

Article

A Fault Diagnosis Scheme for Gearbox Based on Improved Entropy and Optimized Regularized Extreme Learning Machine

Wei Zhang ^{1,†}, Hong Lu ^{1,†} , Yongquan Zhang ¹, Zhangjie Li ^{1,*}, Yongjing Wang ², Jun Zhou ¹, Jiangnuo Mei ¹ and Yuzhan Wei ¹

¹ School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China

² Department of Mechanical Engineering, University of Birmingham, Birmingham B15 2TT, UK

* Correspondence: anhmcy@whut.edu.cn

† These authors contributed equally to this work.

Abstract: The performance of a gearbox is sensitive to failures, especially in the long-term high speed and heavy load field. However, the multi-fault diagnosis in gearboxes is a challenging problem because of the complex and non-stationary measured signal. To obtain fault information more fully and improve the accuracy of gearbox fault diagnosis, this paper proposes a feature extraction method, hierarchical refined composite multiscale fluctuation dispersion entropy (HRCMFDE) to extract the fault features of rolling bearing and the gear vibration signals at different layers and scales. On this basis, a novel fault diagnosis scheme for the gearbox based on HRCMFDE, ReliefF and grey wolf optimizer regularized extreme learning machine is proposed. Firstly, HRCMFDE is employed to extract the original features, the multi-frequency time information can be evaluated simultaneously, and the fault feature information can be extracted more fully. After that, ReliefF is used to screen the sensitive features from the high-dimensional fault features. Finally, the sensitive features are inputted into the optimized regularized extreme learning machine to identify the fault states of the gearbox. Through three different types of gearbox experiments, the experimental results confirm that the proposed method has better diagnostic performance and generalization, which can effectively and accurately identify the different fault categories of the gearbox and outperforms other contrastive methods.

Keywords: hierarchical refined composite multiscale fluctuation dispersion entropy (HRCMFDE); fault diagnosis; gearbox; regularized extreme learning machine; ReliefF; grey wolf

MSC: 68T10



Citation: Zhang, W.; Lu, H.; Zhang, Y.; Li, Z.; Wang, Y.; Zhou, J.; Mei, J.; Wei, Y. A Fault Diagnosis Scheme for Gearbox Based on Improved Entropy and Optimized Regularized Extreme Learning Machine. *Mathematics* **2022**, *10*, 4585. <https://doi.org/10.3390/math10234585>

Academic Editors: António Lopes and Paolo Mercorelli

Received: 9 October 2022

Accepted: 29 November 2022

Published: 3 December 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

As a critical part of the transmit power and motion in mechanical equipment, the gearbox has been widely used in many modern industrial fields such as aerospace, wind power generation, ship, rail transit and construction machinery. However, due to heavy loads and hostile working environments, it is easy to malfunction in the actual working process. These failures will lead to inevitable dynamic behavior and even significant accidents. To avoid losses caused by gearbox failures, accurate and automatic fault detection is of great value to ensure the safe and stable operation of mechanical equipment [1].

The research on gearbox fault diagnosis is mainly based on expert systems [2], analytical models [3] and data-driven methods [4]. The expert system-based method has substantial limitations and primarily relies on the experience of experts for diagnosis. The analytical model-based methods need to establish accurate and systematic mathematical models according to a specific mechanical structure, which is not always possible for complex mechanical systems [5]. The data-driven method analyzes an equipment's operating state through sensor data, which has received much attention in fault diagnosis. When

the gearbox fails, the failed point repeatedly collides with other parts in contact. It will cause nonlinear, non-stationary and multi-frequency complex signals. Therefore, how to extract the fault feature information that can represent the running state from this signal has become the key [6]. Researchers have proposed various state-of-the-art signal analysis methods and applied them to extract gearbox fault features, such as wavelet packet transform (WPT) [7], squared envelope spectrum (SES) [8], empirical mode decomposition (EMD) [9], variational modal decomposition (VMD) [10], machine learning [11] and entropy theories [12].

As a statistical measure, entropy can quantify complexity and detect the dynamic changes of signals through the nonlinear behavior of time series. It has become a hot research topic and study in many necessary fields, such as image processing [13], mechanical fault diagnosis [14], urban systems [15] and biomedical signals [16]. Due to its advantages in nonlinear vibration signals feature extraction, there are many entropy-based methods, such as sample entropy [17], fuzzy entropy [18] and permutation entropy [19]. These entropy-based methods or improved methods have been successfully applied in the field of mechanical equipment fault diagnosis. Feng [20] combines with the sample entropy, and the fault diagnosis of planetary gear under non-stationary operational conditions is realized. Wei [21] proposes an improved fuzzy entropy method for feature extraction of rotating machinery and verifies the effectiveness of the method through experiments. Kuai [22] decomposes the original signal into six intrinsic mode functions and defines the permutation entropies of each intrinsic mode function component as the input for the gearbox fault diagnosis. However, sample entropy has addressed the shortcoming, but the boundary of different categories is fuzzy in practical application. Fuzzy entropy can effectively solve this problem and improve the stability of the calculation results. Permutation entropy only compares the amplitude of time series in the calculation process and ignores the amplitude difference between the same pattern.

To tackle these problems, a method called frequency-based dispersion entropy (FDE) is introduced by Azami [23]. Through the comparative analysis of various kinds of classical signals, FDE has apparent advantages in terms of stability, calculation cost and noise-robustness.

Nevertheless, FDE only measures the randomness and dynamic uncertainty of time series on a single scale. To address the defect, multiscale fluctuation dispersion entropy (MFDE) [24], refined composite multiscale dispersion entropy (RCMDE) [25] and refined composite multiscale fluctuation dispersion entropy (RCMFDE) [26] have been proposed to measure the complexity of time series on multiple scales. However, MFDE, RCMDE and RCMFDE do not comprehensively consider the multiscale feature information of time series at different layers and frequency bands. These also ignore the feature information of different coarse-graining sequences at the same scale during the coarse-graining process, which results in the loss of helpful information and increases entropy estimation deviation. Meanwhile, Yan [27] introduces hierarchical dispersion entropy, and Wang [28] proposes hierarchical fluctuation dispersion entropy. HFDE and HDE can simultaneously extract high-frequency and low-frequency features of the signal. Nonetheless, in the face of complex signals, HFDE and HDE are unstable and have severe feature information loss. To address these shortcomings, this paper combines the advantages of the above methods. Further, it proposes hierarchical refined composite multiscale fluctuation dispersion entropy (HRCMFDE) to extract fault features of the gearbox vibration signals.

The HRCMFDE method extracts the gearbox features information from the time domain signals. The obtained high-dimensional feature vectors contain redundant information, which will drown the sensitive information [29]. In this paper, ReliefF is adopted to screen sensitive information [30], eliminate the correlation among the features and avoid redundancy. In the pattern recognition stage, a regularized extreme learning machine (RELM) [31] is introduced as a classifier. The performance of RELM depends on two parameters, namely, the regularization factor and the number of hidden neurons. To avoid choosing parameter combinations by experience, the grey wolf optimizer (GWO) [32] adapts

tively determines the best parameter combinations of RELM. Therefore, GWO-RELM is also proposed to give full play to the best performance of RELM.

According to the layout of the gearbox, gear trains can be classified into four categories [33]: simple gear train, compound gear train, reverted gear train and planetary gear train. One example is given in Figure 1 for each type of gear train. The types (b), (c) and (d) can be formed by the combination of (a). To verify the applicability and generalization of this method in the field of gearbox fault diagnosis, experimental research on gearboxes with more complex structures (b), (c) and (d) is carried out.

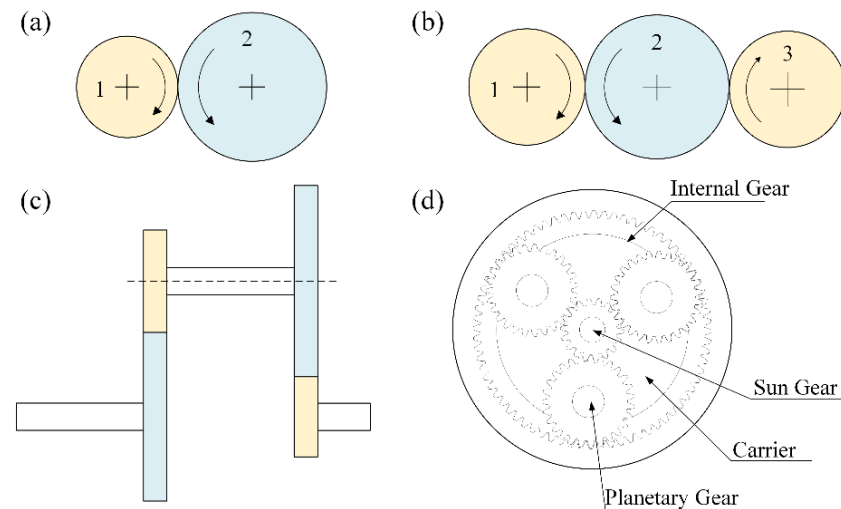


Figure 1. Four types of gear trains: (a) simple gear train, (b) compound gear train, (c) reverted gear train and (d) planetary gear train.

The main contributions of this paper can be summarized as follows:

- (1) A novel HRCMFDE method is employed to calculate the entropy value of the gearbox original vibration signals distributed over multiscale and multi-level fault feature extraction;
- (2) A novel fault diagnosis scheme for gearbox fault diagnosis is proposed based on HRCMFDE, ReliefF and GWO-RELM;
- (3) Experiment studies of the gearbox with single and compound failures are carried out. The results validate that the proposed method has a better detection ability than the existing four entropy-based approaches.

The rest of this paper is organized as follows. Section 2 presents the mathematical modelling and parameter selection of the HRCMFDE algorithm. Section 3 provides the steps of the proposed method in detail and includes the principle of GWO-RELM. Section 4 is the experimental verification. A series of gearbox experiments verify the superiority and generalization of the proposed method. Section 5 draws the conclusions.

2. HRCMFDE

RCMFDE do not comprehensively consider the multiscale feature information of time series at different layers and frequency bands, which inevitably leads to the loss of potential effective information. The paper puts forward hierarchical refined composite multiscale fluctuation dispersion entropy (HRCMFDE). By referring to the process of hierarchical analysis, the multi-frequency information of time can be evaluated simultaneously by constructing operators of different frequency bands, and the feature information can be extracted more fully.

2.1. Fluctuation Dispersion Entropy (FDE)

For random series $X = x_1, x_2, \dots, x_N$, its features are calculated as follows:

- (1) Obtaining the time series $Z = z_1, z_2, \dots, z_N$ by mapping each element in $X = x_1, x_2, \dots, x_N$ to different classes from 1 to c based on Equations (1) and (2),

$$y_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x(t)} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (1)$$

$$z_i = R(cy_i + 0.5) \quad (2)$$

where σ and μ denote the standard deviation and mean of x_i , R represents the rounding function and c stands for class, respectively;

- (2) Defining the vector Z based on embedding dimension m and time delay λ by Equation (3).

$$Z_k^{m,\lambda,c} = \{z_k, z_{k+\lambda}, \dots, z_{k+(m-1)\lambda}\} (k = [1, 2, \dots, N - (m-1)\lambda]) \quad (3)$$

The new series $\bar{Z}_k^{m,\lambda,c} = \{\bar{z}_{k,1}, \bar{z}_{k,2}, \dots, \bar{z}_{k,m-1}\}$ is gained on the basis of $Z_k^{m,\lambda,c} = \{z_k, z_{k+\lambda}, \dots, z_{k+(m-1)\lambda}\}$, where $\bar{z}_{k,m-1} = z_{k+(m-1)\lambda} - z_{k-1+(m-1)\lambda} + c$, $\bar{z}_{k,1} = v_0, \bar{z}_{k,2} = v_1, \dots, \bar{z}_{k,m-1} = v_{m-2}$. Consequently, the number of possible fluctuation dispersion modes is equal to $(2c-1)^{m-1}$. The probability of each mode can be calculated by:

$$p(\pi_{v_0 v_1 \dots v_{m-2}}) = \frac{\text{Number}\{j | j \leq N - (m-1)\lambda, \bar{Z}_k^{m,\lambda,c} \text{ has type } \pi_{v_0 v_1 \dots v_{m-2}}\}}{N - (m-1)\lambda}; \quad (4)$$

- (3) The FDE of series X can be computed as follows:

$$FED(X, m, c, \lambda) = - \sum_{\pi=1}^{(2c-1)^{m-1}} p(\pi_{v_0 v_1 \dots v_{m-2}}) \ln p(\pi_{v_0 v_1 \dots v_{m-2}}). \quad (5)$$

2.2. Refined Composite Multiscale Fluctuation Dispersion Entropy (RCMFDE)

The traditional coarse-graining multiscale method intercepts non-overlapping fragments, and the relationship between adjacent elements of each fragment is not fully considered. With the increase of the scale factor, the stability of the calculated results becomes worse. Therefore, the refined composite multiscale method is introduced, which is summarized as follows:

- (1) The original signal $X = \{x_1, x_2, \dots, x_N\}$ is continuously divided into a small sequence of length τ by the initial point in order $[1, \tau]$ and then taking the average of each small sequence. These means are arranged sequentially to obtain τ scale coarse-graining time series. The q^{th} coarse-graining time series $x_q^\tau = \{x_{q,1}^{(\tau)}, x_{q,2}^{(\tau)}, \dots\}$ in the τ scale is as follows:

$$x_{q,j}^{(\tau)} = \frac{1}{\tau} \sum_{b=q+\tau(j-1)}^{q+\tau j-1} x_b, 1 \leq j < \lfloor N/\tau \rfloor, 1 \leq q \leq \tau; \quad (6)$$

- (2) Then, for each scale factor, calculate the probability of each fluctuation dispersion mode occurring in the q^{th} coarse-graining time series x_q^τ . The average of the dispersion pattern π of the coarse-graining time series in the τ scale is as follows:

$$p(\pi_{v_0 v_1 \dots v_{m-2}}) = \frac{1}{\tau} \sum_{q=1}^{\tau} p_q^{(\tau)}; \quad (7)$$

- (3) The RCMFDE of series X can be computed as follows:

$$RCMFED(X, m, c, \lambda, \tau) = - \sum_{\pi=1}^{(2c-1)^{m-1}} \bar{p}(\pi_{v_0 v_1 \dots v_{m-2}}) \ln \bar{p}(\pi_{v_0 v_1 \dots v_{m-2}}). \quad (8)$$

2.3. Hierarchical Refined Composite Multiscale Fluctuation Dispersion Entropy (HRCMFDE)

RCMFDE ignores the feature information of different coarse-graining sequences at the same scale, which results in the loss of useful information and the increase of entropy estimation deviation. Therefore, hierarchical refined composite multiscale fluctuation dispersion entropy is proposed to extract the fault feature of vibration signals at different hierarchical layers and scales. The detailed flow of the HRCMFDE is required below.

