

Article

A Fault Diagnosis Method of Rolling Bearing Based on Attention Entropy and Adaptive Deep Kernel Extreme Learning Machine

Weiyu Wang ^{1,2} , Xunxin Zhao ^{1,2}, Lijun Luo ^{1,2}, Pei Zhang ^{1,2}, Fan Mo ^{1,2}, Fei Chen ^{3,*} , Diyi Chen ³, Fengjiao Wu ³ and Bin Wang ^{3,*} 

¹ Wuling Power Corporation Ltd., Changsha 410004, China

² Hydropower Industry Innovation Center of State Power Investment Corporation Limited, Changsha 410004, China

³ Department of Power and Electrical Engineering, Northwest A&F University, Xianyang 712100, China

* Correspondence: chenfei@nwfau.edu.cn (F.C.); binwang@nwsuaf.edu.cn (B.W.)

Abstract: To address the difficulty of early fault diagnosis of rolling bearings, this paper proposes a rolling bearing diagnosis method by combining the attention entropy and adaptive deep kernel extreme learning machine (ADKELM). Firstly, the wavelet threshold denoising method is employed to eliminate the noise in the vibration signal. Then, the empirical wavelet transform (EWT) is utilized to decompose the denoised signal, and extract the attention entropy of the intrinsic mode function (IMF) as the feature vector. Next, the hyperparameters of the deep kernel extreme learning machine (DKELM) are optimized using the marine predators algorithm (MPA), so as to achieve the adaptive changes in the DKELM parameters. By analyzing the fault diagnosis performances of the ADKELM model with different kernel functions and hidden layers, the optimal ADKELM model is determined. Compared with conventional machine learning models such as extreme learning machine (ELM), back propagation neural network (BPNN) and probabilistic neural network (PNN), the high efficiency of the method proposed in this paper is verified.

Keywords: rolling bearing; fault diagnosis; empirical wavelet transform; attention entropy; marine predators algorithm; deep kernel extreme learning machine



Citation: Wang, W.; Zhao, X.; Luo, L.; Zhang, P.; Mo, F.; Chen, F.; Chen, D.; Wu, F.; Wang, B. A Fault Diagnosis Method of Rolling Bearing Based on Attention Entropy and Adaptive Deep Kernel Extreme Learning Machine. *Energies* **2022**, *15*, 8423. <https://doi.org/10.3390/en15228423>

Academic Editor: Helena M. Ramos

Received: 21 September 2022

Accepted: 8 November 2022

Published: 10 November 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 key component of rotating machinery, the safe operation of rolling bearings directly affects the operating efficiency of rotating machinery, and timely detection of the bearing faults has important practical significance to maintaining the operation of rotating machinery [1].

Generally speaking, rolling bearing faults are mostly caused by surface defects of the inner ring, outer ring and rolling element. When the machine runs with these faults, there will be strong pulse signals, and these signals are used as important indicators to evaluate the degree of mechanical faults [2]. Considering that the vibration signal is the carrier of fault information, the fault diagnosis based on the vibration signals has become an important approach, and fault diagnosis generally includes the steps signal processing, feature extraction and pattern recognition [3]. The vibration signal fault signals are generally not very stable, which affects the extraction of the fault features [4]. Therefore, the decomposition methods, such as empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD), have been widely used in the bearing signal processing [5,6], structural safety level assessment [7] and high-speed railway grid fault identification [8]. For example, in [6], based on EEMD and singular value entropy theory, singular value entropy is utilized to effectively distinguish different bearing fault states. Although EMD and EEMD have achieved great performances in processing of fault signals, they also have various problems, such as modal aliasing, insufficient theoretical foundation, and inability to select modal component independently. To address these problems, the empirical wavelet

transform (EWT) was proposed. The EWT algorithm combined the advantages of wavelet transform and EMD, the signal spectrum is adaptively segmented, and appropriate empirical wavelet filter banks are constructed according to Meyer wavelet construction method to extract empirical mode components (EWFs) with tight support, effectively avoiding mode aliasing and end effect [9]. Because of its good adaptability and low computational complexity, EWT has been widely used in the field of fault diagnosis [10–14]. In this paper, EWT is introduced to decompose the bearing fault signal, and reduce the influence of signal instability on feature extraction.

Feature extraction is a key step in the bearing fault diagnosis, and the quality of extracted features directly affects the performance of fault recognition [15]. In the field of fault diagnosis, different entropies are often used as the features of the fault signals and wait to be extracted. For example, Liu et al. [16] and Zair et al. [17] used the sample entropy and fuzzy entropy as the features of vibration signals to effectively distinguish different classes of bearing faults. However, it takes a long time to process long time series when the sample entropy and fuzzy entropy are used [18]. Compared with the above two entropies, the permutation entropy proposed by Bandt [19] is widely used in fault diagnosis due to its simple computation [20,21]. However, the permutation entropy tends to ignore the difference between signal amplitudes, which tends to cause the loss of effective information. To overcome the shortcomings of the above entropies, Yang et al. proposed the attention entropy—a new measure of signal complexity [22]. Different from the traditional entropy which focuses on the frequency distribution of all data in a time series, the attention entropy only focuses on the frequency distribution in the interval between peaks in the time series. Therefore, the attention entropy has the advantages of not need for parameter adjustment, fast computation speed, and strong robustness to the length of time series. This paper puts forward a feature extraction method combining EWT and the attention entropy, which uses EWT to decompose the vibration signal, and then extracts the attention entropy of the intrinsic mode function (IMF) as the feature vector.

Fault diagnosis is essentially pattern recognition. Utilizing the fault features, the traditional classifier algorithms, such as support vector machines (SVM) and artificial neural network (ANN), have been widely used in fault pattern recognition [23–27]. However, both SVM and ANN are shallow neural networks, and their diagnostic accuracy is often unsatisfactory. Compared with SVM and ANN, the extreme learning machine (ELM) can provide better performances in terms of the training speed and generalization ability. To address the classification accuracy problem caused by random initialization of the ELM algorithm, Huang et al. [28] proposed the kernel extreme learning machine (KELM), which uses kernel mapping to replace random mapping, thus effectively improving the performance of the ELM model. With the continuous development of deep learning, various deep learning models have been successively developed, and the classic models include deep auto-encoders (DAE), convolutional neural network (CNN) and deep belief network (DBN). Among them, DAE is an unsupervised feature learning model, which extracts the deep features of the input data by transforming the feature space of the input data [29]. The deep extreme kernel learning machine (DKELM) is a deep neural network model based on KELM by combining the DAE model [30]. Compared with the KELM model, DKELM can mine the feature information at a deeper level, so as to improve the accuracy of the model. Therefore, DKELM has been put into broad applications in the financial market prediction [31], hyperspectral image classification [32], multi-classification problems [33], and water quality prediction [34].

Even though the DKELM model can mine the deep features of the data, the settings of the hyperparameters, such as the number of hidden layer nodes, the regularization parameters of the hidden layer, the kernel parameters, and the kernel function penalty coefficient, lead to highly random accuracy of the DKELM model. To address this problem, this paper introduces the marine predators algorithm (MPA) [35] to optimize the hyperparameters of the DKELM model, so as to achieve the adaptive changes in the parameters of the DKELM model, and significantly reduce the adjustment time of the DKELM parameters.

By analyzing the fault diagnosis performances of the ADKELM model with different kernel functions and different hidden layers, the optimal ADKELM model is determined. The simulation results show that the ADKELM model proposed in this paper outperforms the DKELM model.

This paper proposes a bearing fault diagnosis method by combining the attention entropy and ADKELM model. First, to address the problem that high noise covers effective fault signals, the wavelet threshold denoising method is used to effectively eliminate the influence of noise. Second, EWT is used to decompose the noise reduction signal into IMF of different frequencies, and the impulse values of different IMF are captured through the attention entropy. At the same time, the mentioned attention entropy is used as the feature vector. Then, the MPA algorithm is introduced to optimize the hyperparameters of the DKELM model, so as to realize the adaptive adjustment of the parameters of the DKELM model. Finally, with the efficient fault feature capture capability of AE and powerful recognition performance of ADKELM, accurate recognition of rolling bearing faults can be realized. The results of simulation experiments show that the proposed method can achieve great diagnosis accuracy.

Section 2 mainly introduces the basic principle of the algorithm. Section 3 describes the simulation experiments. Finally, Section 4 draws the conclusions.

2. Fault Diagnosis Method of EWT-AE-ADKELM

2.1. Empirical Wavelet Transform (EWT)

Gilles [9] proposed the EWT based on the wavelet analysis framework. Reasonable segmentation of the signal spectrum is critical to EWT. A set of wavelet filters can be constructed to further extract different AM-FM signals in the signal. Suppose that the signal Fourier support $[0, \pi]$ is divided into N continuous parts, and the midpoint ω_n ($\omega_0 = 0, \omega_n = \pi$) between the adjacent maximum points is used as the boundary of the segments. Then, the n -th segment can be expressed as:

$$\begin{cases} \Lambda_n = [\omega_{n-1}, \omega_n], n = 1, 2, \dots, N \\ \cup_{n=1}^N \Lambda_n = [0, \pi] \end{cases} \quad (1)$$

where Λ_n represents the n -th segment frequency band.

