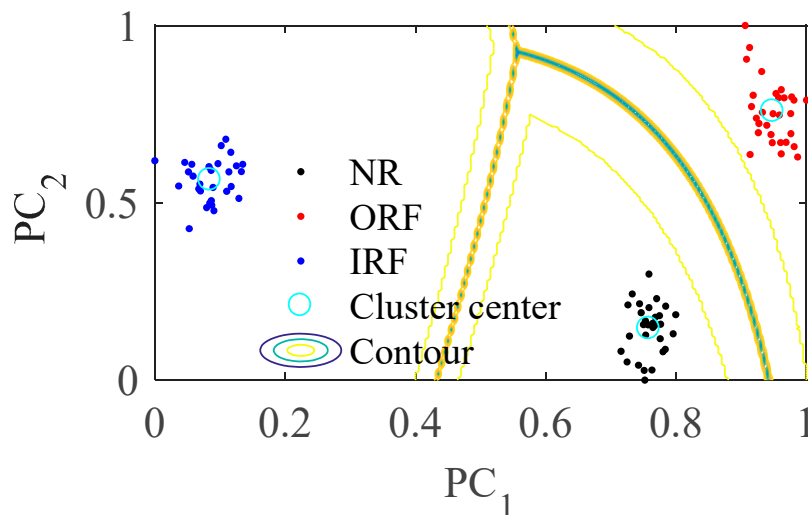
The Table 5 are the parameters of MPE in various states, which are obtained by PSO algorithm. Then GG is used to cluster the samples. The results with PSO-MPE are show in Figure 18.

Table 5. The PSO-MPE parameters.

Fault Types	$m$	$L$
NR	4	3818
ORF	5	3745
IRF	5	3918



(a) Parameters optimized MPE of four state signals



(b) GG clustering results based on PSO-MPE

Figure 18. The results with PSO-MPE.

It can be seen from Figure 18a, the proposed method can effectively distinguish the three states. The values of NR and ORF is obviously separated while they are not in Figure 17a. Compared with Figure 17b, the samples of each state in Figure 18b are obviously

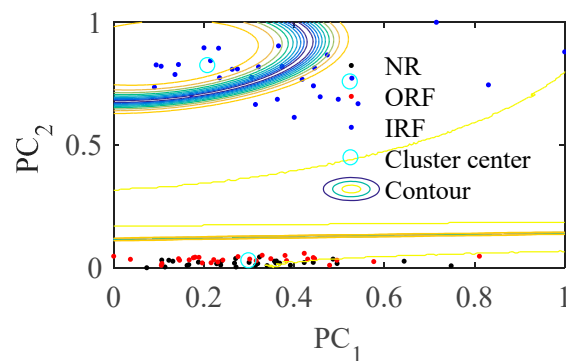
separated, classified with its own cluster centers and the distance between different classes becomes larger and the distance within classes becomes smaller, respectively.

According to Table 6, The fault recognition rate of rolling bearing based on the proposed method is 100%, which improves a lot than the 78.89% of the MPE. Moreover, the *PAC* and *PAE* of PSO-MPE are better than those of MPE, which prove the necessity and advantage of combination PSO-MPE and GG method, it has better clustering effect and recognition effect.

**Table 6.** Performance comparison of two recognition methods.

Classifier	Evaluating Indicators	MPE	PSO-MPE
GG	<i>PAC</i>	0.9841	1
	<i>PAE</i>	0.2720	0
	Recognition rate	78.89%	100%

In order to prove the robustness advantage of the proposed method, compare it with the feature vector composed of kurtosis and root mean square (RMS). Figure 19 shows the clustering effect of kurtosis and root mean square as feature vector. Compared with the Figure 18b, it is obvious that the three types of samples are not effectively distinguished, because the experimental environment simulates the working condition of high-speed train operation, the collected signal is close to the vibration signal of the train running on the actual line and is seriously disturbed by environmental noise. This index is almost invalid in this case.



**Figure 19.** The GG clustering results with Kurtosis and RMS.

## 5. Conclusions

In this paper, a rolling bearing fault detection method based on the PSO-MPE and GG is proposed. The method can select the important parameters of MPE method adaptively, overcome the disadvantages of fixed MPE parameters and greatly improve the accuracy of fault identification. The method is verified by several experiments. Some conclusions are obtained as follows:

(1) To solve the problem of parameter determination of MPE, fitness function is constructed by skewness of multi-scale permutation entropy, the time series length  $L$  and embedding dimension  $m$  are optimized, the effectiveness of the optimization method is verified by experiments.

(2) Compared with the MPE of fixed parameters, it is proved that parameter optimized MPE can extract fault features accurately and has better classification and recognition rate about the rolling bearing typical faults.