- (1) For random series $X = x_1, x_2, \dots, x_N$ of length N , constructing the operators $Q_0(x)$ and $Q_1(x)$ are as follows:

$$\begin{cases} Q_0(x) = \frac{x_i + x_{i+1}}{2} \\ Q_1(x) = \frac{x_i - x_{i+1}}{2} \end{cases} \quad i = 1, 2, \dots, N-1 \quad (9)$$

where $Q_0(x)$ and $Q_1(x)$ contain the low-frequency information and high-frequency information of $X = x_1, x_2, \dots, x_N$, respectively;

- (2) Then, the matrix form of the operator Q_t^k ($t = 0$ or 1) at the hierarchical layer k is written as follows:

$$Q_t^k = \begin{bmatrix} \frac{1}{2} & \underbrace{0 \dots 0}_{2^{k-1}-1} & \frac{(-1)^t}{2} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \underbrace{0 \dots 0}_{2^{k-1}-1} & \frac{(-1)^t}{2} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{2} & \underbrace{0 \dots 0}_{2^{k-1}-1} & \frac{(-1)^t}{2} \end{bmatrix}_{(N-2^{k-1}+1) \times (N-2^{k-1}+1)} ; \quad (10)$$

- (3) Moreover, for a given vector $[v_1, v_2, \dots, v_k]$ of length k , the variable e can be calculated as follows:

$$e = \sum_{m=1}^k 2^{k-m} v_m \quad (11)$$

where $v_m \in \{0, 1\}$ ($m = 1, 2, \dots, k$) is the operator Q_0 or Q_1 at the m -th layer, according to Equation (10), a unique vector correspondence exists for any given non-negative integer e ;

- (4) The hierarchical components of the series X are represented as follows:

$$X_{k,e} = Q_{vk}^k \cdot Q_{v_{k-1}}^{k-1} \cdot \dots \cdot Q_{v_1}^1 \cdot X \quad (12)$$

where $X_{k,e}$ represents the hierarchical components at the node e of the k -th layer of series X . When $k = 3$, the hierarchical decomposition process is depicted in Figure 2, where $X_{3,1}$ represents the hierarchical component at node 1 of the 3-rd layer, the corresponding unique vector is $[0, 0, 1]$. $X_{1,0}$ and $X_{1,1}$ represent the high-frequency and low-frequency components in the first layer, respectively;

- (5) The RCMFDE value corresponding to the hierarchical node component $X_{k,e}$ under the scale factor τ is calculated as the HRCMFDE value under the scale factor, which can be expressed as follows:

$$HRCMFED(X, k, m, c, \lambda, \tau) = RCMFED(X_{k,e}, m, c, \lambda, \tau) \quad (13)$$

It can be seen from the above principle description that the HRCMFDE algorithm is optimized based on FDE and RCMFDE successively. The concepts of refined multiscale and hierarchical analysis are introduced, respectively, which can effectively extract information at different hierarchical layers and scales of the original signals. This method has better stability and performance. The flow of the HRCMFDE method is shown in Figure 3.

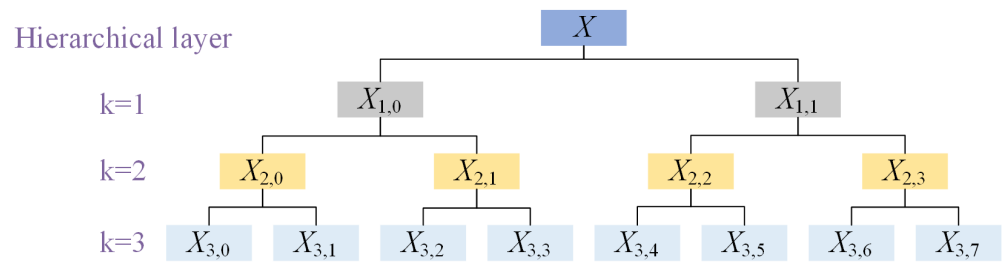


Figure 2. The hierarchical decomposition when $k = 3$.

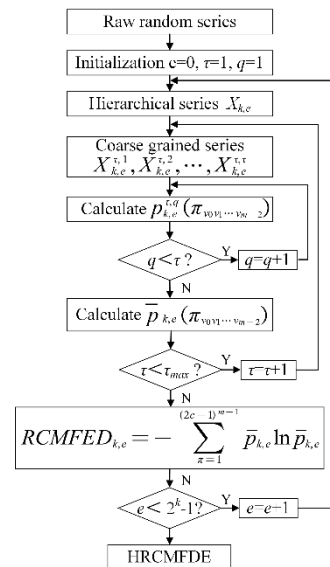


Figure 3. The flowchart of HRCMFDE.

2.4. Parameters Selection

Six main parameters in HRCMFDE need to be set manually: the series length N , the hierarchical layer k , the embedding dimension m , the class c , the time delay λ and the scale factor τ . Selecting proper parameters can process the original signals more effectively, which extracts the fault information more accurately:

- (1) In these parameters, if the scale factor τ is too large, redundant information will be easily generated. However, if τ is too small, obtaining helpful fault feature information from the original signals is challenging. If the hierarchical layer k is too small, it incompletely extracts high-frequency and low-frequency information of the signals. Nevertheless, the computational efficiency will be affected if it is too large. To extract valuable features, the literature results [34] set $\tau = 8$, $k = 3$, which can meet the requirement of gearbox fault diagnosis. Hence, using the HRCMFDE method, 64 features can be extracted from each group of signal samples, and the corresponding feature vector under the sample is constructed;
- (2) For the time delay λ and the series length N , the literature [35] indicates that the time delay λ and the series length N have less impact on the feature extraction result;
- (3) For the embedding dimension m and class c , the influence of different parameter values is analyzed using the distance measure. Assuming that the gearbox parts have n different health states, and each state has M samples of sample length N , the distance measure (average Euclidean distance) would be introduced as follows:

$$AED(x, y) = \frac{1}{n} \sum_{i=1}^M \sqrt{\sum_{j=1}^{k \cdot \tau_{\max}} (HRCMFDE_{x,i}(j) - HRCMFDE_{y,i}(j))^2} \quad (14)$$

$$Value_{AED} = \sum_{x=1}^n \sum_{y=1, y \neq x}^n AED(x, y) \quad (15)$$

where x and y denote the AED values of the x -th and the y -th states, $Value_{AED}$ is the AED value corresponding to the parameter (m, c) , respectively.

Then, repeating the calculation for different parameter combinations, m is determined according to the criterion $N / \tau_{max} > (2c - 1)^{m-1}$ and, according to the literature results [36], set $c \in [4, 8]$. The (m, c) combination corresponding to the maximum $Value_{AED}$ value is the best (m, c) combination.

3. The Proposed Gearbox Intelligent Fault Diagnosis Method

According to the proposed feature extraction method, there is still redundant information in the feature vectors, affecting the recognition accuracy and increasing the calculation cost. Therefore, this section mainly introduces a feature dimension reduction method which removes redundant high-dimensional information and realizes the screening of sensitive features. At the same time, an improved classification method is introduced to realize the final fault diagnosis.

3.1. Grey Wolf Optimizer

Grey wolf optimizer (GWO) is one of the most popular metaheuristic algorithms in the recent decade, which Australian scholar Mirjalili proposes. The introduction of this algorithm is detailed in the literature and will not be described in this paper [37].

The algorithm program of GWO can be described in Algorithm 1.

Algorithm 1: GWO

- (1) Initialize the grey wolf population $X_i (i = 1, 2, \dots, n)$;
 - (2) Initialize α , A and C;
 - (3) Calculate the objective values for each search agent
 X_α = the best search agent
 X_β = the second-best search agent
 X_δ = the third-best search agent;
 - (4) for $t = 1$: max number of iterations
for each search agent
Update the position of the current search agent by Equations (21)–(26)
end for
Update α , A and C
Calculate the fitness of all search agents
Update X_α , X_β , and X_δ
end for
 - (5) Return X_α .
-

The steps of the GWO are as follows:

During hunting, the encircling behavior of grey wolves can be defined as:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}_p(t)| \quad (16)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (17)$$

where t is the current iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}_p$ is the position vector of the prey and $\vec{X}$ indicates the position vector of a grey wolf. The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (18)$$

$$\vec{a} = 2 - 2 \left(\frac{1}{e - 1} \times \left(e^{\frac{t}{m}} - 1 \right) \right) \quad (19)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (20)$$

where r_1, r_2 are random vectors in $[0,1]$, t is the number of the current iteration and m is the maximum number of iterations.

The mathematical model of individual grey wolf tracking prey is described in Equations (21) and (22).

$$\begin{cases} \vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \\ \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \\ \vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \end{cases} \quad (21)$$

$$\begin{cases} \vec{X}_1 = \vec{X}_\alpha - A_1 \cdot \vec{D}_\alpha \\ \vec{X}_2 = \vec{X}_\beta - A_2 \cdot \vec{D}_\beta \\ \vec{X}_3 = \vec{X}_\delta - A_3 \cdot \vec{D}_\delta \end{cases} \quad (22)$$

A proportional weight based on the modulus of the guide position vector is introduced. By adjusting the weights, the global and local search ability of the algorithm is dynamically balanced, and the convergence of the algorithm is accelerated. The calculation formulas are as follows:

$$v_1 = \frac{|\vec{X}_1|}{|\vec{X}_1| + |\vec{X}_2| + |\vec{X}_3|} \quad (23)$$

$$v_2 = \frac{|\vec{X}_2|}{|\vec{X}_1| + |\vec{X}_2| + |\vec{X}_3|} \quad (24)$$

$$v_3 = \frac{|\vec{X}_3|}{|\vec{X}_1| + |\vec{X}_2| + |\vec{X}_3|} \quad (25)$$

$$\vec{X}(t+1) = \frac{v_1 \cdot \vec{X}_1 + v_2 \cdot \vec{X}_2 + v_3 \cdot \vec{X}_3}{3} \quad (26)$$

Equation (21) defines the step length and direction of grey wolf individuals to α , β and δ , Equations (21) and (22) define the final position of X_α .

3.2. Regularized Extreme Learning Machine

The regularized extreme learning machine is used as a classification algorithm, and low dimensional feature vectors of test samples are inputted to realize the fault diagnosis of the gearbox. RELM introduces the concept of regularization based on the extreme learning machine (ELM), which is an improved method based on ELM. ELM is a fast-training algorithm for SLFN proposed by Huang [38]. SLFN has been widely used in many fields with its better learning ability, and the structure is shown in Figure 4.

Assuming a training dataset $\{(x_i, t_i)\}$, where $x_i \in R^n$, $t_i \in R^m$ and $i = 1, 2, \dots, N$. The activation function is $g(x)$ and the number of hidden nodes is k . The training steps of the ELM algorithm are as follows:

- (1) Randomly set input weights w_j and hidden layer biases b_j :

$$(w_j, b_j), j = 1, 2, \dots, k; \quad (27)$$

- (2) The output of SLFN can be formulated as follows:

$$O_j = \sum_{j=1}^k \beta_j \cdot g(w_j \cdot x_i + b_j), j = 1, 2, \dots, N \quad (28)$$

where β_j is the set of values of connection weights between the hidden layer and the output layer. The output Equations for the input samples can be represented as $H\beta = T$, where:

$$\beta = [\beta_1^T \dots \beta_k^T]_{k \times m} \quad (29)$$

$$T = [y_1^T \dots y_N^T]_{N \times m} \quad (30)$$

$$H(w_1, w_2, \dots, w_k, b_1, b_2, \dots, b_k, x_1, x_2, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_k \cdot x_1 + b_k) \\ \vdots & \vdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_k \cdot x_N + b_k) \end{bmatrix}_{N \times k}; \quad (31)$$

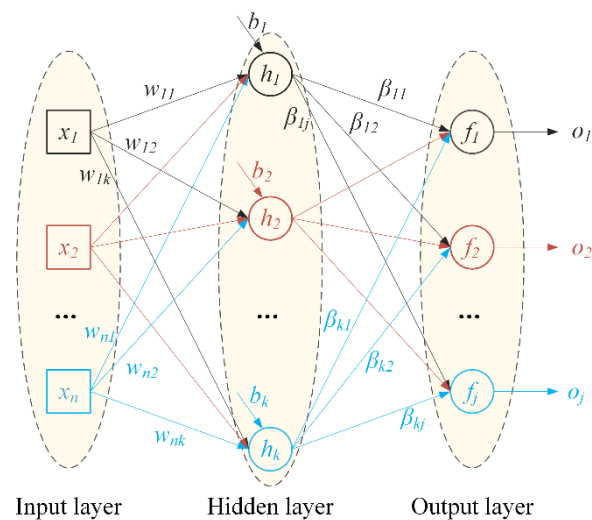


Figure 4. The structure of the SLFN.