Based on the Λ_n segmented, a band-pass filter is constructed. According to the construction idea of Meyer wavelet, the empirical wavelet function $\hat{\phi}_n(\omega)$ and the empirical scale function $\hat{\psi}(\omega)$ can be obtained by Equations (2) and (3):

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & \omega_n - \tau_n \geq |\omega| \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right], & \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$\hat{\psi}(\omega) = \begin{cases} 1, & \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1} \\ \cos \left(\frac{\pi}{2} \beta \left(\frac{1}{2\tau_{n+1}} (|\omega| - \omega_n + \tau_n) \right) \right), & \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_{n+1} + \tau_{n+1} \\ \sin \left(\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} (|\omega| - \omega_{n+1} + \tau_{n+1}) \right) \right), & \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $\tau_n = \gamma \omega_n$, $0 < \gamma < 1$; $\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right)$. In Equation (3), $\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

According to the construction idea of wavelet transform, the detail coefficient of the empirical wavelet transform of signal $x(t)$ can be obtained by Equation (4):

$$W_f^e(n, t) = \langle x(t), \psi_n(t) \rangle = \int x(\tau) \overline{\psi_n(\tau - t)} dt = F^{-1}[x(\omega) \hat{\psi}_n(\omega)] \quad (4)$$

where $W_f^e(n, t)$ represents the detail coefficient; $\langle \cdot \rangle$ stands for the inner product; $F^{-1}[\cdot]$ represents the inverse Fourier transform; $\hat{\psi}_n(\omega)$ is the Fourier transform of $\psi_n(\omega)$; $\overline{\psi_n(t)}$ is the complex conjugate function of $\psi_n(t)$.

The approximate coefficient calculation method of the empirical wavelet transform is as follows:

$$W_f^e(0, t) = \langle x(t), \varphi_1(t) \rangle = \int x(\tau) \overline{\varphi_1(\tau - t)} dt = F^{-1}[x(\omega) \hat{\varphi}_1(\omega)] \quad (5)$$

where $W_f^e(0, t)$ is the approximate coefficient; $\hat{\varphi}_1(\omega)$ represents the Fourier transform of $\varphi_1(\omega)$; $\overline{\varphi_1(t)}$ is the complex conjugate function of $\varphi_1(t)$.

The reconstructed signal of the original signal $x(t)$ is as follows:

$$x(t) = W_f^e(0, t) * \varphi_1(t) + \sum_{n=1}^N W_f^e(n, t) * \psi_n(t) = F^{-1} \left[\hat{W}_f^e(0, \omega) \hat{\varphi}_1(\omega) + \sum_{n=1}^N \hat{W}_f^e(n, \omega) \hat{\psi}_n(\omega) \right] \quad (6)$$

where $*$ represents the convolution symbol. According to Equation (6), the empirical mode $x_k(t)$ can be obtained by the empirical wavelet decomposition:

$$\begin{cases} x_0(t) = W_f^e(0, t) * \varphi_1(t) \\ x_k(t) = W_f^e(k, t) * \psi_k(t) \end{cases} \quad (7)$$

2.2. Attention Entropy (AE)

Unlike the traditional entropy which analyzes the frequency distribution of all signal points in the time series, the attention entropy effectively distinguish different time series signals by analyzing the frequency distribution in the interval between key data points in the signal [22]. Therefore, this paper uses the empirical mode components obtained by the EWT decomposition, calculates the attention entropy of the component and uses it as the feature vector. The attention entropy generally consists of the following three steps:

- (1) Defined the key patterns.
- (2) Calculated the interval between two adjacent key patterns.
- (3) Calculated the Shannon entropy of the interval.

Each point in the vibration signal can be regarded as a state of the system. The peak points represent the local upper and lower limits of the state, which make them the potential key modes. Therefore, the peak is defined as the potential critical mode point. Based on the following four different strategies: the interval from local maximum to local maximum, the interval from local maximum to local minimum, the interval from local minimum to local maximum, and the interval from local minimum to local minimum, the entropy of the interval distribution of key mode points is calculated using the Shannon entropy formula.

$$H(x) = - \sum_{i=1}^n p(x_i) \log_2(p(x_i)) \quad (8)$$

where $H(x)$ represents the Shannon entropy of the signal; $p(x_i)$ represents the probability of interval distribution, and its expression is $p(x_i) = x_i / \sum_{i=1}^n x_i$; x_i represents the number

of intervals between the i -th key modes. Finally, the average value of the entropies under the four strategies is obtained and used as the attention entropy.

$$AE(x) = \sum_{j=1}^4 H_j(x) \quad (9)$$

In Equation (9), $AE(x)$ represents the attention entropy of the signal; $H_j(x)$ represents the Shannon entropy of the signal under the j -th strategy.

2.3. Adaptive Deep Extreme Kernel Learning Machine (ADKELM)

2.3.1. Extreme Learning Machine (ELM)

As a feedforward neural network, the extreme learning machine is different from the BP neural network which needs to repeatedly adjust the weights and biases of the input layer and the hidden layer. ELM randomly selects the weights and biases of the input layer and the hidden layer, and according to the principle of least squares, the weight between the hidden layer and the output layer is directly determined. The sample dataset $\{X_N, Y_N\}$ is input into the ELM, where $X_n = [x_{n1}, x_{n2}, \dots, x_{ns}] \in \mathfrak{R}^{1 \times s}$ is the input vector, and $Y_n = [x_{n1}, x_{n2}, \dots, x_{nm}] \in \mathfrak{R}^{1 \times m}$ is the output vector, N is the number of samples. The ELM output can be obtained by Equation (10):

$$\sum_{j=1}^L \beta_j g(W_j \cdot X_n + b_j) = o_t \quad (10)$$

where β_j is output weight of the j -th hidden node; $g(\cdot)$ represents the activation function; W_j is input weight of the j -th hidden node; b_j is input bias of the j -th hidden node. The training objective of the ELM algorithm is to minimize the error between the actual output and the expected output.

$$\sum_{j=1}^L \beta_j g(W_n \cdot X_n + b_n) = o_n, \quad n = 1, \dots, N \quad (11)$$

Rewrite Equation (11) into the matrix representation:

$$H\beta = Y \quad (12)$$

In Equation (12), $H = \begin{bmatrix} g(W_1 X_1 + b_1) & \cdots & g(W_L X_1 + b_L) \\ \vdots & & \vdots \\ g(W_1 X_N + b_1) & \cdots & g(W_L X_N + b_L) \end{bmatrix}_{N \times L}$, and H represents

the output matrix of the hidden layer node; the output weight matrix is $[\beta_1, \beta_2, \dots, \beta_L]_{L \times m}^T$; the expected output matrix Y is $[Y_1, Y_2, \dots, Y_N]_{N \times m}^T$.

By solving Equation (12), the output weight matrix β can be obtained as:

$$\beta = H^\dagger Y \quad (13)$$

where H^\dagger is the Moore–Penrose generalized inverse of the output matrix H . To improve the generalization ability of ELM, the regularization parameter C is introduced, and Equation (13) is rewritten to obtain a new β :

$$\beta = H^T \left(\frac{I}{C} + HH^T \right)^{-1} Y \quad (14)$$

where I represents the identity matrix. The output weight β can be obtained by Equation (13). Input the new sample dataset into the ELM network to obtain:

$$o(g) = h(g)H^T \left(\frac{I}{\theta} + HH^T \right)^{-1} Y \tag{15}$$

where $o(g)$ is the actual output value of the new sample dataset g ; $h(g)$ is the random mapping matrix of the hidden layer.

2.3.2. Autoencoder-Extreme Learning Machine (AE-ELM)

As an unsupervised learning algorithm, the auto-encoder maps the input features to the hidden layer through the encoder, and then uses the decoder to reconstruct the feature vector. In this way, the features of the data can be effectively learned through the encoding and decoding processes. According to Equation (16), the AE-ELM model randomly generates orthogonal input weights and biases:

$$\begin{cases} W^T W = I \\ b^T b = 1 \end{cases} \tag{16}$$

The input sample dataset X is mapped to the hidden layer through encoding, and the output matrix can be calculated by $X = g(WX + b)\beta$:

$$\beta = H^T \left(\frac{I}{C} + HH^T \right)^{-1} X \tag{17}$$

The transposed matrix β^T of β is used as the input weight of the next layer, and $H\beta$ is the input matrix of the next network layer. Figure 1 shows the basic network structure of AE-ELM.