(3) The effectiveness and robustness of the proposed method is verified by several rolling bearing experiments, of which the signals are simple to complex. Meanwhile, compared with the feature vector composed of root mean square and kurtosis, the proposed method shows advantages when the vibration signal contains more noise components and

serious environmental interference, the proposed method has more accurate and stable performance in fault diagnosis.

**Author Contributions:** Conceptualization: H.W.; data curation: Q.L. and Y.L.; investigation: Q.L. and Y.L.; methodology: H.W.; software: S.Y.; validation: Y.L.; writing—original draft: H.W.; writing—review and editing: Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** The present work is supported by the National Natural Science Foundation of China (Nos. 11790282, 12032017, 11802184, 11902205 and 12002221), S&T Program of Hebei (20310803D), and Natural Science Foundation of Hebei Province (No A2020210028).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data are not publicly available due to laboratory restrictions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Similarities and Differences between This Paper and Other References.

References	Similarities	Differences
Reference [26]: LMD-MPE		The local mean decomposition (LMD) is used to denoise, then calculate MPE of the product functions, SVM is used to classify. LMD-MPE and LMD-MPE are compared to prove the advantages of MPE
Reference [27]: VMD-MPE-FCM		The variational mode decomposition (VMD) is used to denoise and obtain intrinsic mode functions (IMF), calculate MPE of these IMFs, FCM is used to cluster. The parameters influence of MPE is not considered.
Reference [28]: MPE-LS-SVM		The MPE of the bearing vibration signal in different scales is calculated, the Laplacian score (LS) is used to refine the feature vector, SVM is used to classify. The parameters of MPE is fixed.
Reference [29]: EEMD-MPE-SA-SVM	The MPE of vibration signal is extracted as the feature for fault subsequent classification or clustering	A number of intrinsic mode functions (IMFs) are obtained using ensemble empirical mode decompose (EEMD), the multi-scale IMF permutation entropy are extracted, SA-SVM is used to classify. The parameters ( $m, t$ ) of MPE is fixed.
Reference [30]: CEEMD-MPE-GK		Complementary Ensemble Empirical Mode Decomposition (CEEMD) is used to denoise and obtain intrinsic mode functions (IMF), calculate MPE of the modal IMF, the GK is used to fault type recognition. The parameters influence of MPE is not considered.
This paper: PSO-MPE-GG		The embedding dimension $m$ , delay time $t$ and the length $L$ of time series sample are considered when calculate MPE, PSO and skewness are used to determine the selection of parameters, which increases the robustness of the MPE. GG is used for cluster of fault types, which increases the accuracy.

## References

- Wang, L.; Liu, Z. An improved local characteristic-scale decomposition to restrict end effects, mode mixing and its application to extract incipient bearing fault signal. *Mech. Syst. Signal Process.* **2021**, *156*, 107657. [\[CrossRef\]](#)
- Saleem, R.; Hassan, E.; Kashif, J.; Tufail, S. Vibration feature extraction and analysis for fault diagnosis of rotating machinery-A Literature Survey. *Asia Pac. J. Multidiscip. Res.* **2017**, *5*, 103–110.