- (3) Obtaining the output weights matrix β by solving the least multiplication solution of the following Equation:

$$\tilde{\beta} = \min_{\beta} \|H\beta - T\| = H^+ T \quad (32)$$

where H^+ is Moore-Penrose generalized inverse matrix;

- (4) Building the model of regularized extreme learning machine by the following Equation:

$$\tilde{\beta} = (\theta I + H^T H)^{-1} H^T T \quad (33)$$

where H^T is the Transposed matrix, θ is the regularization factor and I is the identity matrix, using the non-singular matrix $(H^T H)^{-1} H^T$ to replace the matrix H^+ . RELM can avoid overfitting and enhance the generalization ability of the model, improving the accuracy of the actual prediction. All in all, RELM has a more stable performance than ELM.

The algorithm program of RELM can be described in Algorithm 2.

3.3. Hybrid GWO-RELM

The training of RELM requires randomly setting the number of hidden neurons and constantly adjusting the number n of hidden neurons to search for a better value. If the value of n is too large it will increase the possibility of overfitting and take too much time. On the contrary, achieving the best accuracy and stability is difficult. Moreover, the value of θ depends on the input sample and needs to be set according to the results of many experiments.

To overcome the problems mentioned above and improve the efficiency of RELM, a hybrid means that combines GWO with RELM is required. The goal of the GWO algorithm

is to optimize the parameters to find the best set of n and θ by avoiding over-fitting and improving generalization ability.

Algorithm 2: RELM

Input: a training set $\{(x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, 2, \dots, N\}$
 the hidden node output function: $g(x)$ // Sigmoidal Function; Radial Basis Function; Triangular Basis Function et al.
 number of hidden node numbers: k
 label vector: $S_k \in R^C$, C is the number of classes
Output: $\tilde{\beta}$
 (1) Randomly assign the weights w_j and biases b_j , ($w_j \in [-1, 1], b_j \in [-1, 1]$);
 (2) Calculate the hidden layer output matrix H ;
 (3) Calculate the output weight $\tilde{\beta}$;
 (4) Return $\tilde{\beta}$.

The fitness function is the essential design problem to be solved in the GWO-RELM application. In the research of this paper, the selection of the commonly used fitness functions is the minimization of the root mean squared error (RMSE) given in Equation (34).

$$F(n, \theta)_{\min} = \sqrt{\sum_{i=1}^N (T_i - P_i)^2 / N} \quad (34)$$

where N is the number of training samples, T_i is the actual value and P_i is the predicted value. The steps of the GWO-RELM are shown as follows:

- (1) Building fitness function for optimization parameters n and θ ;
- (2) Setting the initial parameters and taking $[n, \theta]$ as the grey wolf position to generate the initial population;
- (3) Calculating the fitness of individual grey wolves in the population;
- (4) Repeating several iterations and constantly updating the optimal fitness value;
- (5) Outputting the best parameters and the corresponding accuracy.

The flow charts of GWO-RELM are shown in Figure 5.

3.4. ReliefF

The high-dimensional feature vectors extracted by the HRCMFDE method are rich in fault feature information and redundant information. If all the feature information is used for fault diagnosis, the accuracy and efficiency of the diagnosis will be affected. Therefore, according to the importance and sensitivity of each feature, it is essential to reduce the dimension of high-dimensional feature vectors and obtain sensitive low-dimensional feature vectors. This paper uses the ReliefF method for feature dimension reduction; the detailed description of ReliefF is in reference [39].

A sample R_i is randomly selected from the training set for the high-dimensional feature. Then, k nearest neighbour samples are chosen from the samples with the same label, and select k nearest neighbour samples from the different labels. Finally, using Equation (35), update the corresponding weight of the feature constantly, and the calculation is carried out m times until all the samples are successively calculated. The final weight of a single feature is obtained.

$$W^{i+1}(f_l) = W^i(f_l) - \sum_{j=1}^k \frac{\text{diff}(f_l, R_i, H_j)}{mk} + \sum_{C \neq \text{label}(R_i)} \left[\frac{P(C)}{1 - P(\text{label}(R_i))} \right] \sum_{j=1}^k \frac{\text{diff}(f_l, R_i, M_j(C))}{mk} \quad (35)$$

where $W^i(f_l)$ is the weight of the l -th feature f in the i -th sample; $H_j (j = 1, 2, \dots, k)$ is the j -th sample among k nearest neighbour samples of the same kind as R_i ; $P(C)$ is the

probability of label C ; $P(\text{label}(R_i))$ is the probability of samples of the same kind as R_i to the total samples; and $M_i(C)$ represents k nearest neighbour samples different from R . The calculation method of function $\text{diff}(f, R_1, R_2)$ is shown in Equation (36).

$$\text{diff}(f, R_1, R_2) = |R_{1f} - R_{2f}| / (\max(f) - \min(f)) \quad (36)$$

where $\text{diff}(f, R_1, R_2)$ is the normalized distance between sample R_1 and sample R_2 on the f -th feature. R_{1f} and R_{2f} are the f -th feature of samples R_1 and sample R_2 .

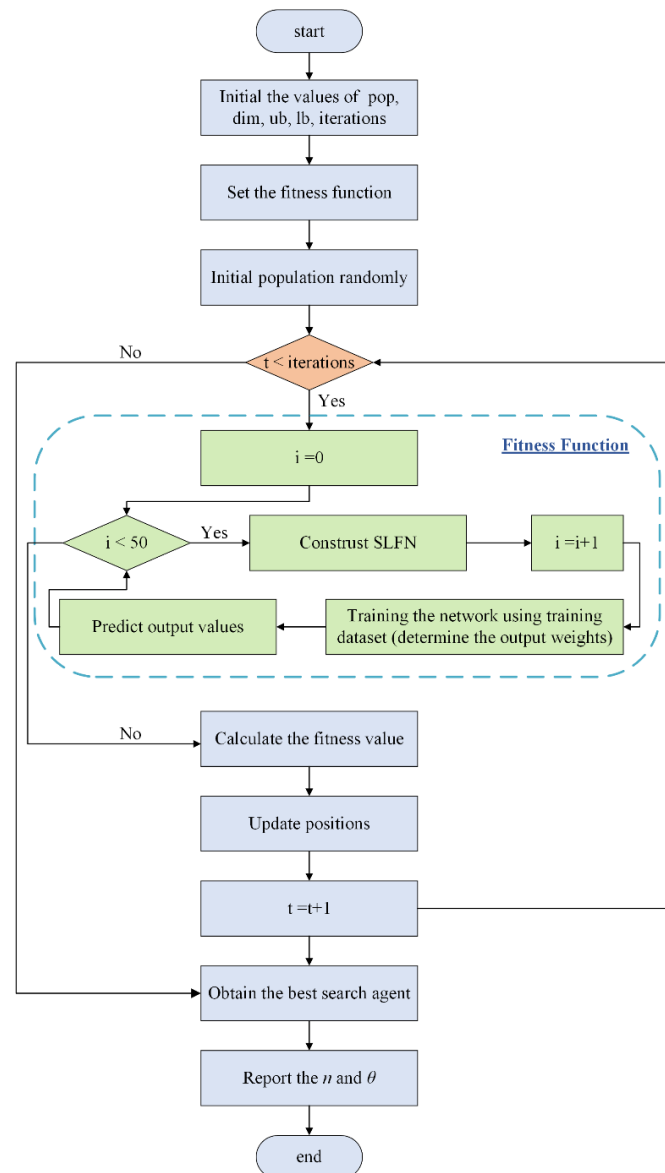


Figure 5. Flowchart of the GWO-RELM algorithm.

3.5. The Proposed Fault Diagnosis Method

To ensure high fault classification accuracy for the gearbox. Based on HRCMFDE, ReliefF and GWO-RELM, a novel gearbox fault diagnosis method is presented in this paper, and the detailed process is shown in Figure 6.

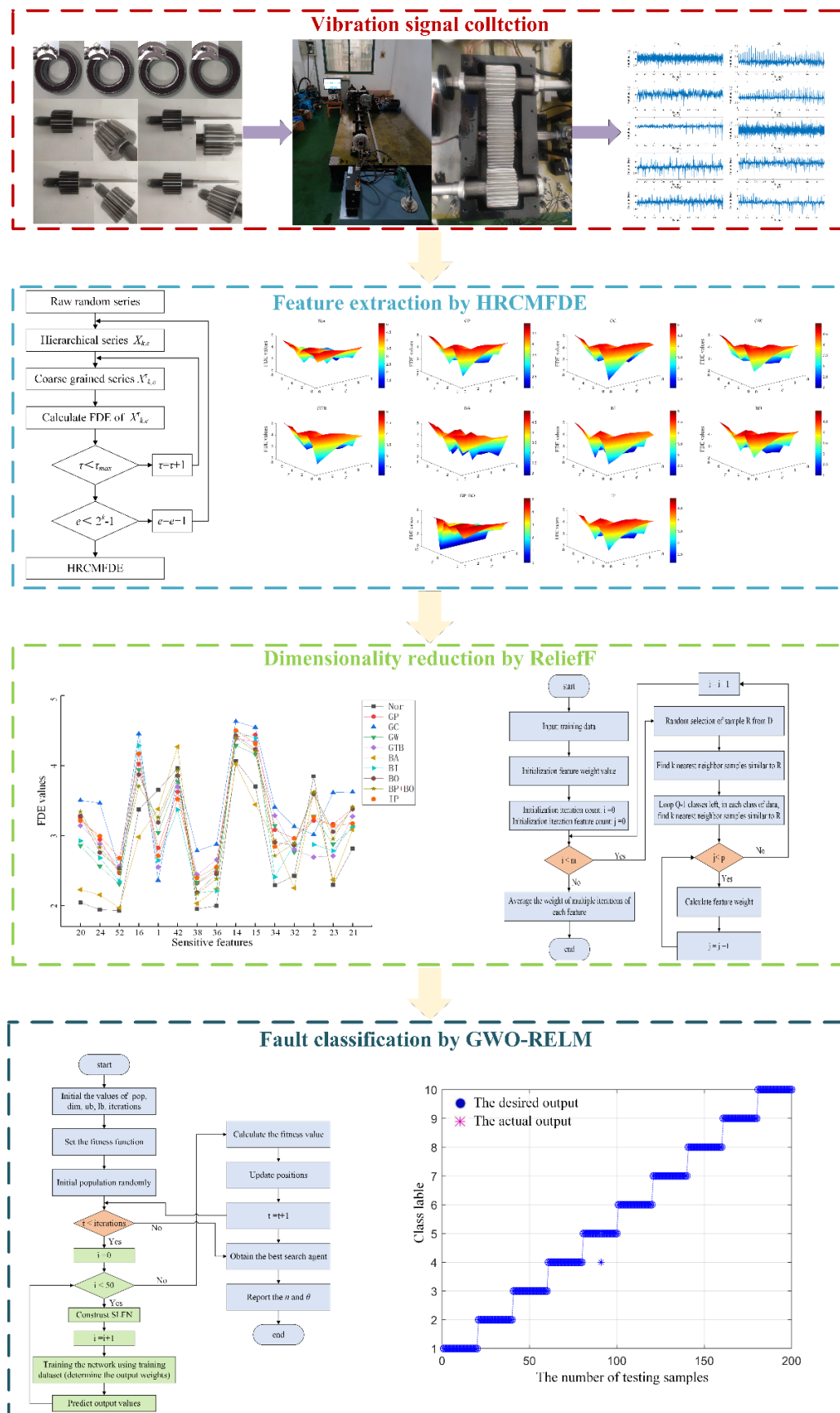


Figure 6. Flowchart of the proposed fault diagnosis method.

- (1) Collecting the vibration signals. The various fault states of gears and rolling bearings in the gearbox are collected by accelerometers;
- (2) Determining the optimal parameters of HRCMFDE. The features under different (m, c) combinations are extracted, respectively, and the (m, c) combination corresponding to the maximum $Value_{AED}$ value is taken as the optimal parameter;
- (3) Extracting fault features. To extract the fault feature information of the gearbox completely, the HRCMFDE method is employed to calculate the entropy value, and the feature set with a length of 64 is obtained;
- (4) Feature dimension reduction. ReliefF is utilized to extract sensitive feature information and remove redundant features;
- (5) Fault classification. The obtained low-dimensional sensitive feature information is inputted into GWO-RELM to identify the health conditions of the gearbox.

4. Experimental Verification

In this section, to verify the diagnostic effectiveness and generalization of the above methods, the gearboxes of three structural types as shown in Figure 1b–d, are selected to carry out experimental testing.