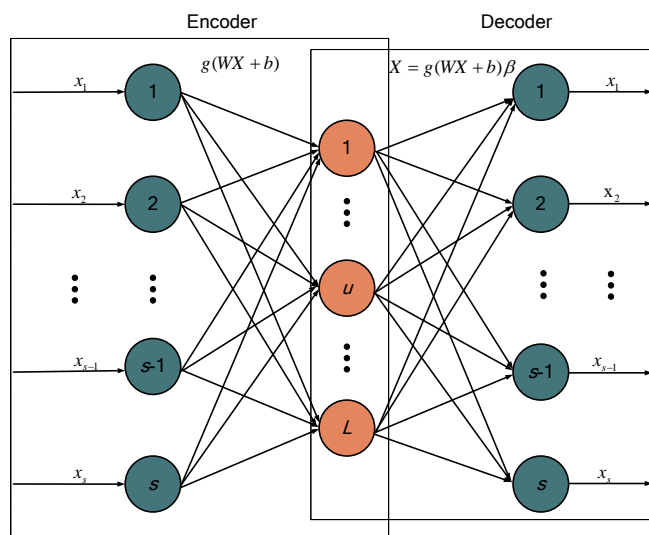


Figure 1. The Structure of AE-ELM.

2.3.3. Deep Extreme Kernel Learning Machine (DKELM)

The DK-ELM model is a deep model based on KELM by combining multiple AE-ELM models, and a neural network composed of multiple hidden layers is constructed by stacking multiple AE-ELM models. The DK-ELM model can efficiently extract the effective features of the samples by mapping the initial features to the new feature space. Moreover, to increase the stability of the samples, a kernel function is introduced to eliminate the shortcomings of random mapping. As shown in Figure 2, the principle of DK-ELM is as follows:

(1) Input the original feature matrix X , and calculate the output matrix $H(1)$ and reconstruction matrix β^1 of the first hidden layer according to Section 3.2.

(2) $H(1)$ is used as the input of the second hidden layer, the transposed matrix $(\beta^1)^T$ of β^1 is used as the output matrix $H(2)$ of the second hidden layer, and calculate β^2 according to Equation (17).

(3) Repeat the operation in Step 2: use h_λ as the input of the $(\lambda + 1)$ -th hidden layer, and calculate the output matrix $H^{\lambda+1}$ and reconstruction matrix $\beta^{\lambda+1}$ of the $(\lambda + 1)$ -th hidden layer.

(4) Map the feature samples processed by the AE-ELM model using the kernel function to obtain the kernel matrix.

$$\Omega_{ELMx_p,x_q} = h(x_p)h(x_q) = K(x_p, x_q) \tag{18}$$

where x_p and x_q are the p -th and q -th input samples; $K(x_p, x_q)$ is the kernel function, and we select radial basis function (RBF), polynomial kernel function (POLY) and wavelet kernel function (WAVE) as kernel function. The expression of RBF is $k(x, y) = \exp(-\gamma\|x - y\|^2)$, and the expression of POLY is $k(x, y) = (ax^T y + c)^d$, the expression of WAVE is $k(x, y) = \prod_{i=1}^d \cos\left(\frac{1.75(x - y)}{\sigma}\right) \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right)$. γ, a, c and σ is parameters of different kernel functions.

(5) Combining the kernel function, calculate the last output weight β and the last output $g(x)$ according to Equation (19):

$$\begin{cases} \beta = \left(\frac{I}{\theta} + \Omega_{ELM}\right)^{-1} Y \\ g(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix} \cdot \beta \end{cases} \tag{19}$$

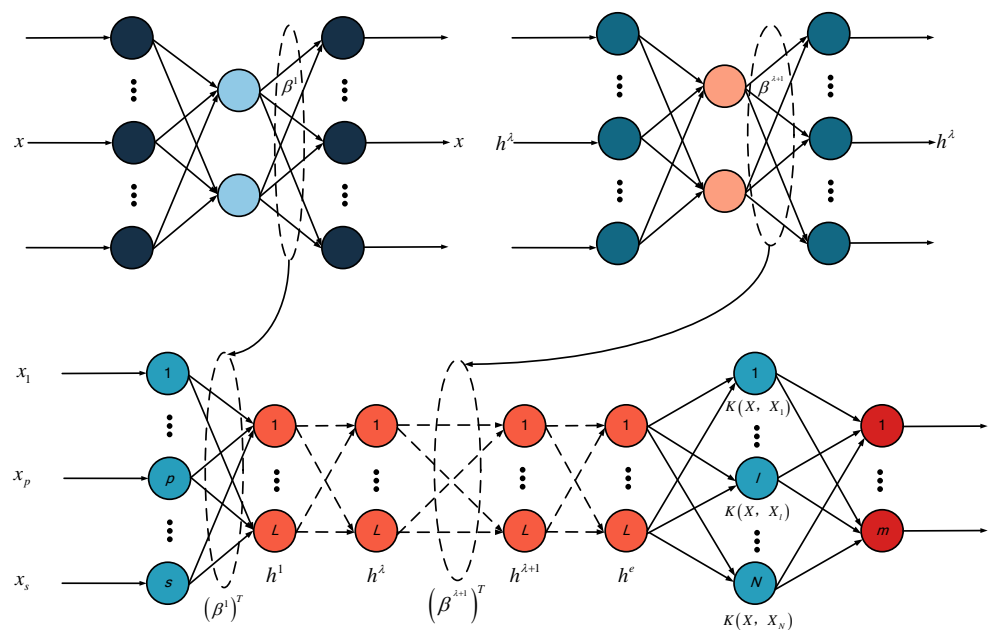


Figure 2. The structure of DK-ELM.

2.4. Marine Predators Algorithm (MPA)

Inspired by the optimal predation strategies (Levy flight and Brownian motion) of marine predators and the speed change strategy after predator and prey encounter, Faramarzi

developed the marine predators' algorithm (MPA) by simulating the predation process of marine predators and considering the influence of eddy currents. The specific process of MPA is as follows:

- (1) Set the population number pop and the maximum iterations M_Iter .
- (2) Initialize the population parameters, calculate the fitness value of the population, and find the individual X_{opt} in the population under the optimal fitness value.
- (3) If the current iteration number t satisfies $t < M_Iter/3$, update the locations of individuals in the population according to Equation (20); otherwise, go to Step 4.

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_B \otimes (\overrightarrow{X}_{opt} - \overrightarrow{R}_B \otimes \overrightarrow{X}_i) \\ \overrightarrow{X}_i = \overrightarrow{X}_i + 0.5R \otimes \overrightarrow{stepsize}_i \end{cases} \quad (20)$$

where $stepsize$ is the step size; R_B is a vector containing random numbers based on normal distribution, which is used to represent the Brownian motion; \otimes is the Kronecker operator; R is a random vector within the range of $(0, 1)$.

- (4) If the current iteration number t satisfies $M_Iter/3 \leq t < 2 \times M_Iter/3$, the update method of population particles divided into two parts. For the first half of the population, the locations of population particles should be updated according to Equation (21); for the remaining population, the locations should be updated according to Equation (22); otherwise, go to Step 5.

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_L \otimes (\overrightarrow{X}_{opt} - \overrightarrow{R}_B \otimes \overrightarrow{X}_i) \\ \overrightarrow{X}_i = \overrightarrow{X}_i + 0.5R \otimes \overrightarrow{stepsize}_i \quad i = 1, \dots, pop/2 \end{cases} \quad (21)$$

where R_L is a random vector representing the Levy flight.

$$\begin{cases} \overrightarrow{stepsize}_p = \overrightarrow{R}_B \otimes (\overrightarrow{R}_B \otimes \overrightarrow{X}_{opt} - \overrightarrow{X}_p) \\ CF = \left(1 - \frac{t}{M_iter}\right)^{2t/M_Iter} \\ \overrightarrow{X}_i = \overrightarrow{X}_{opt} + 0.5 \cdot CF \otimes \overrightarrow{stepsize}_i \quad i = pop/2, \dots, pop \end{cases} \quad (22)$$

where CF is a dynamic parameter that controls the step size.

- (5) If the current iteration number t satisfies $t \geq 2 \times M_Iter/3$, update the locations of the population particles according to Equation (23); otherwise, go to Step 6.

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_L \otimes (\overrightarrow{R}_L \otimes \overrightarrow{X}_{opt} - \overrightarrow{X}_i) \\ \overrightarrow{X}_i = \overrightarrow{X}_{opt} + 0.5 \cdot CF \otimes \overrightarrow{stepsize}_i \end{cases} \quad (23)$$

- (6) Output the best location X_{best} of the population particle.

2.5. The Proposed Method

Combining the auto-encoder and KELM, DK-ELM can obtain more effective feature samples by transforming the sample feature space, thus improving the accuracy of KELM. However, in the DK-ELM model, the number of nodes at each hidden layer, the regularization parameters of the AE-ELM model, the kernel parameters and penalty coefficient of the KELM at the top layer all affect the performances of the DK-ELM model. If the parameters are set based experiences, it will not only take a lot of time, but the best combination of parameters may not even be found. To address this problem, this paper introduces the MPA algorithm to optimize the parameters of the DK-ELM model, which can achieve adaptive selection of the DK-ELM parameters, and effectively improve the performances of the DK-ELM.