3. Yang, S.; Gu, X.; Liu, Y.; Hao, R.; Li, S. A general multi-objective optimized wavelet filter and its applications in fault diagnosis of wheelset bearings. *Mech. Syst. Signal Process.* **2020**, *145*, 106914. [[CrossRef](#)]
4. Yuan, H.; Wu, N.; Chen, X.; Wang, Y. Fault diagnosis of rolling bearing based on shift invariant sparse feature and optimized support vector machine. *Machines* **2021**, *9*, 98. [[CrossRef](#)]
5. Caesarendra, W.; Tjahjowidodo, T. A review of feature extraction Methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines* **2017**, *5*, 21. [[CrossRef](#)]
6. Shen, K.; He, Z.; Chen, X.; Sun, C.; Liu, Z. A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time. *Sensors* **2012**, *12*, 10109–10135. [[CrossRef](#)]
7. Eftekharnajad, B.; Carrasco, M.R.; Charnley, B.; Mba, D. The application of spectral kurtosis on Acoustic Emission and vibrations from a defective bearing. *Mech. Syst. Signal Process.* **2011**, *25*, 266–284. [[CrossRef](#)]
8. Yiakopoulos, C.T.; Gryllias, K.C.; Antoniadis, I.A. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Syst. Appl.* **2011**, *38*, 2888–2911. [[CrossRef](#)]
9. Caesarendra, W.; Kosasih, B.; Tieu, A.K.; Zhu, H.; Moodie, C.A.; Zhu, Q. Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing. *Mech. Syst. Signal Process.* **2016**, *72–73*, 134–159. [[CrossRef](#)]
10. Stazewski, W.J.; Tomlinson, G.R. Application of the wavelet transform to fault detection in a spur gear. *Mech. Syst. Signal Process.* **1994**, *8*, 289–307. [[CrossRef](#)]
11. Žvokelj, M.; Zupan, S.; Prebil, I. Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis. *Mech. Syst. Signal Process.* **2010**, *24*, 1049–1067. [[CrossRef](#)]
12. Zhang, Y.; Zhang, P.; Wang, H. Rolling bearing early fault intelligence recognition based on weak fault feature enhancement in time-time domain. *J. Mech. Eng.* **2016**, *52*, 96–103. [[CrossRef](#)]
13. Zhao, H.; Zuo, S.; Hou, M. A novel adaptive signal processing method based on enhanced empirical wavelet transform technology. *Sensors* **2018**, *18*, 3323. [[CrossRef](#)]
14. Zheng, J.; Cheng, J.; Yang, Y. Multi-scale permutation entropy and its applications to rolling bearing fault diagnosis. *Chin. Mech. Eng.* **2013**, *24*, 2641–2645.
15. Yan, R.; Liu, Y.; Gao, R. Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. *Mech. Syst. Signal Process.* **2012**, *29*, 474–484. [[CrossRef](#)]
16. Canales, D.P.; Ramirez, J.A.; Carlos, J. Identification of dynamic instabilities in machining process using the approximate entropy method. *Int. J. Mach. Tools Man.* **2011**, *51*, 556–564. [[CrossRef](#)]
17. Alkan, Ö.; Albayrak, Ö.; Karadağ, Z. Ranking of renewable energy sources for regions in Turkey by fuzzy entropy based fuzzy copras and fuzzy multimoora. *Renew. Energy* **2020**, *162*, 712–726. [[CrossRef](#)]
18. Aktaruzzaman, M.; Sassi, R. Parametric estimation of sample entropy in heart rate variability analysis. *Biomed. Signal Process. Cont.* **2014**, *14*, 141–147. [[CrossRef](#)]
19. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)]
20. Minhas, A.S.; Singh, G.; Singh, J.; Kankar, P.K.; Singh, S. A novel method to classify bearing faults by integrating standard deviation to refined composite multi-scale fuzzy entropy. *Measurement* **2020**, *154*, 107441. [[CrossRef](#)]
21. Li, Y.; Feng, K.; Liang, X.; Zuo, M. A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy. *J. Sound Vib.* **2019**, *439*, 271–286. [[CrossRef](#)]
22. Sampaio, D.L.; Nicoletti, R. Detection of cracks in shafts with the approximated entropy algorithm. *Mech. Syst. Signal Process.* **2019**, *127*, 286–302. [[CrossRef](#)]
23. Zheng, J.; Chen, M.; Cheng, J.; Yang, Y. Multiscale fuzzy entropy and its application in rolling bearing fault diagnosis. *J. Vib. Eng.* **2014**, *27*, 145–151.
24. Kim, I. Multiscale sample entropy of two-dimensional decaying turbulence. *Entropy* **2021**, *23*, 245. [[CrossRef](#)] [[PubMed](#)]
25. Aziz, W.; Arif, M. Multiscale permutation entropy of physiological times series. In Proceedings of the 2005 Pakistan Section Multitopic Conference, Karachi, Pakistan, 24–25 December 2005; pp. 1–6.
26. Yasir, M.N.; Koh, B.H. Data Decomposition techniques with multi-scale permutation entropy calculations for bearing fault diagnosis. *Sensors* **2018**, *18*, 1278. [[CrossRef](#)]
27. Zhou, C.; Wu, J.; Yuan, X. A Fault diagnosis method combining with quadratic VMD screening, MPE and FCM. *Mech. Sci. Tech. Aerosp. Eng.* **2019**, *38*, 8.
28. Zheng, J.; Cheng, J.; Yu, Y. Multiscale permutation entropy based rolling bearing fault diagnosis. *Shock Vib.* **2014**, *1*, 1–8. [[CrossRef](#)]
29. Yao, D.; Yang, J.; Cheng, X. Railway rolling bearing fault diagnosis based on multi-scale IMF permutation entropy and SA-SVM classifier. *J. Mech. Eng.* **2018**, *54*, 168–176. [[CrossRef](#)]
30. Zhao, R.; Sun, Z. Method of fault identification based on fusion of CEEMD-MPE and GK fuzzy clustering. *J. Vib. Eng.* **2020**, *33*, 629–635.
31. Rao, G. Method for optimal determination of parameters in permutation entropy algorithm. *J. Vib. Shock.* **2014**, *33*, 188–193.
32. Wang, Y.; Li, H.; Ye, P. Fault identification of hydraulic pump based on multi-scale permutation entropy. *China Mech. Eng.* **2015**, *26*, 518–523.