4.1. Experiment 1: Fault Diagnosis of Reverted Gear Train Gearbox

The experiment data comes from the 2009 PHM Challenge gearbox composite fault data set [40]. The experimental platform and its structure principle used in the experiment are shown in Figure 7, which mainly consists of the shaft, bearing, gear and other components.

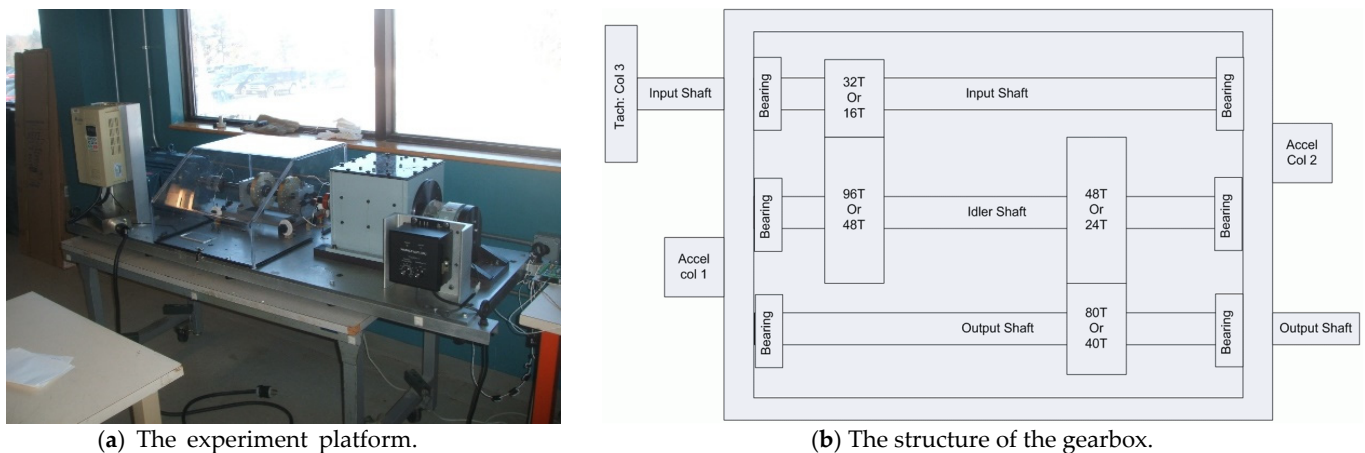


Figure 7. The experimental platform and gearbox structure.

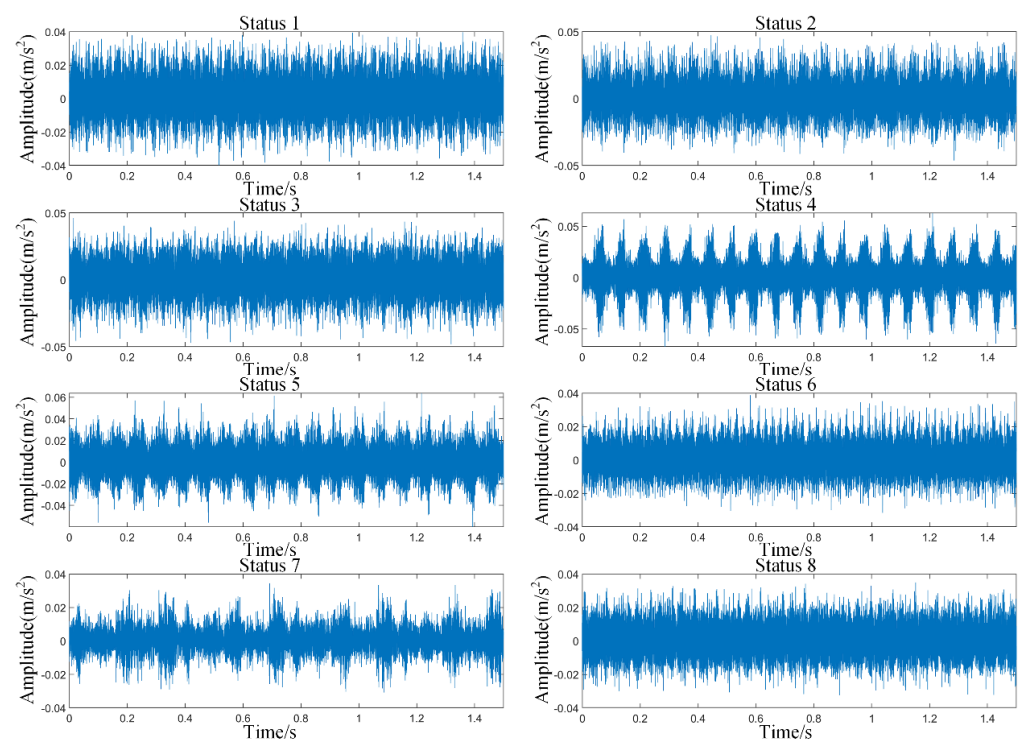
In the study, using the data set of the spur gear for analysis, which includes a normal operation state, single fault state and compound fault, fully reflects the fault state in the actual operation process of the gearbox. The detailed and time domain waveforms are depicted in Table 1 and Figure 8.

The experiment is performed in the input shaft speed is 2400 r/min and the low load, the corresponding number of teeth of the spur gear 1, 2, 3, 4 are 16, 48, 24, 40, respectively. Vibration signals are collected by two accelerometers, and the installation mode as shown in Figure 9, which the paper uses the vibration data obtained by 1 channel, with sample frequency of 66.7 kHz and sampling time is 4 s. For each working status, 60 samples with the length of 2048 are taken, where 40 samples as the training samples and 20 samples as the testing samples.

Table 1. Detailed information on different working statuses.

Labels	Gearbox Status	Gear				Bearing				Shaft	
		1	2	3	4	1	2	3	Others	Input	Output
1	Status 1	Nor	Nor	Nor	Nor	Nor	Nor	Nor	Nor	Nor	Nor
2	Status 2	Chi	Nor	Ecc	Nor	Nor	Nor	Nor	Nor	Nor	Nor
3	Status 3	Nor	Nor	Ecc	Nor	Nor	Nor	Nor	Nor	Nor	Nor
4	Status 4	Nor	Nor	Ecc	Bro	Ball	Nor	Nor	Nor	Nor	Nor
5	Status 5	Chi	Nor	Ecc	Bro	Inner	Ball	Outer	Nor	Nor	Nor
6	Status 6	Nor	Nor	Nor	Bro	Inner	Ball	Outer	Nor	Imb	Nor
7	Status 7	Nor	Nor	Nor	Nor	Inner	Nor	Nor	Nor	Nor	Key
8	Status 8	Nor	Nor	Nor	Nor	Nor	Ball	Outer	Nor	Imb	Nor

Nor = Normal; Chi = Chipped; Ecc = Eccentric; Bro = Broken; Imb = Imbalance; Key = Keyway Sheared.

**Figure 8.** Waveforms corresponding to different states.

The performance of the proposed method is verified by experimental data. Firstly, selecting the best (m, c) combination according to the AED method proposed in Section 2.4 is required, and 50 samples are randomly selected for each working state of the gearbox. The results of $Value_{AED}$ under different (m, c) combinations are illustrated in Table 2. It can be found that, with the increase of (m, c) , the $Value_{AED}$ also increases, the Euclidean distance between samples of different states becomes larger and the separability is constantly enhanced. Hence, selecting $m = 3$ and $c = 8$. Comprehensively, the final parameters are set to $k = 3$, $\tau = 8$, $\lambda = 1$, $N = 2048$, $m = 3$ and $c = 8$.

Table 2. The $Value_{AED}$ under different (m, c) combinations.

(m, c)	$Value_{AED}$	(m, c)	$Value_{AED}$
(2,4)	40.72	(3,4)	76.15
(2,5)	44.18	(3,5)	82.94
(2,6)	44.93	(3,6)	85.07
(2,7)	46.37	(3,7)	88.08
(2,8)	46.73	(3,8)	88.60

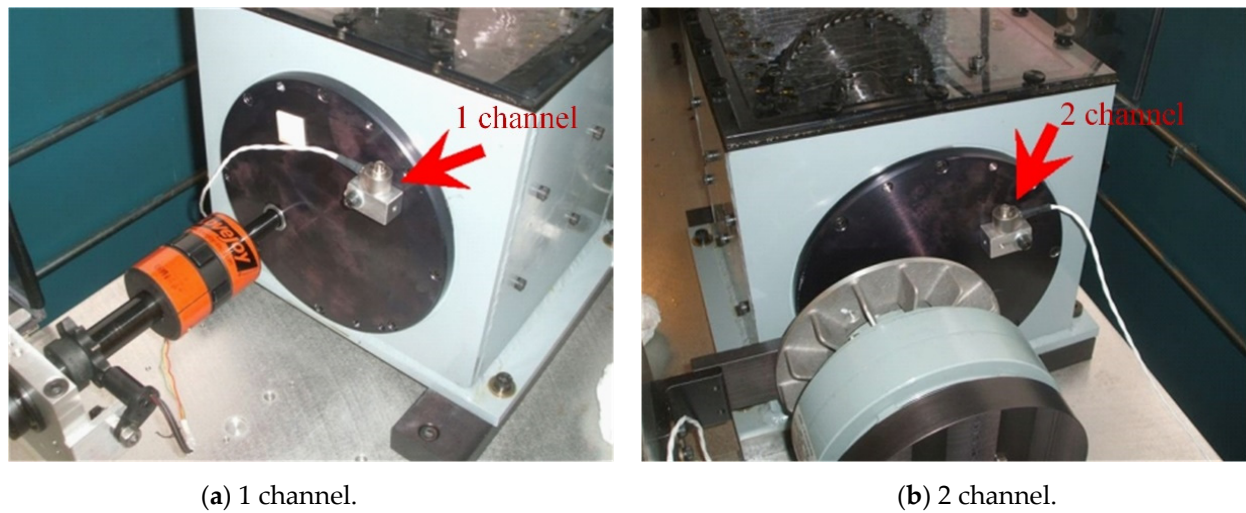


Figure 9. Accelerometers installation location.

The results of the feature extraction of the training samples are shown in Figure 10, and the low-dimensional features after feature reduction are displayed in Figure 11, respectively. It can be seen that there are differences in features of different states, but it is difficult to distinguish them directly. Therefore, it is necessary to rely on a classification algorithm to identify the states. The sensitive feature vectors are inputted into GWO-RELM, and the final setting optimization parameters are set to $n = 113$, $\theta = 0.509$.

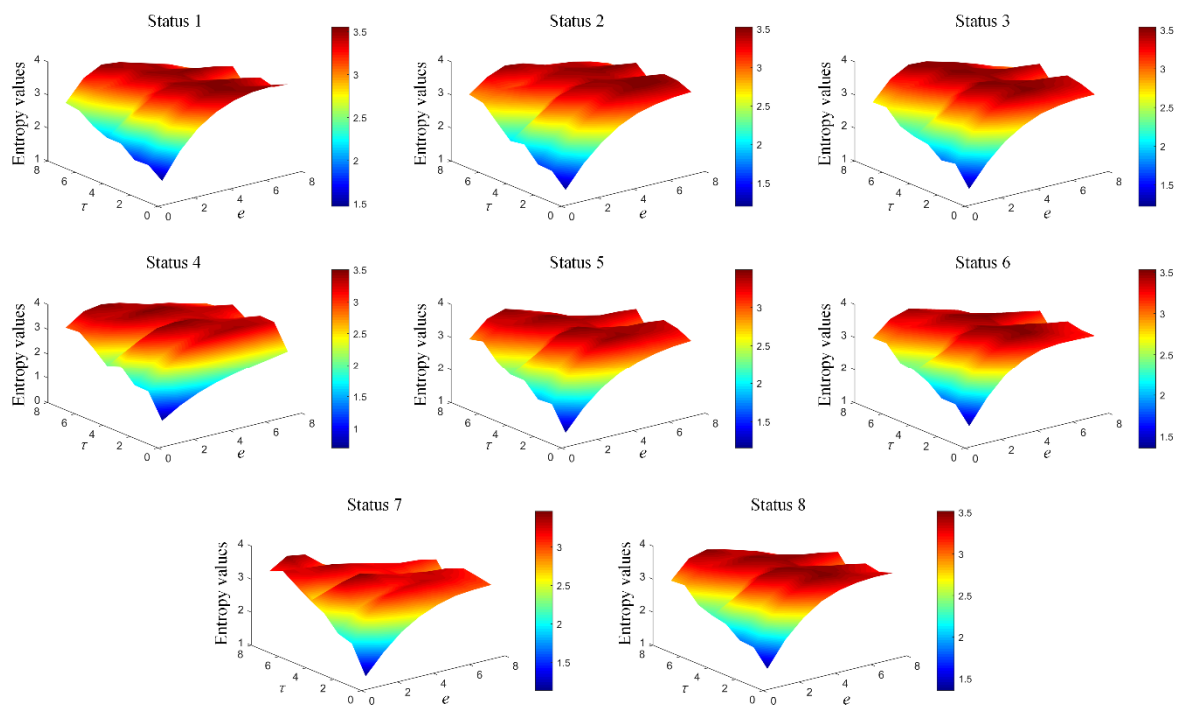


Figure 10. Raw fault features corresponding to different states.

To further verify the performance of the method, the samples of the training and testing are inputted into the optimized RELM model for state recognition. The results are shown in Figure 12. Among all test samples, 159 samples are identified successfully, and only one sample is incorrectly identified ("Status 3" is identified as "Status 1"). In the experiment, the recognition accuracy of all samples in different operating states of the gearbox is

99.38%. It is indicated that the proposed method can effectively realize the fault diagnosis of the gearbox under various working conditions, such as single fault and compound fault.