As shown in Figure 3, the main steps of the proposed method are as follows:

- (1) The bearing vibration signals are denoised by the wavelet threshold denoising method.
- (2) Utilize EWT to decompose the bearing signal into different empirical modal components, and extract the attention entropy of the components as the feature samples of ADK-ELM.
- (3) Divide the feature samples into the training set, test set and validation set according to the ratio of 3:1:1.
- (4) Set the misjudgment rate of the validation set as the fitness function; select the number of hidden layers and the kernel function of the DK-ELM model; set the upper and lower limits of the parameters, in which process, the node number at the hidden layer should be within the range of (0, 100), the regularization parameters and penalty coefficient of the AE-ELM model should be within the range of (10^{-3} , 10^3), and the kernel parameters should be within the range of (10^{-7} , 10^{-3}).
- (5) In the MPA algorithm, the population number of particle is 20, and the maximum number of iterations M_Iter is 50.
- (6) According to the rules of MPA population optimization described in Section 2.4, find the optimal parameter combination.
- (7) Substitute the optimal parameters into the DK-ELM model to obtain the output of the test set.

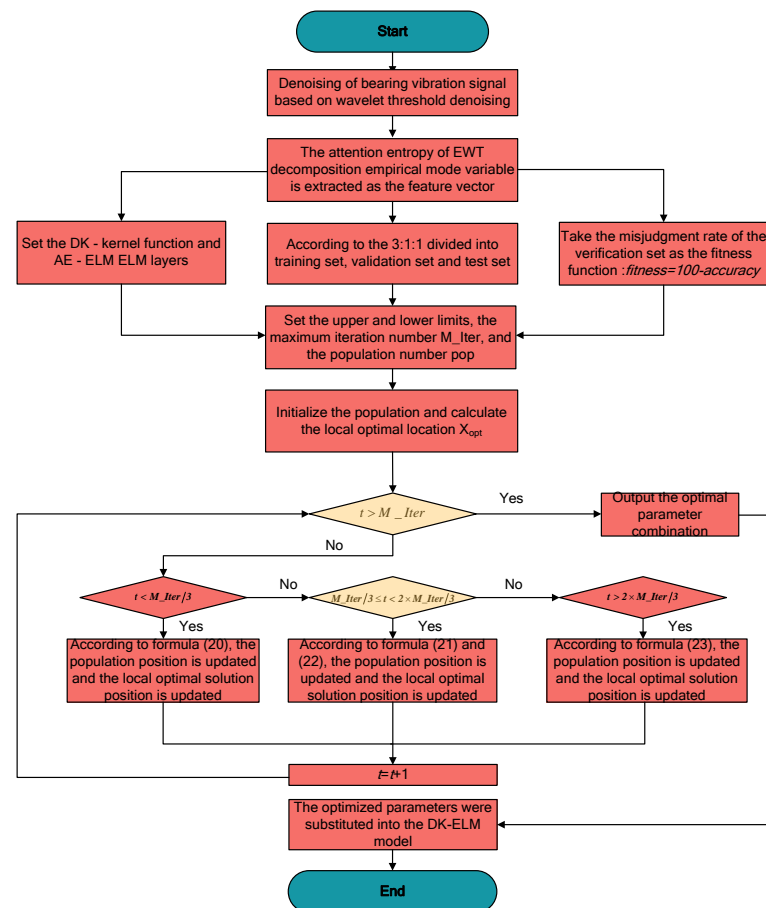


Figure 3. The flow chart of proposed diagnostic method of rolling bearing.

3. Simulation Experiments

To verify the rationality of the method proposed in this paper, and the bearing fault data from Western Reserve University [36] and the bearing fault data of Cincinnati, USA [37] are used as the research objects. The rationality of the proposed method is verified based on recognition of different fault states.

3.1. Bearing Fault Data from Western Reserve University

3.1.1. Fault Data Collection

The fault experiment platform of Western Reserve University is mainly composed of motor, bearing, torque sensor, encoder, power converter and other electronic control equipment. The tested bearing is the drive end bearing, and the sampling frequency of bearing fault signal is 12 kHz, motor speed is 1797 r/min. As shown in Table 1, according to different bearing diameters and different locations of fault and defect, the fault signals are divided into 10 different states. In the experiment, 500 sets of samples under various bearing conditions were collected, and each sample contained 4086 collection points. To facilitate the simulation experiment, the data state labels were coded.

Table 1. Bearing status information (Western Reserve University).

Fault Location	Depth of Bearing Fault Defects	Number of Samples	Fault Type	Label
Outer race	0.1778	500	OR_07	1000000000
Outer race	0.3556	500	OR_14	0100000000
Outer race	0.5334	500	OR_21	0010000000
Inner race	0.1778	500	IR_07	0001000000
Inner race	0.3556	500	IR_14	0000100000
Inner race	0.5334	500	IR_21	0000010000
Ball	0.1778	500	B_07	0000001000
Ball	0.3556	500	B_14	0000000100
Ball	0.5334	500	B_21	0000000010
Normal	/	500	NOR	0000000001

3.1.2. Data Processing

The bearing signal often contains a lot of noise, which affects the extraction of effective fault information. To address this problem, this paper adopts the wavelet threshold denoising method, and sets the wavelet basis function as ‘sym8’. The threshold function is an adaptive threshold function, soft threshold is used, and the number of decomposition layers is as 5. By comparing Figures 4 and 5, it can be seen that the noise part in the vibration signal is effectively eliminated, and the effective information is highlighted by the wavelet threshold.

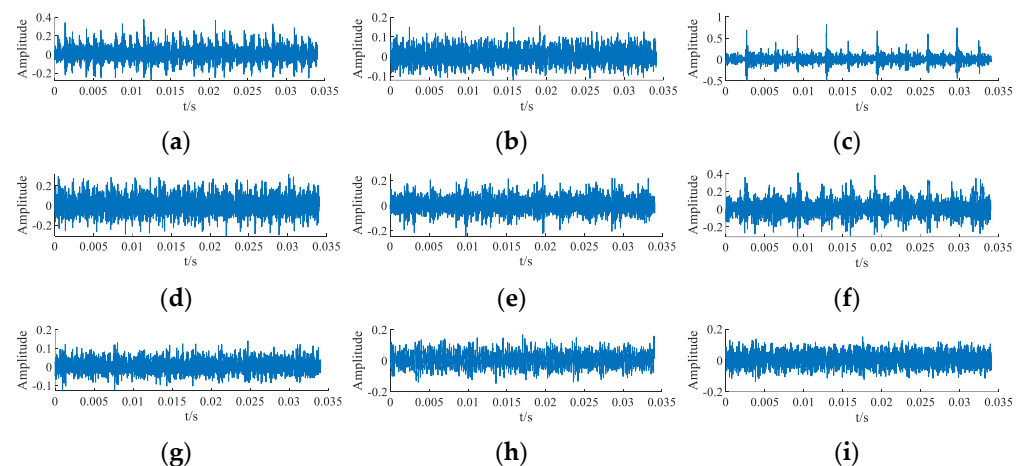


Figure 4. The original vibration signal of rolling bearings: (a) The original vibration signal of OR_07. (b) The original vibration signal of OR_14. (c) The original vibration signal of OR_21. (d) The original vibration signal of IR_07. (e) The original vibration signal of IR_14. (f) The original vibration signal of IR_21. (g) The original vibration signal of B_07. (h) The original vibration signal of B_14. (i) The original vibration signal of B_21.

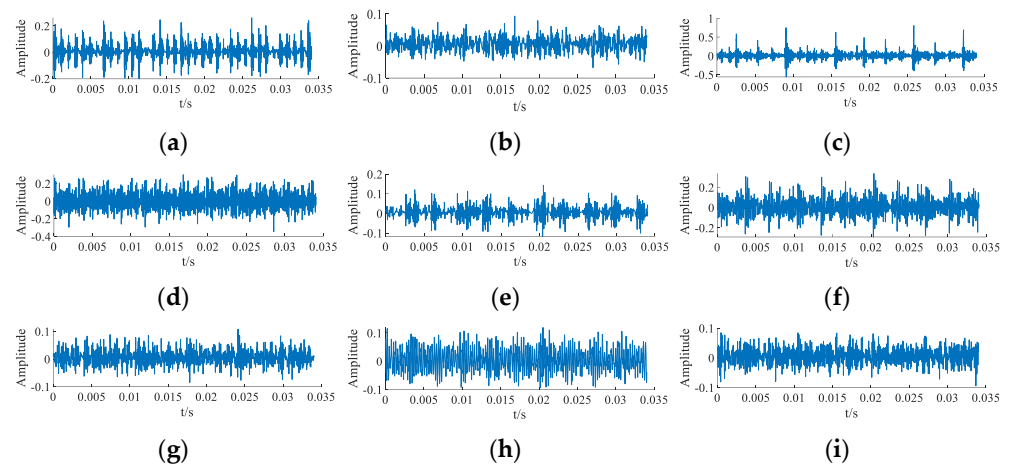


Figure 5. The denoised vibration signal of rolling bearings: (a) The denoised vibration signal of OR_07. (b) The denoised vibration signal of OR_14. (c) The denoised vibration signal of OR_21. (d) The denoised vibration signal of IR_07. (e) The denoised vibration signal of IR_14. (f) The denoised vibration signal of IR_21. (g) The denoised vibration signal of B_07. (h) The denoised vibration signal of B_14. (i) The denoised vibration signal of B_21.