33. Shao, K.; Fu, W.; Tan, J.; Wang, K. Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing. *Measurement* **2021**, *173*, 108580. [[CrossRef](#)]
34. Zhang, H.; Chen, X.; Zhang, X.; Ye, B.; Wang, X. Aero-engine bearing fault detection: A clustering low-rank approach. *Mech. Syst. Signal Process.* **2020**, *138*, 106529. [[CrossRef](#)]
35. Fu, W.; Wang, K.; Zhang, C.; Tan, J. A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine. *Trans. Inst. Meas. Control.* **2019**, *41*, 4436–4449. [[CrossRef](#)]
36. Albert, P.; Nanjappan, M. An efficient kernel FCM and artificial fish swarm optimization-based optimal resource allocation in Cloud. *J. Circuit Syst. Comp.* **2020**, *29*, 2050253. [[CrossRef](#)]
37. Krishnapuram, R.; Kim, J. A note on the Gustafson-Kessel and adaptive fuzzy clustering algorithms. *IEEE Trans. Fuzzy Syst.* **1999**, *7*, 453–461. [[CrossRef](#)]
38. Hou, M.; Hu, X. Roller bearings faults diagnosis method based on VMD and BSE and GG clustering. *J. World Sci. Res.* **2020**, *6*, 231–245.
39. Davari, A.; Marhaban, M.H.; Noor, S. Parameter estimation of K-distributed sea clutter based on fuzzy inference and Gustafson-Kessel clustering. *Fuzzy Set. Syst.* **2010**, *163*, 45–53. [[CrossRef](#)]
40. Wang, S.; Li, L.; Zhang, S. Mechanical fault diagnosis method based on EEMD sample entropy and GK fuzzy clustering. *Chin. Mech. Eng.* **2013**, *24*, 3036–3040.
41. Yang, Z.; Fan, L.; Yang, C. Identification of partial discharge in gas insulated switchgears based on GK fuzzy clustering and LS-SVM. *Power Syst. Prot. Control* **2014**, *42*, 38–45.
42. Chen, D.; Zhang, Y.; Yao, C. Fault diagnosis based on FVMD multi-scale permutation entropy and GK fuzzy clustering. *J. Mech. Eng.* **2018**, *54*, 16–27. [[CrossRef](#)]
43. Silva, S.D.; Junior, M.D.; Junior, V.L.; Brennan, M.J. Structural damage detection by fuzzy clustering. *Mech. Syst. Signal Process.* **2008**, *22*, 1636–1649. [[CrossRef](#)]
44. Zhang, L.; Li, P.; Li, M. Fault diagnosis of rolling bearing based on ITD fuzzy entropy and GG clustering. *Chin. J. Sci. Instr.* **2014**, *35*, 2624–2632.
45. Zhang, L.; Huang, W.; Xiong, G. Assessment of rolling element bearing fault severity using multi-scale entropy. *J. Vib. Shock.* **2014**, *33*, 185–189.
46. Chen, C.-Y.; Ye, F. Particle swarm optimization algorithm and its application to clustering analysis. In Proceedings of the 17th Conference on Electrical Power Distribution, Tehran, Iran, 2–3 May 2021; pp. 789–794.
47. Zhang, S.; Bao, H.; Li, P.; Li, X.; Jiang, W. Fault diagnosis of rolling bearings based on RQA and GG clustering. *Chin. Mech. Eng.* **2015**, *26*, 1385–1390.
48. Song, X.; Shi, M.; Wu, J.; Sun, W. A new fuzzy c-means clustering-based time series segmentation approach and its application on tunnel boring machine analysis. *Mech. Syst. Signal Process.* **2019**, *133*, 106279. [[CrossRef](#)]
49. Giannella, C.R. Instability results for Euclidean distance, nearest neighbor search on high dimensional Gaussian data. *Inform. Process. Lett.* **2021**, *169*, 106–115. [[CrossRef](#)]
50. Loparo, K. Bearings Vibration Data Set, Case Western Reserve University. Available online: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 10 August 2021).
51. Liu, W.; Yang, S.; Li, Q.; Liu, Y.; Hao, R.; Gu, X. The Mkurtogram: A novel method to select the optimal frequency band in the AC domain for railway wheelset bearings fault diagnosis. *Appl. Sci.* **2021**, *11*, 9. [[CrossRef](#)]