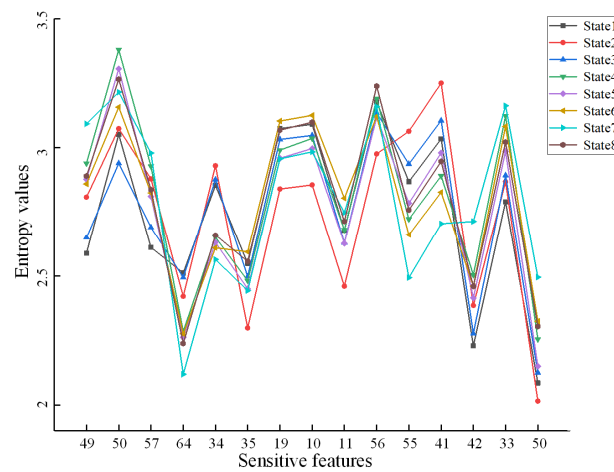


Figure 11. Sensitive fault features of different states.

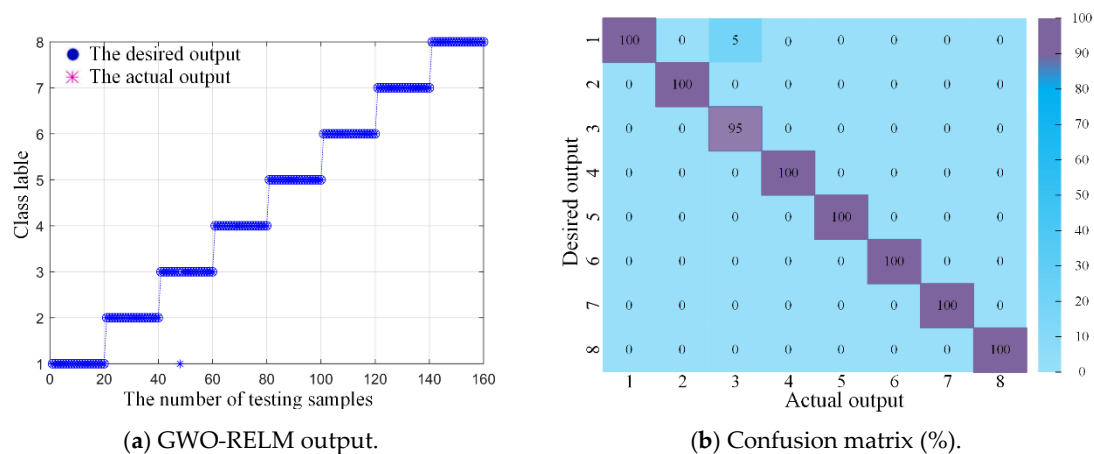


Figure 12. Identification results of the proposed method.

Then, HRCMFDE is compared with the existing RCMFDE, MFDE, RCMDE and MDE. The parameters are also set to $k = 3$, $\tau = 8$, $\lambda = 1$, $N = 2048$, $m = 3$ and $c = 8$. For each model, the experiment is repeated 50 times, and the results are shown in Table 3 and Figure 13, where 'Time' in Table 3 refers to the time consumed by a single sample to extract the high-dimensional features. The following conclusions can be found:

- (1) The correlations can be found by SD values as follows: The SD of MFDE and RCMFDE is smaller than MDE and RCMDE, respectively. It can be seen that FDE has better feature evaluation performance than DE, considering the fluctuation characteristics of the vibration signals. The SD of RCMFDE and RCMDE is smaller than MFDE and MDE, respectively, which means that the refined composite entropy-based method has better stability. The SD of HRCMFDE is small than RCMFDE, indicating that the hierarchical entropy-based method further improves the stability. Among these methods, HRCMFDE has the best stability and apparent advantages. The reason is that the refined composite multiscale entropy only analyzes the low-frequency signals and often ignores the high-frequency signals, resulting in a relatively large limitation of feature extraction performance;
- (2) The MFDE model has the fastest calculation speed, but the diagnostic accuracy is insufficient, and the significant variance indicates a lack of stability. Although the HRCMFDE model has the lowest computational efficiency, it is still acceptable in

practical applications. Ultimately, HRCMFDE has the highest accuracy. The reason is that the information extracted from low-frequency to high-frequency is the most extensive, and the feature information contained is the richest. Hence, it has the best stability, the highest diagnostic accuracy and a longer calculation time. The above analysis shows that the HRCMFDE model proposed in this paper has apparent advantages in the separability and stability of features. The HRCMFDE model can be effectively applied in gearbox fault diagnosis.

Table 3. The performance comparison between different feature extraction models.

Feature Extraction Method	Time (s)	Recognition Accuracy (%)			
		Max	Min	Mean	SD
HRCMFDE	1.623	100.00	97.50	98.64	0.46
RCMFDE	0.961	98.13	91.25	94.53	2.69
MFDE	0.017	81.25	71.88	76.48	5.16
RCMDE	1.021	95.63	87.50	91.99	2.96
MDE	0.125	78.13	67.50	72.66	5.51

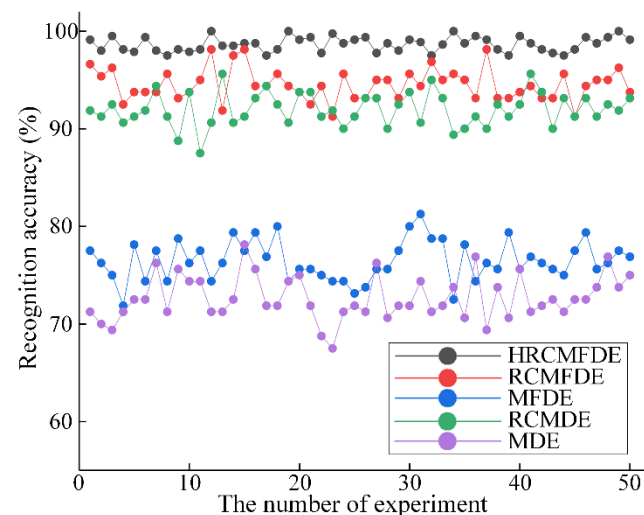


Figure 13. The accuracy of different approaches.

Then, the effectiveness of the dimension reduction method in the proposed approach is studied. The high-dimensional feature vectors are directly inputted into GWO-RELM to identify the health conditions of the gearbox. The results shown in Figure 14 are obtained according to Tables 3 and 4.

Table 4. The performance comparison between different methods without ReliefF.

Method	Recognition Accuracy (%)			
	Max	Min	Mean	SD
HRCMFDE + ReliefF + GWO – RELM	95.63	90.63	93.44	1.06
RCMFDE + ReliefF + GWO – RELM	91.25	83.13	87.78	3.29
MFDE + ReliefF + GWO – RELM	63.13	43.75	56.77	6.06
RCMDE + ReliefF + GWO – RELM	91.25	81.88	85.15	4.17
MDE + ReliefF + GWO – RELM	51.25	35.00	42.23	5.81

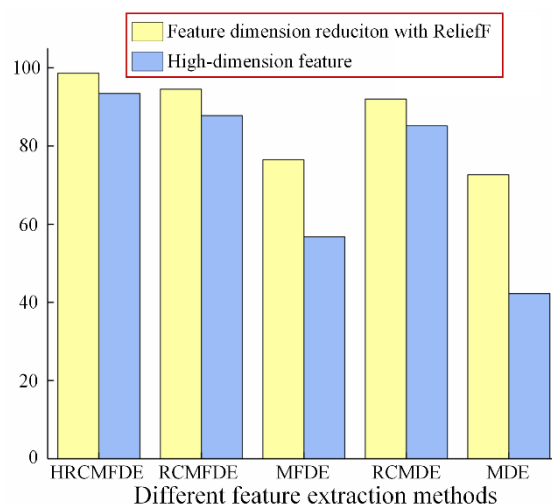


Figure 14. Comparison of identification accuracy before and after ReliefF.

Obviously, after ReliefF, the SD of different methods is significantly reduced, and the average recognition accuracy is significantly improved. It indicates that the low-dimensional feature vectors obtained by ReliefF strengthen the stability and accuracy of the recognition and are more suitable for the recognition of the operating state of the gearbox. All in all, ReliefF is an essential process for gearbox fault diagnosis.

After that, several classification approaches widely studied in current fault diagnosis algorithms are selected and compared with the method proposed in this paper. The results of the feature extraction model under different classification approaches are shown in Table 5. Therefore, the proposed classification method has better classification performance.

Table 5. Comparison of models under different classification methods.

Different Classification Methods	HRCMFDE	RCMFDE	MFDE	RCMDE	MDE
mRMR (%)	98.32	91.76	78.58	92.25	67.74
SVM (%)	97.99	93.74	78.07	91.89	66.62
PNN (%)	94.01	89.14	75.41	89.80	64.20
ELM (%)	96.89	93.47	77.71	92.14	65.35
GWO-RELM (%)	99.87	94.86	80.71	93.13	70.06

Finally, the HRCMFDE, RCMFDE and RCMDE models with better feature extraction performance are further evaluated. Commonly used indicators used in fault diagnosis methods to evaluate the superiority of model performance include Precision (P), Recall (R), Accuracy (Acc) and F1 score (F1) [41]. Precision is the ratio of the actual positive samples predicted in the test model to the predicted positive samples, which indicates the proportion of the real positive samples in the prediction results of the model. The Recall is the ratio of the number of true positive samples predicted by the model and the number of true positive samples in the samples. The Accuracy and F1 score are used to measure the overall performance of the model. The higher the index, the stronger the fault diagnosis capability of the model and the better the overall performance.

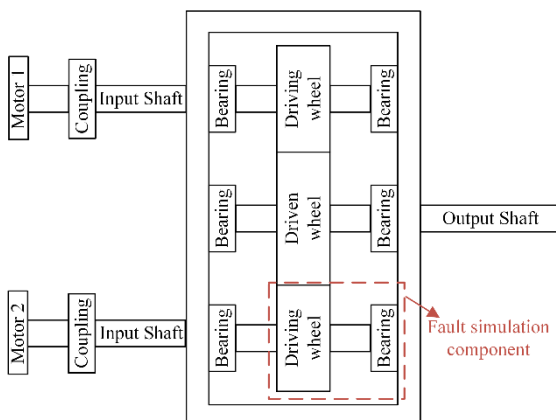
Each state is taken as a positive class, and the corresponding four indicators under this state are calculated successively. Each group of experiments is carried out 50 times. The average values are recorded in Table 6, where the status corresponds to various fault states in Table 1 and OM is the overall means. Compared with the RCMDE model, the RCMFDE model has better feature extraction performance and higher comprehensive scoring. Compared with the RCMFDE model, the P-means, R-means, Acc-means and F-means of the HRCMFDE model are increased by 3.89%, 4.16%, 1.04% and 4.17%, respectively. It shows that the HRCMFDE model has superior performance, higher accuracy and stability.

4.2. Experiment 2: Fault Diagnosis of Compound Gear Train Gearbox

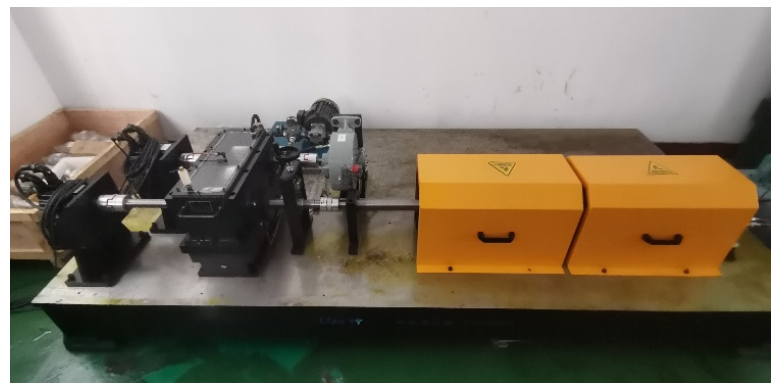
In Section 4.1, the validity of the proposed method is verified by the typical vibration signals in the reverted gear train gearbox. Then, this method is used in another experiment to verify the effectiveness further and provide an effective state diagnosis method for the compound gear train gearbox, which offers the basis for other studies on the experimental platform. The structure of the experimental platform and gearbox is shown in Figure 15, which adopts a dual-input single-output fault diagnosis platform to collect the vibration signals of the gearbox in different working conditions. The platform mainly consists of driving motors, gears, bearings, transmission shafts and other components.

Table 6. Comparison of different models.

Status	HRCMFDE				RCMFDE				RCMDE			
	P (%)	R (%)	Acc (%)	F1 (%)	P (%)	R (%)	Acc (%)	F1 (%)	P (%)	R (%)	Acc (%)	F1 (%)
G1	97.79	99.00	99.44	97.79	98.54	97.25	99.47	97.84	94.53	85.50	97.56	89.68
G2	100.00	100.00	100.00	100.00	99.75	97.75	99.69	98.72	94.15	92.25	98.31	93.15
G3	97.84	97.00	99.47	97.84	86.02	96.50	97.56	90.87	84.56	83.75	96.00	83.98
G4	98.70	97.50	99.69	98.70	98.72	93.25	99.00	95.87	93.35	94.50	98.47	93.89
G5	98.63	98.75	99.66	98.63	88.17	80.50	96.19	84.08	83.13	83.50	95.78	83.24
G6	99.51	99.75	99.88	99.51	97.65	99.25	99.59	98.41	96.45	100.00	99.53	98.18
G7	99.22	98.50	99.81	99.22	99.25	94.25	99.19	96.66	99.26	98.75	99.75	98.99
G8	98.06	99.25	99.50	98.06	90.53	97.75	98.44	93.98	90.52	96.50	98.28	93.38
OM	98.72	98.72	99.68	98.72	94.83	94.56	98.64	94.55	91.99	91.84	97.96	91.81



(a) GWO-RELM output.