3.1.3. Data Processing

In this paper, EWT is used to decompose the denoised vibration signal into 5 different IMF components, and the attention entropy of IMF is extracted and used as the feature input vector of AK-DELM. As shown in Figure 6, the fault signal is decomposed into the modal components with different amplitudes by EWT, and the local peaks of the signal are obvious, which is convenient for the extraction of the attention entropy of the modal components.

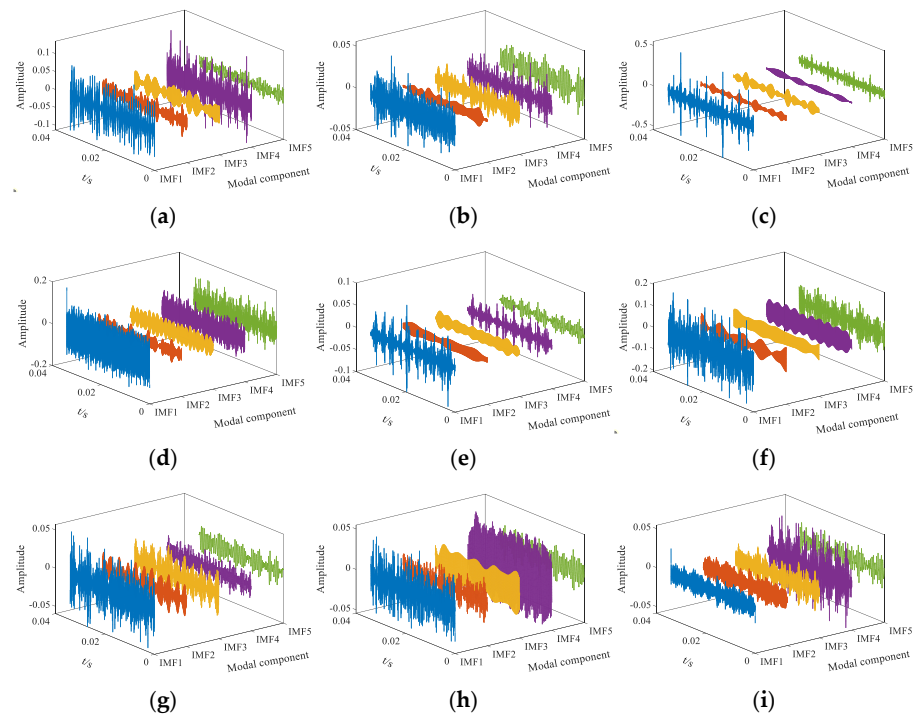


Figure 6. The modal components of different vibration signals by EWT: (a) The modal components of OR_07 by EWT. (b) The modal components of OR_14 by EWT. (c) The modal components of OR_21 by EWT. (d) The modal components of IR_07 by EWT. (e) The modal components of IR_14 by EWT. (f) The modal components of IR_21 by EWT. (g) The modal components of B_07 by EWT. (h) The modal components of B_14 by EWT. (i) The modal components of B_21 by EWT.

As shown in Table 2, the fault signal is decomposed into different IMF components via EWT. By analyzing the attention entropy distribution of the IMF components, we can find the attention entropy distributions of different fault signals present certain differences. The results of simulation experiments show that the fault categories can be effectively distinguished by using the attention entropy.

Table 2. The distribution of attention entropy about different types of rolling bearing signal.

	IMF1	IMF2	IMF3	IMF4	IMF5
OR_07	1.125	0.727	0.886	1.969	2.835
OR_14	2.901	0.867	0.575	2.951	2.018
OR_21	3.395	1.245	1.076	1.613	2.991
IR_07	2.558	0.858	0.987	1.740	3.363
IR_14	3.218	0.924	0.576	2.680	3.525
IR_21	3.428	1.179	0.433	0.994	2.587
B_07	2.261	1.763	1.271	3.474	2.611
B_14	3.260	3.593	0.904	1.270	3.068
B_21	1.274	0.941	1.746	1.071	3.013
NOR	1.200	1.722	0.861	1.414	2.907

3.1.4. DKELM Model Optimized by MPA

The attention entropies of the extracted modal components are used as the feature input vectors, which are divided into the training set, validation set, and test set according to the ratio of 3:1:1. The MPA algorithm is employed to optimize the number of nodes at each AE-ELM layer in DK-ELM, the regularization parameters of the AE-ELM model, as well as the kernel parameters and penalty coefficient of the top-layer KELM, which has achieved the adaptive setting of DK-ELM model parameters. To make it easier to remember, the DK-ELM model improved with different types of kernel functions is named AD-Kernel-ELM. In order to investigate the performances of ADK-ELM with different kernel functions, the three kernel functions of RBF, WAVE, and POLY are selected in this paper.

The parameters of the DK-ELM model with different kernel functions are optimized by MPA. By comparing the optimization processes of the three models of D-WAVE-ELM, D-PLOY-ELM, and D-RBF-ELM using MPA, it is found that MPA has the best optimization effects on the D-RBF-ELM model. As shown in Figure 7, compared with the other two models, D-RBF-ELM can find the optimal solutions at different AE-ELM layers. Therefore, the model with better performances can be trained by optimizing D-RBF-ELM.

3.1.5. ADKELM Model Diagnosis

To analyze the performances of ADK-ELM models with different kernel functions and different AE-ELM layers, the three kernel functions of RBF, WAVE, and POLY are used in this paper. For the ADK-ELM model in each case, the experiment was repeated 10 times, so as to analyze the fault diagnosis performances of the ADK-ELM model with different kernel functions. According to Table 3 and Figure 8, the ADK-ELM models with three different kernel functions achieved the best performances with the following numbers of AE-ELM layers: the AD-POLY-ELM model achieved an accuracy of 97.09% with two AE-ELM layers; the AD-WAVE-ELM model achieved an accuracy of 98.74% with two AE-ELM layers; the AD-RBF-ELM model achieved an accuracy of 99.1% with four AE-ELM layers. Therefore, the AD-RBF-ELM model had the best performance with four AE-ELM layers.

According to Section 3.1.4, by optimizing the D-RBF-ELM model using MPA, in the finally obtained AD-RBF-ELM model, the node numbers of various AE-ELM layers are 36, 30, 100, and 79, respectively; the regularization parameter of the AE-ELM model is 497.27; the penalty coefficient of the radial basis kernel function is 784.55; the parameter of the kernel function is 1.031×10^{-7} . As shown in Figure 9 AD-RBF-ELM achieved 100% accuracy on the training set; after inputting the samples from the test set into trained AD-RBF-ELM model, an accuracy of 99.1% was achieved on the test set. Among these samples,

seven IR_14 samples were misdiagnosed as B_07, one IR_14 sample was misdiagnosed as OR_14, and one B_21 sample was misdiagnosed as B_07.

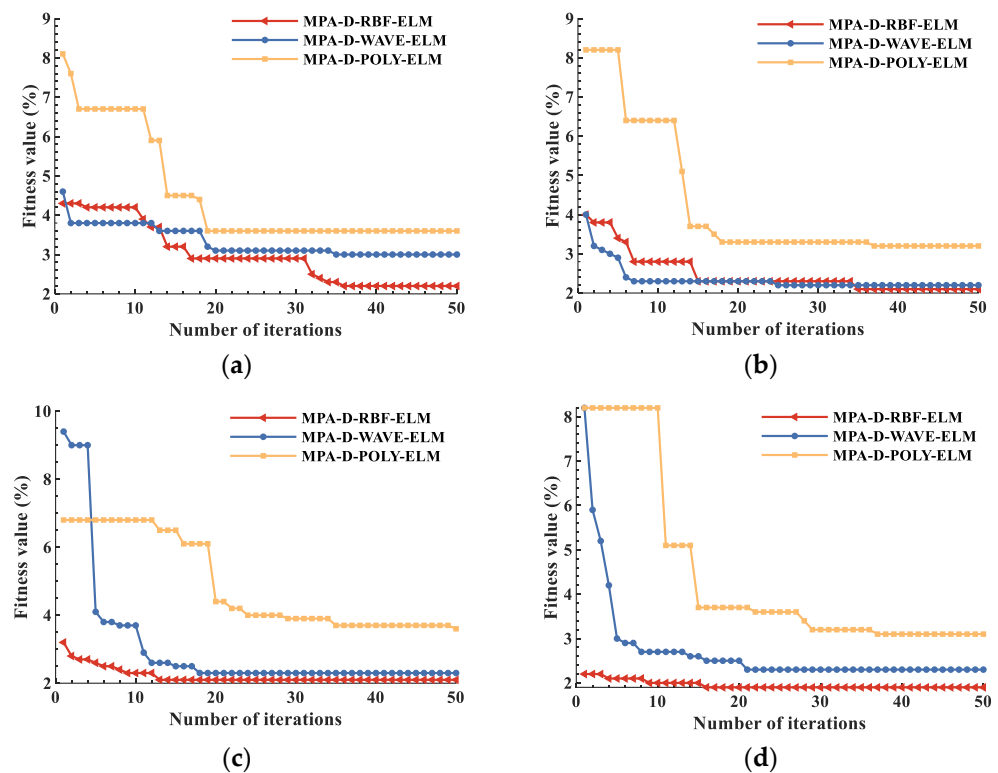


Figure 7. Iteration curves of different DK-ELM optimized by MPA (Western Reserve University): (a) Iteration curves of different DK-ELM optimized by MPA (The model of one layer AE-ELM). (b) Iteration curves of different DK-ELM optimized by MPA (The model of two layers AE-ELM). (c) Iteration curves of different DK-ELM optimized by MPA (The model of three layers AE-ELM). (d) Iteration curves of different DK-ELM optimized by MPA (The model of four layers AE-ELM).

Table 3. Diagnosis performances of ADK-ELM model under different kernel functions (Western Reserve University).

Kernel Function	Number of Hidden Layers	Average Accuracy on the Training Set	Average Accuracy on the Test Set	Maximum Accuracy on the Test Set	Minimum Accuracy on the Test Set	Standard Deviation
RBF_Kernel	1	100.00	97.41	99.00	95.90	1.43
	2	99.70	98.64	99.20	98.30	0.29
	3	99.41	98.74	99.10	98.50	0.22
	4	100.00	99.10	99.30	99.00	0.10
POLY_Kernel	1	96.89	95.90	96.30	95.60	0.22
	2	97.41	97.09	97.70	96.30	0.45
	3	97.15	96.07	97.30	94.40	0.83
	4	97.50	96.85	97.30	96.10	0.39
WAVE_Kernel	1	98.22	97.75	98.80	96.10	0.70
	2	99.56	98.74	99.10	98.20	0.31
	3	100.00	98.53	99.00	98.20	0.25
	4	100.00	98.64	99.00	97.6	0.42

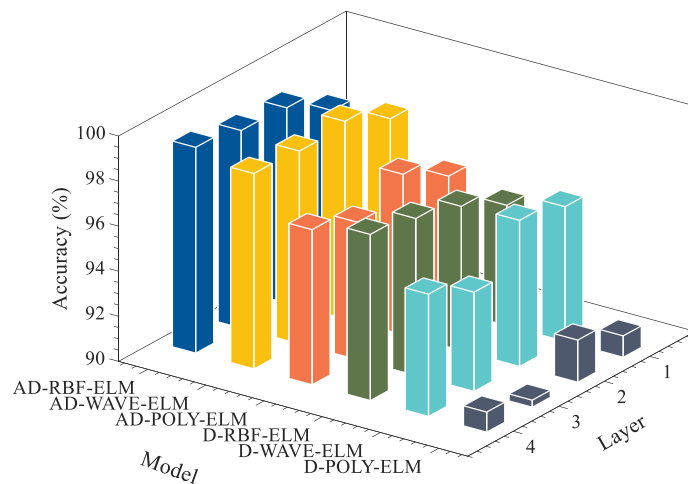


Figure 8. The accuracy rates of different DK-ELM models (Western Reserve University).

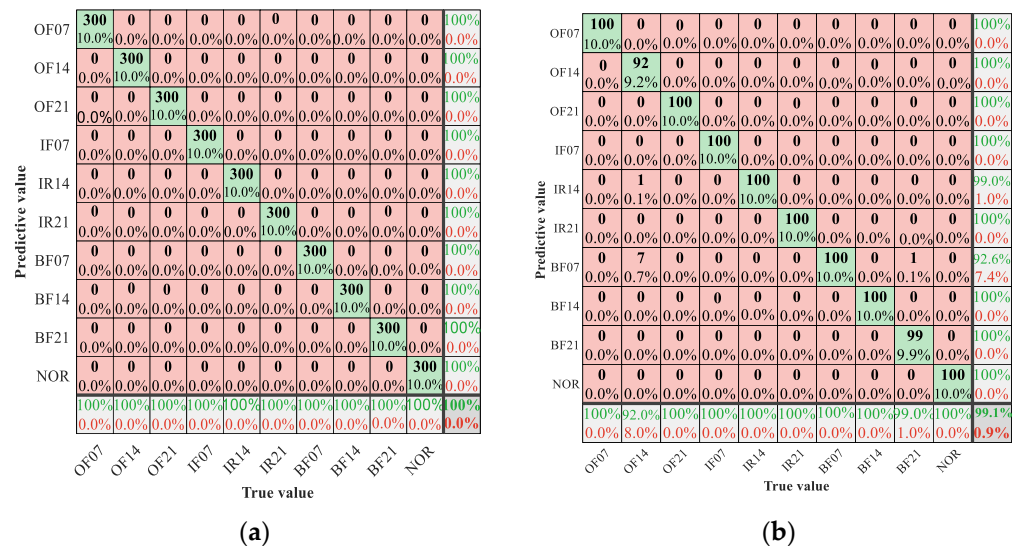


Figure 9. The diagnostic confusion matrix of AD-RBF-ELM model (Western Reserve University): (a) The confusion matrix of training set. (b) The confusion matrix of testing set.

Meanwhile, to verify the superiority of the method proposed in this paper, our method is compared with the DK-ELM, ELM, BPNN and PNN algorithms. As can be seen from Table 4, the proposed method achieved a diagnosis rate of 99.1%; the not optimized DK-ELM model achieved a diagnosis rate of 97.25%; the ELM and PNN models obtained the diagnosis rates of 94.76% and 81.39%, respectively; the BPNN model only achieved an accuracy of 37.53%. Compared with other traditional classifier algorithms (AD-WAVE-ELM, AD-POLY-ELM, D-RBF-ELM, D-WAVE-ELM, D-POLY-ELM, ELM, PNN, and BPNN), the proposed method improves the diagnostic rate by 0.36%, 2.01%, 1.76%, 2.65%, 7.26%, 15.5%, 17.74%, and 61.57%. Based on the comparison of different models, the rationality and superiority of the method proposed in this paper are verified.

Table 4. Diagnostic rate of different models (Western Reserve University).

Model	AD-RBF-ELM	AD-WAVE-ELM	AD-POLY-ELM	D-RBF-ELM	D-WAVE-ELM	D-POLY-ELM	ELM	PNN	BPNN
Accuracy	99.1	98.74	97.09	97.34	96.45	91.84	83.6	81.36	37.53

3.2. Bearing Fault Data of Cincinnati Fault Data

In this paper, the bearing dataset provided by the Intelligent Maintenance System Center (IMS) of the University of Cincinnati is used as the research object, and the rationality of the proposed method is verified by analyzing the recognition performances of four different states of bearing inner ring fault, outer ring fault, roller fault and normal state using our method.

The test device includes drive motor, load device, acceleration sensor, and four experimental bearings. In this paper, the data of No. 1 bearing from 0:4 on 23 October to 8:14 on 23 October are used as the normal vibration signal of the bearing; the data of No. 3 and No. 4 bearings from 0:6 on 22 November to 8:16 on 22 November are used as the vibration data of inner ring fault and roller fault, respectively; the data of No. 1 bearing from 21:42 on 15 February to 5:52 on 16 February are used as the vibration data of outer ring fault. As shown in Table 5, a total of 2000 bearing samples were collected in this experiment, and each sample contained 2048 collection points. Similar to the steps in Section 3.1, the samples were denoised, and the attention entropies were extracted by EWT decomposition and used as the feature vectors.

Table 5. Bearing status information (University of Cincinnati).

Fault	Sample Points	Number of Samples	Fault Type	Label
Inner race	2048	500	OR	1000
Outer race	2048	500	IR	0100
Ball	2048	500	Ball	0010
Normal	2048	500	NOR	0001

The feature vectors are divided into the training set, validation set, and test set according to the ratio of 3:1:1. The training set and validation set were input into the MPA-DK-ELM model to optimize the parameters of the DK-ELM model. As shown in Figure 10, compared with other models, MPA can provide the best optimization effects for D-RBF-ELM. With different AE-ELM layers, MPA-D-RBF-ELM can always find the optimal fitness value, verifying that the MPA-optimized D-RBF-ELM can train the model with better performances.

To compare the performances of different DK-ELM models, the experiment on each model was repeated 10 times, so as to analyze the diagnostic performances of different DK-ELM models. As shown in Figure 11, AD-RBF-ELM achieved an average accuracy of 98.3% when the number of AE-ELM layers was 4; AD-WAVE-ELM achieved an average accuracy of 97.7% with one AE-ELM layer; AD-POLY-ELM achieved an average accuracy of 95.725% with one AE-ELM layer. Therefore, by comparing the performances of different DK-ELM models, it is verified that AD-RBF-ELM has the best performance when the number of AE-ELM layers is 4.