(b) Confusion matrix (%).

Figure 15. The experimental platform and gearbox structure.

In the experiment, the driving motor provides power, and the two driving wheels transmit the power to the driven gear to achieve power transmission. The internal structure of the gearbox and the layout of the accelerometer are displayed in Figure 16. In this paper, sensor data of channel 1 are employed for analysis, and ten different working states are simulated by replacing different fault components.

In the experiment, the sampling frequency is 2048 Hz, the sampling time is the 90 s and the driving wheel speed is 1200 r/min. 60 samples with a length of 2048 are taken under each working state, with 40 samples as the training samples and 20 samples as the testing samples. The detailed fault information of the gearbox components is shown in Table 7. The components with various faults of gears and bearings are shown in Figure 17, which includes single and compound faults of gears and gearings. The time-domain waveforms of different states are shown in Figure 18, and the difference between signals of each state can be found.

Firstly, selecting the best (m, c) of HRCMFDE is required. It is observed that $m = 3$ and $c = 8$, based on the result in Table 8, are selected, and the other parameter selection is the same as experiment 1, and the final parameters are set to $k = 3$, $\tau = 8$, $\lambda = 1$, $N = 2048$, $m = 3$ and $c = 8$. The high-dimensional fault features corresponding to the different states extracted by HRCMFDE are shown in Figure 19, and the low-dimensional features after feature reduction are displayed in Figure 20.

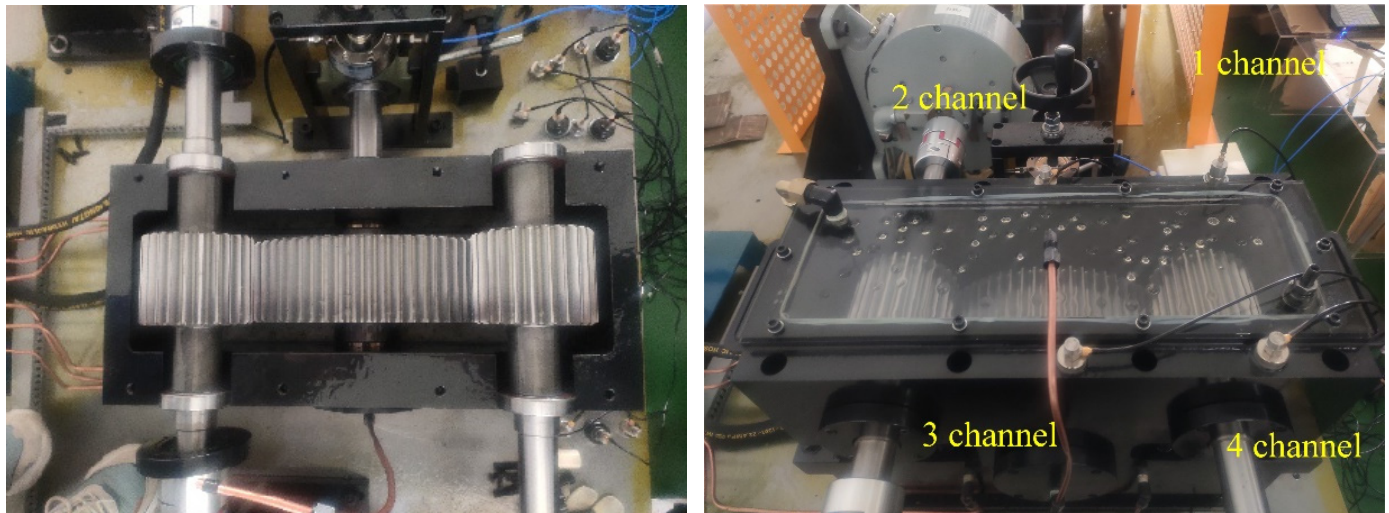


Figure 16. The internal structure of the gearbox and the installation positions of accelerometers.

Table 7. The detailed fault information of gearbox components.

Labels	Gearbox States	Abbreviation	Fault Type	Size of Training Samples	Size of Testing Samples
1	Normal	Nor	–	40	20
2	Gearwheel pitting	GP	Single	40	20
3	Gearwheel crack	GC	Single	40	20
4	Gearwheel wear	GW	Single	40	20
5	Gearwheel tooth breaking	GTB	Single	40	20
6	Bearing ball pitting	BA	Single	40	20
7	Bearing inner pitting	BI	Single	40	20
8	Bearing outer pitting	BO	Single	40	20
9	Gearwheel pitting + Bearing outer pitting	GP + BO	Compound	40	20
10	Bearing inner and outer pitting	IP	Single	40	20

Table 8. The $Value_{AED}$ under different (m, c) combinations.

(m, c)	$Value_{AED}$	(m, c)	$Value_{AED}$
(2,4)	80.99	(3,4)	175.04
(2,5)	162.18	(3,5)	226.21
(2,6)	106.38	(3,6)	220.90
(2,7)	164.04	(3,7)	247.56
(2,8)	121.94	(3,8)	286.67

Secondly, the sensitive feature vectors of training samples are inputted into GWO-RELM, and the final setting optimization parameters are set to $n = 121$, $\theta = 0.91$.

Then, the low-dimensional sensitive feature vectors obtained from the gearbox in ten different states are inputted into the optimized RELM model for training and testing. The final identification result is shown in Figure 21. It can be seen that, among 160 testing samples, only one sample is misidentified (“GTB” is identified as “GW”), and the overall recognition accuracy reaches 99.38%. The comparison results of different feature extraction

models and classification methods are shown in Tables 9 and 10 and Figure 22, and the conclusions obtained are similar to that of experiment 1.

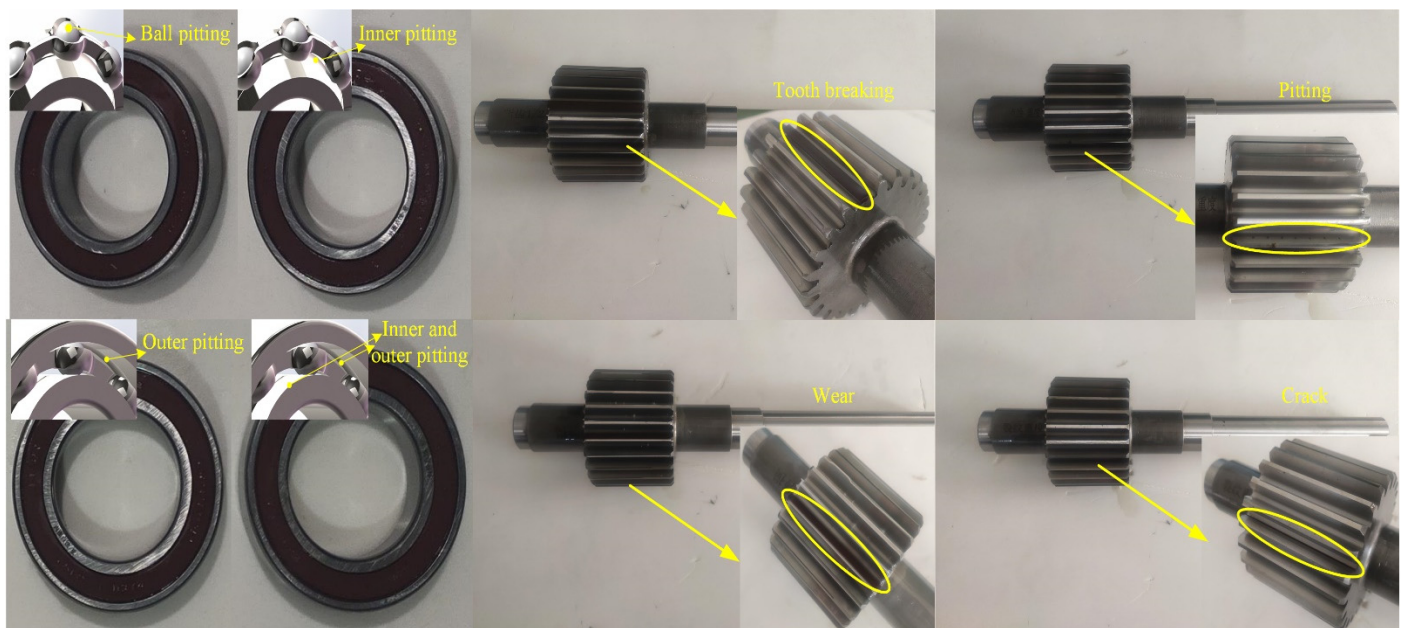


Figure 17. The components with various faults.

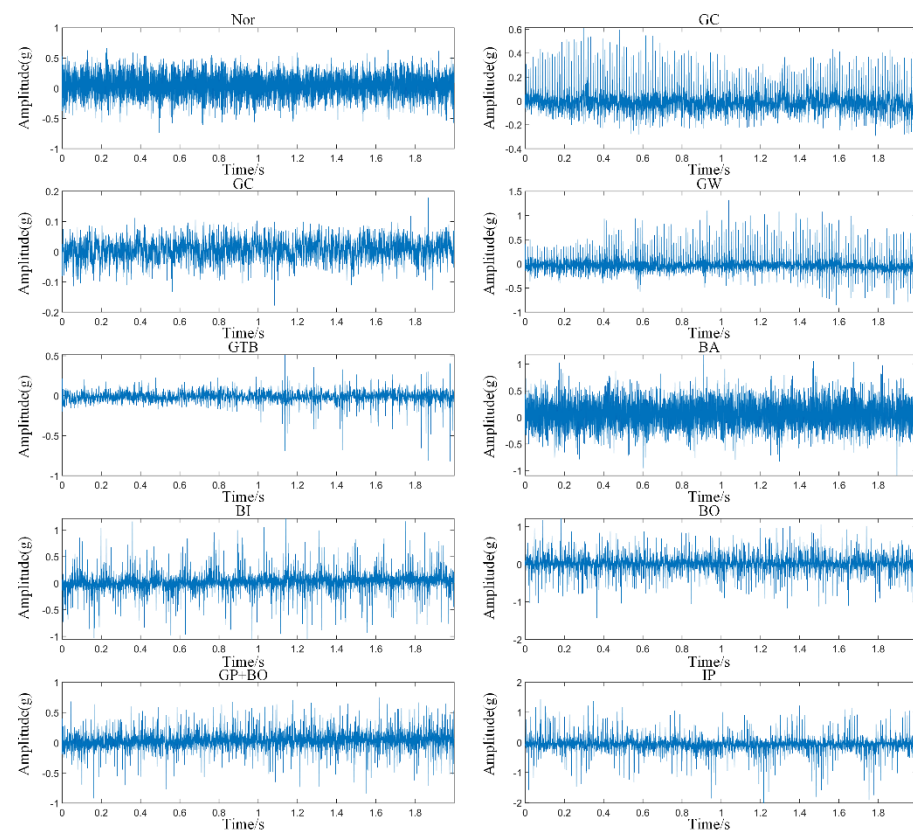


Figure 18. Waveforms corresponding to different states.

Table 9. The performance comparison between different feature extraction models.