As shown in Figure 12, AD-RBF-ELM achieved 100% accuracy on the training set, and by inputting the test set samples into the trained AD-RBF-ELM model, an accuracy of 99% was obtained on the test set. Among these samples, one normal sample was misdiagnosed as IR, and three IRs were misdiagnosed as Nor. The proposed method can completely identify the roller and inner ring fault signals. According to Table 6, the proposed method can provide better diagnostic performances than other traditional classifiers. Compared with other traditional classifier algorithms (AD-WAVE-ELM, AD-POLY-ELM, D-RBF-ELM, D-WAVE-ELM, D-POLY-ELM, ELM, PNN, and BPNN), the proposed method improves the diagnostic rate by 0.6%, 2.575%, 2.1%, 1.8%, 6%, 12.31%, 8.045%, and 12.75%.

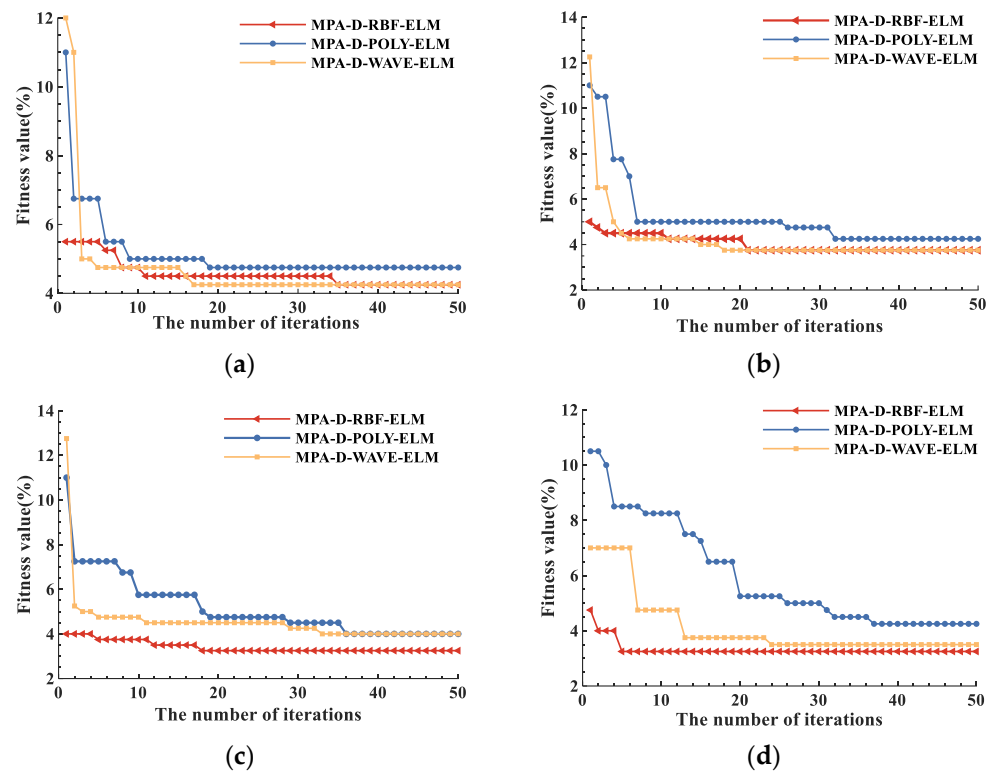


Figure 10. Iteration curves of different DK-ELM optimized by MPA (University of Cincinnati):
 (a) Iteration curves of different DK-ELM optimized by MPA (The model of one layer AE-ELM).
 (b) Iteration curves of different DK-ELM optimized by MPA (The model of two layers AE-ELM).
 (c) Iteration curves of different DK-ELM optimized by MPA (The model of three layers AE-ELM).
 (d) Iteration curves of different DK-ELM optimized by MPA (The model of four layers AE-ELM).

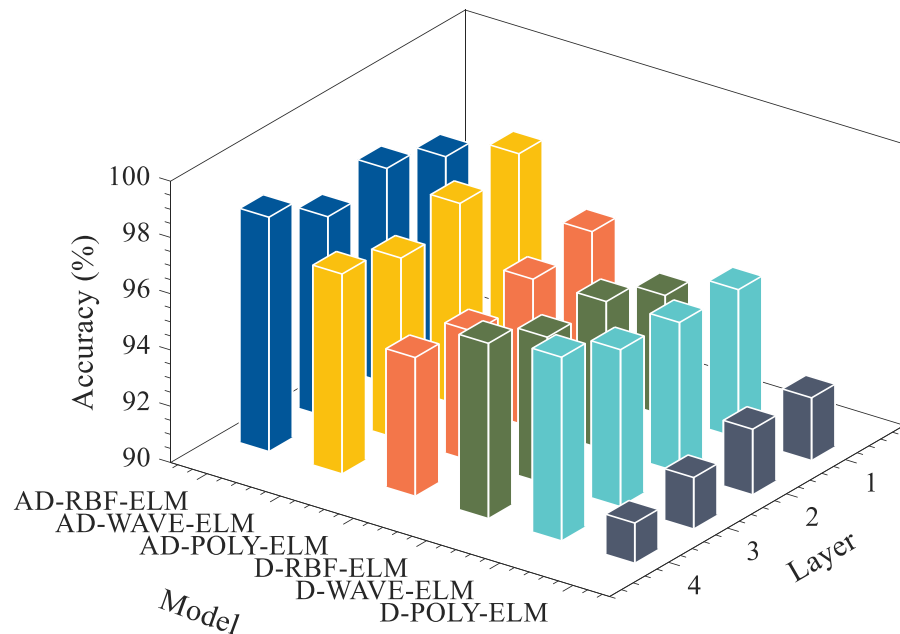


Figure 11. The accuracy rates of different DK-ELM models (University of Cincinnati).

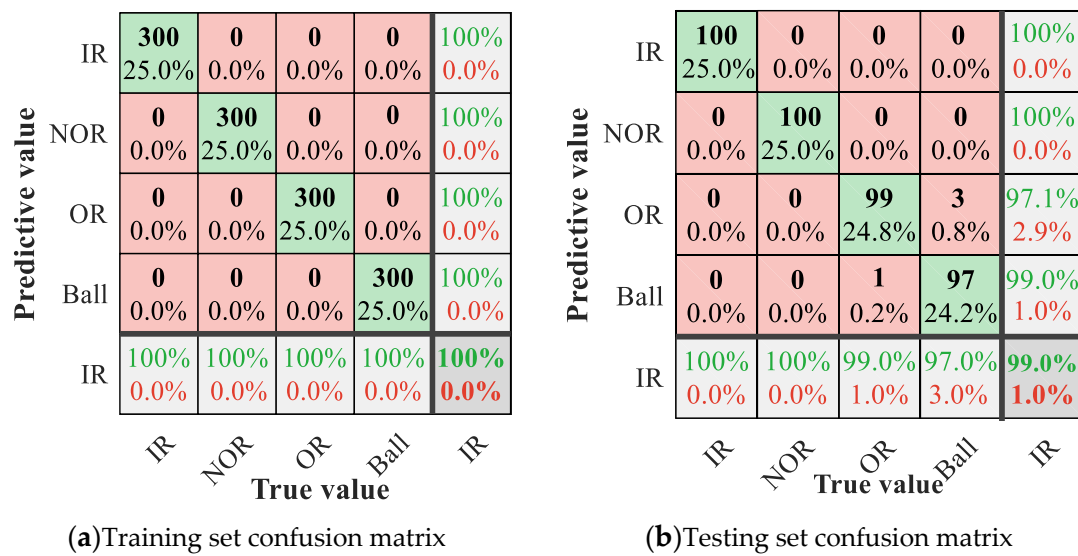


Figure 12. The diagnostic confusion matrix of AD-RBF-ELM model (University of Cincinnati): (a) The confusion matrix of training set. (b) The confusion matrix of testing set.

Table 6. Diagnostic rate of different models (University of Cincinnati).

Model	AD-RBF-ELM	AD-WAVE-ELM	AD-POLY-ELM	D-RBF-ELM	D-WAVE-ELM	D-POLY-ELM	ELM	PNN	BPNN
Accuracy	98.3	97.7	95.725	96.2	96.5	92.3	85.99	90.255	85.55

4. Conclusions

To address the difficulty of early fault diagnosis of rolling bearings, this paper proposes a bearing fault diagnosis model combining the IMF attention entropy and the adaptive deep kernel extreme learning machine. First, the wavelet threshold denoising method is adopted to effectively eliminate the noise in the vibration signal. Second, the denoised signal is decomposed by EWT, and the attention entropies of the IMF components are extracted and used as the feature vectors. Then, to address the difficulty to determine the parameters of DK-ELM, the MPA algorithm is employed for adaptive setting of the parameters of the DK-ELM model. Finally, the ADK-ELM model is used to achieve effective recognition of the bearing faults. The main conclusions of this paper are as follows:

(1) As the traditional entropy is extremely sensitive to the parameters, this paper introduces the attention entropy for feature extraction of the IMF components. The simulation results show that the attention entropy can effectively distinguish various fault signals.