Method	Time (s)	Recognition Accuracy (%)			
		Max	Min	Mean	SD
HRCMFDE	1.574	100.00	97.50	98.85	0.60
RCMFDE	0.879	96.88	91.25	93.47	1.79
MFDE	0.018	82.50	71.25	76.65	5.90
RCMDE	1.224	94.34	88.75	91.28	1.51
MDE	0.131	78.13	65.63	72.15	6.25

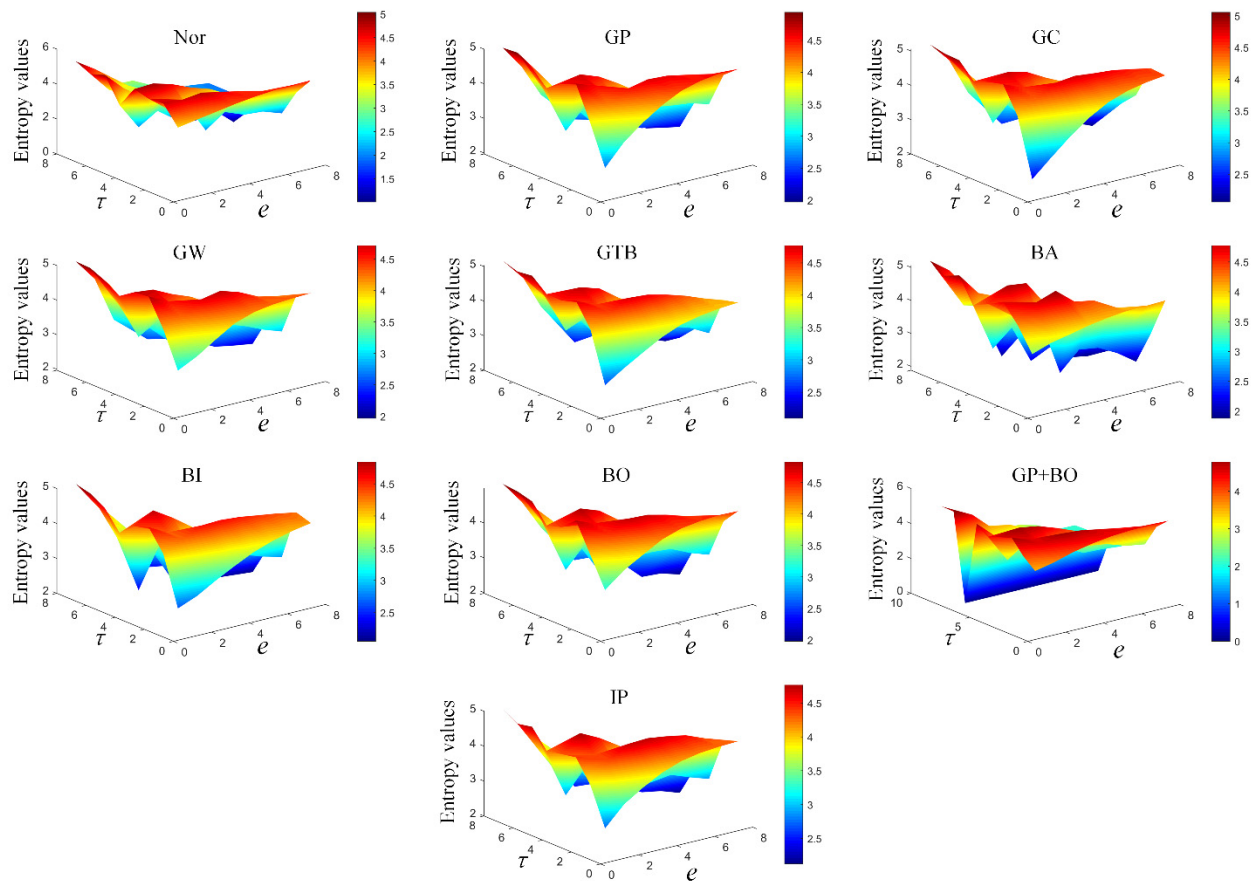
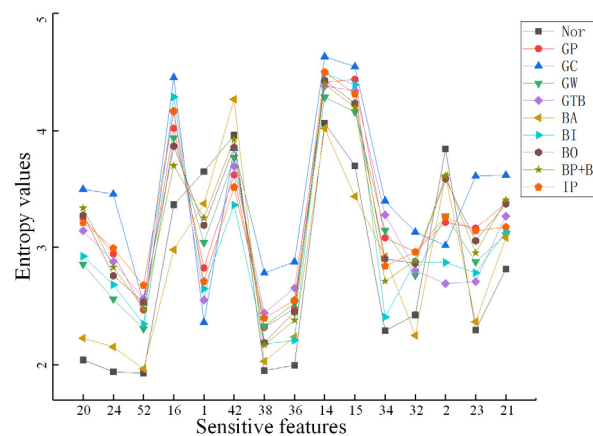
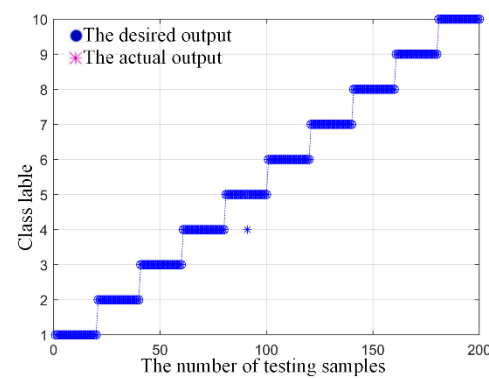
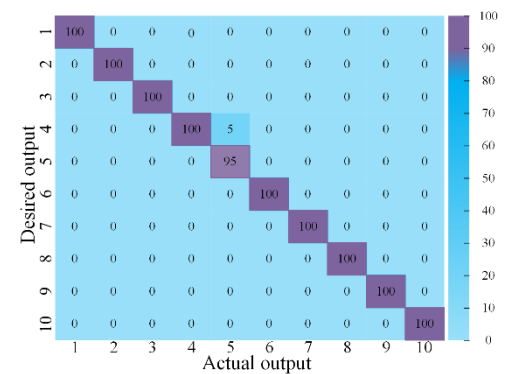
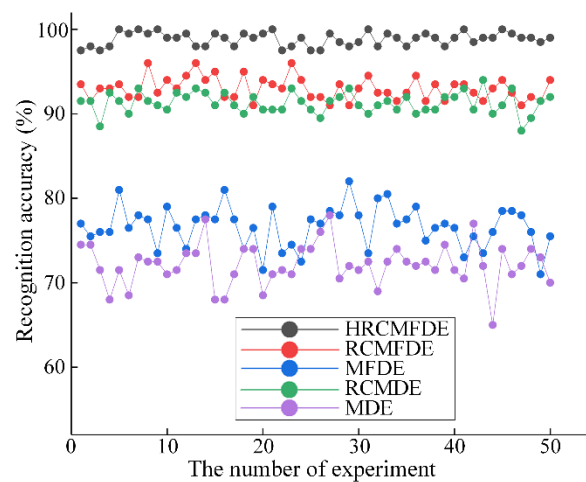
**Figure 19.** Raw fault features corresponding to different states.**Figure 20.** Sensitive fault features of different states.

Table 10. Comparison of models under different classification methods.

Different Classification Methods	HRCMFDE	RCMFDE	MFDE	RCMDE	MDE
mRMR (%)	96.43	92.11	73.24	90.01	71.14
SVM (%)	97.01	92.01	71.99	90.13	70.79
PNN (%)	94.32	86.14	70.41	87.71	67.28
ELM (%)	96.69	91.12	74.11	89.84	69.11
GWO-RELM (%)	98.85	93.47	76.65	91.28	72.15

**(a)** GWO-RELM output.**(b)** Confusion matrix (%).**Figure 21.** Identification results of the proposed method.**Figure 22.** The accuracy of different feature extraction approaches.

Similar to experiment 1, the feature extraction capability of those models is further evaluated, and the results are shown in Table 11. It can be found that the four indexes in the HRCMFDE model have improved and still have better comprehensive performance and stability. In addition, the HRCMFDE model performs better in the feature extraction of bearing faults than gear faults.

Aiming at the compound gear train experimental platform, it is shown that this method can effectively identify the running state of the fault and provide a diagnosis method for the state monitoring of the experimental platform, which offers a basis for other studies on the experimental platform.

4.3. Experiment 3: Fault Diagnosis of Planetary Gear Train Gearboxes

In experiment 1 and experiment 2, the proposed method is used to realize the state identification of two different types of gearboxes, respectively. The results are satisfactory, which proves the application potential of the presented approach in the field of gearbox fault diagnosis.

Table 11. Comparison of different models.

Status	HRCMFDE				RCMFDE				RCMDE			
	P (%)	R (%)	Acc (%)	F1 (%)	P (%)	R (%)	Acc (%)	F1 (%)	P (%)	R (%)	Acc (%)	F1 (%)
G1	99.76	100.00	99.98	99.88	100.00	100.00	100.00	100.00	100.00	99.50	99.95	99.74
G2	96.46	98.25	99.45	97.30	91.80	75.50	96.85	82.75	83.16	70.25	95.55	75.94
G3	100.00	99.75	99.98	99.87	87.97	91.50	97.88	89.63	85.54	90.25	97.48	87.77
G4	99.02	95.25	99.43	97.02	100.00	90.50	99.05	94.87	87.78	83.50	97.18	85.46
G5	97.85	97.25	99.50	97.49	82.87	75.50	95.98	78.94	77.95	69.25	94.95	73.24
G6	100.00	99.25	99.93	99.62	100.00	100.00	100.00	100.00	99.52	100.00	99.95	99.76
G7	100.00	100.00	100.00	100.00	76.29	100.00	96.88	86.52	84.71	100.00	98.18	91.68
G8	97.49	100.00	99.73	98.69	100.00	100.00	100.00	100.00	93.79	100.00	99.33	96.76
G9	99.76	99.25	99.90	99.49	99.76	100.00	99.98	99.88	100.00	100.00	100.00	100.00
G10	99.31	100.00	99.93	99.64	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OM	98.97	98.90	99.78	98.90	93.87	93.30	98.66	93.26	91.24	91.28	98.26	91.04

In this experiment, the planetary gearbox data from Southeast University is taken as an example to verify further the effectiveness of the proposed method [42]. These data are collected from the Drivetrain Dynamic Simulator (DDS). There is a classic planetary gearbox fault diagnosis state simulation experimental platform employed by many scholars to research the fault diagnosis method. A detailed description of this experiment is shown in [42]. This paper uses the data of channel 2, and the working condition is 20 HZ-0 V. The different fault types of gears and bears in the gearbox are shown in Table 12.

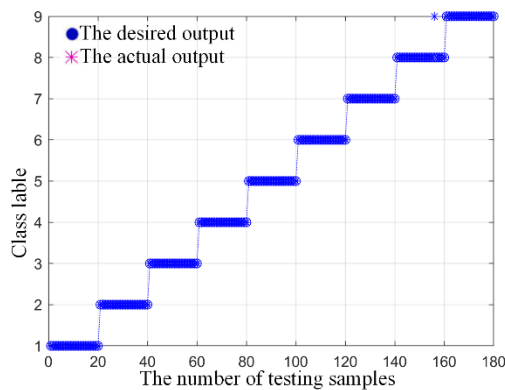
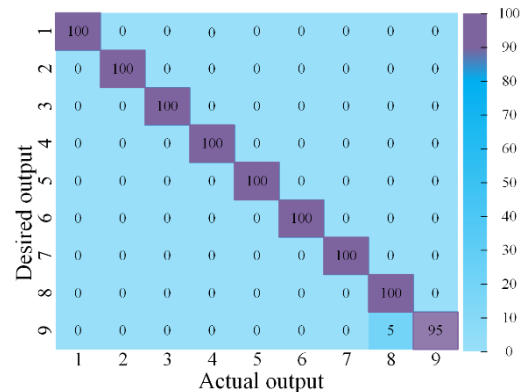
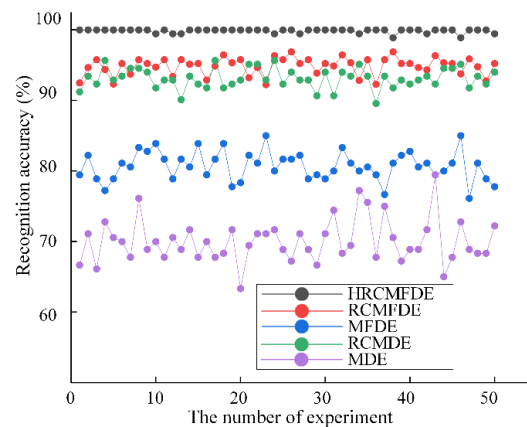
Table 12. The detailed fault information of gearbox components.

Labels	Status	Abbreviation	Description
1	Normal	Nor	Healthy status
2	Bearing Ball	BB	Crack occurs in the ball
3	Bearing Combination	BC	Crack occurs in both the inner and outer ring
4	Bearing Inner	BI	Crack occurs in the inner ring
5	Bearing Outer	BO	Crack occurs in the outer ring
6	Gearwheel Chipped	GC	Crack occurs in the gear feet
7	Gearwheel Miss	GM	Missing one foot in the gear
8	Gearwheel Root	GR	Crack occurs in the root of gear feet
9	Gearwheel Surface	GS	Wear occurs in the surface of gear

In the experiment, 60 samples are selected for each group, with 40 samples as the training samples and 20 samples as the testing samples. The parameter determination process in the proposed method is consistent with that in experiment 1 and experiment 2. The final parameters are set to $k = 3$, $\tau = 8$, $\lambda = 1$, $N = 2048$, $m = 3$, $c = 8$, $n = 113$, $\theta = 0.72$. Using the fault diagnosis method proposed in Section 3.5, the final identification result is shown in Figure 23. It can be seen that, among 180 testing samples, only one sample is misidentified (“GS” is identified as “GR”), and the overall accuracy reaches 99.44%. The comparison results of different methods are shown in Table 13 and Figure 24. In addition, the comparison of recognition results of the feature extraction model under different classification methods is shown in Table 14. It can be seen that the proposed method still has a more precise diagnosis accuracy and a more stable performance.

Table 13. The performance comparison between different methods.

Method	Time (s)	Recognition Accuracy (%)			
		Max	Min	Mean	SD
HRCMFDE	1.391	100.00	98.89	99.89	0.08
RCMFDE	0.803	96.89	92.22	94.86	1.79
MFDE	0.018	85.00	76.11	80.71	4.22
RCMDE	1.129	95.56	89.67	93.13	2.03
MDE	0.124	79.44	63.33	70.06	9.49

**(a)** GWO-RELM output.**(b)** Confusion matrix (%).**Figure 23.** Identification results of the proposed method.**Figure 24.** The accuracy of different approaches.**Table 14.** Comparison of models under different classification methods.

Different Classification Methods	HRCMFDE	RCMFDE	MFDE	RCMDE	MDE
mRMR (%)	98.32	91.76	78.58	92.25	67.74
SVM (%)	97.99	93.74	78.07	91.89	66.62
PNN (%)	94.01	89.14	75.41	89.80	64.20
ELM (%)	96.89	93.47	77.71	92.14	65.35
GWO-RELM (%)	99.89	94.86	80.71	93.13	70.06

The HRCMFDE, RCMFDE and RCMDE models with better feature extraction performance are compared, and the results are shown in Table 15. The method proposed can fully identify testing samples of six different states. Meanwhile, the remaining methods have no way of completely identifying testing samples of any states. Compared with the RCMFDE model, the P-means, R-means, Acc-means and F-means are increased by 4.70%,

4.89%, 1.09% and 4.89%, respectively. Once again, the superior feature extraction ability of the HRCMFDE model is highlighted.