(2) The MPA optimization algorithm is used to optimize the node number at the hidden layers of the DK-ELM model, the regularization parameters of the AE-ELM, the kernel parameters of the kernel function and the penalty coefficient, which can achieve adaptive adjustment of the parameters of the DK-ELM model.

(3) The diagnostic performances of the ADK-ELM model with different kernel functions and different hidden layers are investigated. The analysis of simulation results shows that AD-RBF-ELM model achieved the best diagnostic performance with four hidden layers.

Author Contributions: Conceptualization, B.W.; methodology, W.W., X.Z. and F.C.; software, L.L. and P.Z.; validation, F.M., D.C., F.W. and B.W.; funding acquisition, B.W. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the scientific research foundation of the Young Scholar Project of Cyrus Tang Foundation, the Shaanxi Province Key Research and Development Plan (Number 2021NY-181), the coordinates scientific research projects of State Power Investment Corporation Limited (Number TC2020SD01), and the National Natural Science Foundation of China (Numbers 51909222 and 51509210).

Data Availability Statement: Not applicable.

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

References

1. Hu, Z.; Wang, Y.; Ge, M.; Liu, J. Data-driven fault diagnosis method based on compressed sensing and improved multiscale network. *IEEE Trans. Ind. Electron.* **2019**, *67*, 3216–3225. [[CrossRef](#)]
2. Liang, M.; Bozchalooi, I.S. An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection. *Mech. Syst. Signal Process.* **2010**, *24*, 1473–1494. [[CrossRef](#)]
3. Song, L.; Wang, H.; Chen, P. Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1887–1899. [[CrossRef](#)]
4. Pan, H.; Zheng, J.; Liu, Q. A novel roller bearing condition monitoring method based on RHLCD and FVPMCD. *IEEE Access* **2019**, *7*, 96753–96763. [[CrossRef](#)]
5. Wang, Z.; Han, Z.; Liu, Q.; Ning, S. Weak fault diagnosis for rolling element bearing based on MED-EEMD. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 70–78.
6. Zhang, C.; Zhao, R.; Deng, L. Fault diagnosis method of rolling bearing based on the singular value entropy of EEMD. *J. Vib. Meas. Diagn.* **2019**, *39*, 353–358+446–447. (In Chinese)
7. Wang, J.; Lin, J. Structural health monitoring of periodic infrastructure: A review and discussion. *Data Min. Struct. Dyn. Anal.* **2019**, 25–40. [[CrossRef](#)]
8. Song, Y.; Liu, Z.; Ronnquist, A.; Navik, P.; Liu, Z. Contact wire irregularity stochastics and effect on high-speed railway pantograph–catenary interactions. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8196–8206. [[CrossRef](#)]
9. Gilles, J. Empirical wavelet transform. *IEEE Trans. Signal Process.* **2013**, *61*, 3999–4010. [[CrossRef](#)]
10. Li, J.; Wang, H.; Wang, X.; Zhang, Y. Rolling bearing fault diagnosis based on improved adaptive parameterless empirical wavelet transform and sparse denoising. *Measurement* **2020**, *152*, 107392. [[CrossRef](#)]
11. Sahani, M.; Dash, P.K. Fault location estimation for series-compensated double-circuit transmission line using EWT and weighted RVFLN. *Eng. Appl. Artif. Intell.* **2020**, *88*, 103336. [[CrossRef](#)]
12. Li, L.; Guo, A.; Chen, H. Feature extraction based on EWT with scale space threshold and improved MCKD for fault diagnosis. *IEEE Access* **2021**, *9*, 45407–45417. [[CrossRef](#)]
13. Ding, Y.; Ma, L.; Wang, C.; Tao, L. An EWT-PCA and extreme learning machine based diagnosis approach for hydraulic pump. *IFAC-PapersOnline* **2020**, *53*, 43–47. [[CrossRef](#)]
14. Wang, X.; Tang, G.; Wang, T.; Zhang, X.; Peng, B.; Dou, L.; He, Y. Lkurtogram guided adaptive empirical wavelet transform and purified instantaneous energy operation for fault diagnosis of wind turbine bearing. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 3506619. [[CrossRef](#)]
15. Fu, W.; Tan, J.; Li, C.; Zou, Z.; Li, Q.; Chen, T. A hybrid fault diagnosis approach for rotating machinery with the fusion of entropy-based feature extraction and SVM optimized by a chaos quantum sine cosine algorithm. *Entropy* **2018**, *20*, 626. [[CrossRef](#)]
16. Liu, X.; Zhang, X.; Luan, Z.; Xu, X. Rolling bearing fault diagnosis based on EEMD sample entropy and PNN. *J. Eng.* **2019**, *2019*, 8696–8700. [[CrossRef](#)]
17. Zair, M.; Rahmoune, C.; Benazzouz, D. Multi-fault diagnosis of rolling bearing using fuzzy entropy of empirical mode decomposition, principal component analysis, and SOM neural network. *J. Mech. Eng. Sci.* **2019**, *233*, 3317–3328. [[CrossRef](#)]
18. 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](#)]
19. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)]
20. Jiang, L.; Tan, H.; Li, X.; Lei, J. Fault recognition of spiral bevel gears based on CEEMDAN permutation entropy and SVM. *J. Vib. Meas. Diagn.* **2021**, *41*, 33–40+198–199. (In Chinese)
21. Wang, Z.; Yao, L.; Chen, G.; Ding, J. Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals. *ISA Trans.* **2021**, *114*, 470–484. [[CrossRef](#)] [[PubMed](#)]
22. Yang, J.; Choudhary, G.L.; Rahardja, S.; Franti, P. Classification of interbeat interval time-series using attention entropy. In *IEEE Transactions on Affective Computing*; IEEE: Piscataway, NJ, USA, 2020. [[CrossRef](#)]
23. Jiang, Q.; Jia, M.; Hu, J.; Xu, F. Machinery fault diagnosis using supervised manifold learning. *Mach. Syst. Signal Process.* **2009**, *23*, 2301–2311. [[CrossRef](#)]
24. Zhang, X.; Li, C.; Wang, X.; Wu, H. A novel fault diagnosis procedure based on improved symplectic geometry mode decomposition and optimized SVM. *Measurement* **2021**, *173*, 108644. [[CrossRef](#)]
25. Yao, L.; Fang, Z.; Xiao, Y.; Hou, J.; Fu, Z. An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy* **2021**, *214*, 118866. [[CrossRef](#)]

26. Yang, Z.; Kong, C.; Wang, Y.; Rong, X.; Wei, L. Fault diagnosis of mine asynchronous motor based on MEEMD energy entropy and ANN. *Comput. Electr. Eng.* **2021**, *92*, 107070. [[CrossRef](#)]
27. Shifat, T.A.; Hur, J.W. ANN assisted multi sensor information fusion for BLDC motor fault diagnosis. *IEEE Access* **2021**, *9*, 9429–9441. [[CrossRef](#)]
28. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Cybern.* **2012**, *42*, 513–529. [[CrossRef](#)]
29. Kong, X.; Mao, G.; Wang, Q.; Ma, H.; Yang, W. A multi-ensemble method based on deep auto-encoders for fault diagnosis of rolling bearings. *Measurement* **2020**, *151*, 107132. [[CrossRef](#)]
30. Ding, S.; Zhang, N.; Xu, X.; Guo, L.; Zhang, J. Deep extreme learning machine and its application in EEG classification. *Math. Probl. Eng.* **2015**, *2015*, 129021. [[CrossRef](#)]
31. Mohanty, D.K.; Parida, A.K.; Khuntia, S.S. Financial market prediction under deep learning framework using auto encoder and kernel extreme learning machine. *Appl. Soft Comput.* **2021**, *99*, 106898. [[CrossRef](#)]
32. Li, J.; Xi, B.; Du, Q.; Song, R.; Li, Y.; Ren, G. Deep kernel extreme-learning machine for the spectral-spatial classification of hyperspectral imagery. *Remote Sens.* **2018**, *10*, 2036. [[CrossRef](#)]
33. Cheng, Y.; Zhao, D.; Wang, Y.; Pei, G. Multi-label learning with kernel extreme learning machine autoencoder. *Knowl. -Based Syst.* **2019**, *178*, 1–10. [[CrossRef](#)]
34. Liu, H.; Zhang, Y.; Zhang, H. Prediction of effluent quality in papermaking wastewater treatment processes using dynamic kernel-based extreme learning machine. *Process Biochem.* **2020**, *97*, 72–79. [[CrossRef](#)]
35. Faramarzi, A.; Heidarinejad, M.; Mirjalili, S.; Gandomi, A.H. Marine predators algorithm: A nature-inspired metaheuristic. *Expert Syst. Appl.* **2020**, *152*, 113377. [[CrossRef](#)]
36. Case Western Reserve University Bearing Data Center Website. 2004. Available online: <http://csegroups.case.edu/bearingdatacenter/home> (accessed on 1 May 2022).
37. IMS Bearings Data Set. 2013. Available online: <http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/> (accessed on 11 May 2022).