This section uses gearbox signals from three different types to discuss the performance and generalization of the presented method in the field of the gearbox. In experiment 1, the effectiveness and stability of the proposed method in feature extraction and sensitive information screening for single and compound faults are verified by the reverted gear train gearbox vibration signals. In experiment 2, the proposed method is applied to the constructed fault simulation platform, which proves that it still has excellent stability and diagnostic accuracy. In experiment 3, the proposed method is successfully used for planetary gearbox fault diagnosis. Analyzing the results of the three experiments shows that this method has excellent practicability and superior performance and can effectively identify single or compound faults in the gearbox. Meanwhile, the method possesses good generalization performance and is suitable for various structural types of gearboxes.

Table 15. Comparison of different models.

Status	HRCMFDE				RCMFDE				RCMDE			
	P (%)	R (%)	Acc (%)	F1 (%)	P (%)	R (%)	Acc (%)	F1 (%)	P (%)	R (%)	Acc (%)	F1 (%)
G1	100.00	100.00	100.00	100.00	100.00	93.25	99.25	96.48	100.00	92.00	99.11	95.81
G2	100.00	100.00	100.00	100.00	89.92	95.00	98.25	92.37	98.59	100.00	99.83	99.27
G3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.52	99.75	99.92	99.63
G4	100.00	100.00	100.00	100.00	93.65	95.25	98.75	94.43	95.28	92.75	98.67	93.95
G5	100.00	100.00	100.00	100.00	95.00	95.00	98.89	94.99	99.51	96.75	99.58	98.09
G6	100.00	100.00	100.00	100.00	94.93	88.25	98.17	91.42	91.16	99.75	98.89	95.24
G7	100.00	100.00	100.00	100.00	87.78	95.50	98.00	91.43	83.55	69.25	95.06	75.57
G8	99.76	99.25	99.89	99.49	100.00	99.75	99.97	99.87	90.74	98.00	98.67	94.22
G9	99.29	99.75	99.89	99.51	95.40	93.00	98.72	94.15	81.79	90.50	96.67	85.82
OM	99.89	99.89	99.98	99.89	95.19	95.00	98.89	95.02	93.35	93.19	98.49	93.07

5. Conclusions

In this research, a novel fault diagnosis approach based on HRCMFDE, ReliefF and GWO-RELM is developed and applied to gearbox fault diagnosis. The effectiveness and superiority of the proposed method compared with existing methods are verified, and the generalization of the method is discussed on various gearbox structures. The main conclusions of this paper can be summarized as follows:

- (1) In view of the problem of poor stability and insufficient feature information extraction, a feature extraction method of HRCMFDE is proposed based on the existing techniques, which can effectively extract fault features of gearbox vibration signals at different hierarchical layers and scales;
- (2) ReliefF is utilized to screen sensitive information of high-dimensional information and remove redundant features. GWO-RELM is used to identify the health conditions of the gearbox. Combined with HRCMFDE, ReliefF and GWO-RELM, a novel gearbox fault diagnosis method is proposed. The method is verified by the gearbox fault data set. It shows that the proposed method has superior fault diagnosis performance and can accurately diagnose different working states of gearbox bearings and gears. The proposed method has better diagnostic accuracy and stability than the existing RCMFDE, MFDE, RCMDE and MDE methods;
- (3) The main structural types of gearboxes in practical applications have been tested. The methods proposed in each experiment can effectively identify the fault running state and have been successfully applied to the fault diagnosis of various gearboxes. The results show that this method has excellent practicability and generality, so it can widely apply to gearbox fault diagnosis.

In this preliminary study, the proposed approach is satisfactory and promising in the health condition identification of the gearbox. Moreover, it can be extended to other

complex mechanical systems, such as rotating machinery, hydraulic pumps, etc. In our future work, this research direction will be conducted.

Author Contributions: Conceptualization, H.L. and W.Z.; methodology, Z.L. and W.Z.; software, J.Z.; validation, Z.L. and Y.Z.; data curation, J.M. and Y.W. (Yuzhan Wei); writing-original draft, W.Z.; writing-review and editing, Y.Z., H.L. and Y.W. (Yongjing Wang). All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (52275505) and supported by the research (JZX7Y20220144100101).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Dou, D.; Yang, J.; Liu, J.; Zhao, Y. A rule-based intelligent method for fault diagnosis of rotating machinery. *Knowl.-Based Syst.* **2012**, *36*, 1–8. [\[CrossRef\]](#)
2. Kafeel, A.; Aziz, S.; Awais, M.; Khan, M.A.; Afaq, K.; Idris, S.A.; Mostafa, S.M. An expert system for rotating machine fault detection using vibration signal analysis. *Sensors* **2021**, *21*, 7587. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Song, Y.; Zhong, M.; Xue, T.; Ding, S.; Li, W. Parity space-based fault isolation using minimum error minimax probability machine. *Control Eng. Pract.* **2020**, *95*, 104242. [\[CrossRef\]](#)
4. Kong, Y.; Wang, T.; Chu, F.; Feng, Z.; Selesnick, I. Discriminative dictionary learning-based sparse classification framework for data-driven machinery fault diagnosis. *IEEE Sens. J.* **2021**, *21*, 8117–8129. [\[CrossRef\]](#)
5. Yu, W.; Zhao, C. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE Trans. Ind. Electron.* **2019**, *67*, 5081–5091. [\[CrossRef\]](#)
6. Yu, J. Evolutionary manifold regularized stacked denoising autoencoders for gearbox fault diagnosis. *Knowl.-Based Syst.* **2019**, *178*, 111–122. [\[CrossRef\]](#)
7. Huang, W.; Kong, F.; Zhao, X. Spur bevel gearbox fault diagnosis using wavelet packet transform and rough set theory. *J. Intell. Manuf.* **2018**, *29*, 1257–1271. [\[CrossRef\]](#)
8. Luo, C.; Mo, Z.; Miao, Q. Cyclic harmonic ratio defined in squared envelope spectrum and log-envelope spectrum for gearbox fault diagnosis. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9568–9577. [\[CrossRef\]](#)
9. Inturi, V.; Pratyush, A.S.; Sabareesh, G.R. Detection of local gear tooth defects on a multistage gearbox operating under fluctuating speeds using DWT and EMD analysis. *Arab. J. Sci. Eng.* **2021**, *46*, 11999–12008. [\[CrossRef\]](#)
10. Miao, Y.; Zhao, M.; Yi, Y.; Lin, J. Application of sparsity-oriented VMD for gearbox fault diagnosis based on built-in encoder information. *ISA Trans.* **2020**, *99*, 496–504. [\[CrossRef\]](#)
11. Wang, L.; Cai, G.; Wang, J.; Jiang, X.; Zhu, Z. Dual-enhanced sparse decomposition for wind turbine gearbox fault diagnosis. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 450–461. [\[CrossRef\]](#)
12. Minhas, A.S.; Singh, S. A new bearing fault diagnosis approach combining sensitive statistical features with improved multiscale permutation entropy method. *Knowl.-Based Syst.* **2021**, *218*, 106883. [\[CrossRef\]](#)
13. Wang, L.; Chen, G.; Shi, D.; Chang, Y.; Chan, S.; Pu, J.; Yang, X. Active contours driven by edge entropy fitting energy for image segmentation. *Signal Process.* **2018**, *149*, 27–35. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Wei, Y.; Yang, Y.; Xu, M.; Huang, W. Intelligent fault diagnosis of planetary gearbox based on refined composite hierarchical fuzzy entropy and random forest. *ISA Trans.* **2021**, *109*, 340–351. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Purvis, B.; Mao, Y.; Robinson, D. Entropy and its application to urban systems. *Entropy* **2019**, *21*, 56. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Azami, H.; Rostaghi, M.; Abásolo, D.; Escudero, J. Refined composite multiscale dispersion entropy and its application to biomedical signals. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2872–2879.
17. 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. [\[CrossRef\]](#)
18. Tran, M.Q.; Elsis, M.; Liu, M.K. Effective feature selection with fuzzy entropy and similarity classifier for chatter vibration diagnosis. *Measurement* **2021**, *184*, 109962. [\[CrossRef\]](#)
19. Ruiz-Aguilar, J.J.; Turias, I.; González-Enrique, J.; Urda, D.; Elizondo, D. A permutation entropy-based EMD-ANN forecasting ensemble approach for wind speed prediction. *Neural Comput. Appl.* **2021**, *33*, 2369–2391. [\[CrossRef\]](#)
20. Feng, K.; Wang, K.; Ni, Q.; Zuo, M.; Wei, D. A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions. *J. Sound Vib.* **2017**, *408*, 190–209. [\[CrossRef\]](#)
21. Wei, Y.; Li, Y.; Xu, M.; Huang, W. Intelligent fault diagnosis of rotating machinery using ICD and generalized composite multiscale fuzzy entropy. *IEEE Access* **2018**, *7*, 38983–38995. [\[CrossRef\]](#)

22. Kuai, M.; Cheng, G.; Pang, Y.; Li, Y. Research of planetary gear fault diagnosis based on permutation entropy of CEEMDAN and ANFIS. *Sensors* **2018**, *18*, 782. [CrossRef] [PubMed]
23. Azami, H.; Escudero, J. Amplitude- and frequency-based dispersion patterns and entropy. *Entropy* **2018**, *20*, 210. [CrossRef] [PubMed]
24. Azami, H.; Arnold, S.; Sanei, S.; Chang, Z.; Sapiro, G.; Escudero, J. Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. *IEEE Access* **2019**, *7*, 68718–68733. [CrossRef]
25. Zhang, W.; Zhou, J. A comprehensive fault diagnosis method for rolling bearings based on refined composite multiscale dispersion entropy and fast ensemble empirical mode decomposition. *Entropy* **2019**, *21*, 680. [CrossRef]
26. Zhou, F.; Yang, X.; Shen, J.; Liu, W. Fault diagnosis of hydraulic pumps using PSO-VMD and refined composite multiscale fluctuation dispersion entropy. *Shock. Vib.* **2020**, *2020*, 8840676. [CrossRef]
27. Yan, X.; Liu, Y.; Huang, D.; Jia, M. A new approach to health condition identification of rolling bearing using hierarchical dispersion entropy and improved Laplacian score. *Struct. Health Monit.* **2021**, *20*, 1169–1195. [CrossRef]
28. Ke, Y.; Yao, C.; Song, E.; Dong, Q.; Yang, L. An early fault diagnosis method of common-rail injector based on improved CYCBD and hierarchical fluctuation dispersion entropy. *Digit. Signal Process.* **2021**, *114*, 103049. [CrossRef]
29. Yan, X.; Jia, M. Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection. *Knowl.-Based Syst.* **2019**, *163*, 450–471. [CrossRef]
30. Robnik-Sikonja, M.; Kononenko, I. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Mach. Learn.* **2003**, *53*, 23–69. [CrossRef]
31. Eshtay, M.; Faris, H.; Obeid, N. Improving Extreme Learning Machine by Competitive Swarm Optimization and its application for medical diagnosis problems. *Expert Syst. Appl.* **2018**, *104*, 134–152. [CrossRef]
32. Mirjalili, S.; Saremi, S.; Mirjalili, S.M.; Coelho, L.D.S. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. *Expert Syst. Appl.* **2016**, *47*, 106–119. [CrossRef]
33. Liang, X.; Zuo, M.J.; Feng, Z. Dynamic modeling of gearbox faults: A review. *Mech. Syst. Signal Process.* **2018**, *98*, 852–876. [CrossRef]
34. Yang, C.; Jia, M. Hierarchical multiscale permutation entropy-based feature extraction and fuzzy support tensor machine with pinball loss for bearing fault identification. *Mech. Syst. Signal Process.* **2021**, *149*, 107182. [CrossRef]
35. Zhou, F.; Gong, J.; Yang, X.; Han, T.; Yu, Z. A new gear intelligent fault diagnosis method based on refined composite hierarchical fluctuation dispersion entropy and manifold learning. *Measurement* **2021**, *186*, 110136. [CrossRef]
36. Rostaghi, M.; Azami, H. Dispersion Entropy: A Measure for Time-Series Analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [CrossRef]
37. Shariati, M.; Mafipour, M.S.; Ghahremani, B.; Azarhomayun, F.; Ahmadi, M.; Trung, N.T.; Shariati, A. A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement. *Eng. Comput.* **2022**, *38*, 757–779. [CrossRef]
38. Huang, G.; Zhu, Q.; Siew, C. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
39. Wu, Z.; Wang, X.; Jiang, B. Fault diagnosis for wind turbines based on ReliefF and extreme gradient boosting. *Appl. Sci.* **2020**, *10*, 3258. [CrossRef]
40. PHM Data Challenge. 2009. Available online: <https://www.phmsociety.org/competition/PHM/09> (accessed on 24 April 2016).
41. Zhang, X.; Han, P.; Xu, L.; Zhang, F.; Wang, Y.; Gao, L. Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM. *IEEE Access* **2020**, *8*, 192248–192258. [CrossRef]
42. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2446–2455. [CrossRef]