

## Research Article

# Research on Novel Bearing Fault Diagnosis Method Based on Improved Krill Herd Algorithm and Kernel Extreme Learning Machine

Zhijian Wang <sup>1,2</sup>, Likang Zheng <sup>3</sup>, Junyuan Wang <sup>1</sup>, and Wenhua Du <sup>1</sup>

<sup>1</sup>School of Mechanical Engineering, North University of China, Taiyuan, Shanxi 030051, China

<sup>2</sup>School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shanxi 030619, China

<sup>3</sup>School of Energy and Power Engineering, North University of China, Taiyuan, Shanxi 030051, China

Correspondence should be addressed to Junyuan Wang; wangyi01161013@163.com

Received 10 June 2019; Revised 21 August 2019; Accepted 6 September 2019; Published 28 November 2019

Academic Editor: Danilo Comminiello

Copyright © 2019 Zhijian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, a novel bearing intelligent fault diagnosis method based on a novel krill herd algorithm (NKH) and kernel extreme learning machine (KELM) is proposed. Firstly, multiscale dispersion entropy (MDE) is used to extract fault features of bearings to obtain a set of fault feature vectors composed of dispersion entropy. Then, it is imported into the kernel extreme learning machine for fault diagnosis. But considering the kernel function parameters  $\sigma$  and the error penalty factor  $C$  will affect the classification accuracy of the kernel extreme learning machine, this paper uses the novel krill herd algorithm (NKH) for their optimization. The opposite populations are added to the NKH in the initialization of population to improve its speed and prevent local optimum, and during the period of looking for the optimal solution, the impulse operator is introduced to ensure it has enough impulse to rush out of the local optimal once into the local optimum. Finally, in order to verify the effectiveness of the proposed method, it was applied to the bearing fault experiment of Case Western Reserve University and XJTU-SY bearing data set. The results show that the proposed method not only has good fault diagnosis performance and generalization but also has fast convergence speed and does not easily fall into the local optimum. Therefore, this paper provides a method for fault diagnosis under different loads. Meanwhile, the new method (NKH-KELM) is compared and analyzed with other mainstream intelligent bearing fault diagnosis methods to verify the effectiveness and accuracy of the proposed method.

## 1. Introduction

The rolling bearing is an indispensable part of the most rotating mechanical equipment. The loss of life and property and damage to the environment caused by the failure of the rolling bearing are very serious. In order to avoid bearing failure as much as possible, we can improve the bearing design, study and apply better materials and new technologies in the design and manufacturing stage, strengthen the quality control measures in the production process, and improve the assembly level of the bearing. In addition, another way to prevent catastrophic events of the rotating machinery equipment is to monitor and diagnose the working state of the equipment in real time so as to achieve effective control of the equipment. Therefore, it is of great

significance to monitor the working state of the bearing in real time.

In recent years, various novel methods have been widely used to solve practical engineering problems [1–5]. And learning these novel methods [6–10] provides many ideas for fault diagnosis. In the field of bearing fault diagnosis, novel intelligent fault diagnosis methods emerge one after another in recent years, namely, the method based on statistics having Pearson's correlation coefficient (PCC) [11], the method based on signal processing having modified variable modal decomposition (MVMD) [12], improved ensemble local mean decomposition (IELMD) [13], maximum kurtosis spectral entropy deconvolution (MKSED) [14], regression residual signal based on improved intrinsic time-scale decomposition [15], enhanced singular spectrum

decomposition (ESSD) [16], weighted cyclic harmonic-to-noise ratio [17], time-frequency analysis [18], multipoint optimal minimum entropy deconvolution adjusted (MOMEDA) [19], and so on [20]. In recent years, with the development of big data, machine learning methods and deep learning methods have been widely used to solve practical engineering problems [21–26]. Machine learning methods or deep learning methods were applied in the field of bearing fault diagnosis, including the support vector machine (SVM) [27, 28], BP neural network (BP) [29], deep convolutional transfer learning network (CNN) [30], and kernel extreme learning machine (ELM) [31, 32]. These methods are good for fault diagnosis in most cases, but some are more subjective in choosing parameters. For example, the method used in reference [31] is highly subjective in selecting input weights and biases. Therefore, it is necessary to introduce the parameter optimization algorithm in fault diagnosis.

In the era of big data, many engineering problems need feature extraction. For example, Yin et al. used multiobjective feature extraction optimization to detect M/OD HVI damages [33]. And, in the bearing fault diagnosis, the first step is how to collect useful information from the fault bearing. At present, there are many practical engineering problems that take vibration information as useful information, and there are endless methods for collecting vibration information [34–38]. And most scholars also collect bearing vibration information in the bearing fault diagnosis [39]. Then, how to extract the characteristic information from the vibration signal is the primary problem of intelligent fault diagnosis based on machine learning. As a tool to describe the uncertainty of signals, entropy has been widely used in recent years to extract fault features of rolling bearings [40, 41], such as sample entropy (SampEn) [42] and permutation entropy (PerEn) [43]. However, SampEn and PerEn need to calculate the distance between the two embedded dimensions and the embedded vector of each sample separately, which greatly increases the time for extracting features. In addition, most useful information cannot be extracted from a single time scale when analyzing the time series. In engineering practice, the optimal time scale of the original signal is often unknown. In order to solve this problem, the multiscale operation is introduced into feature extraction. Multiscale allows entropy to be extended to multiple time scales to provide an additional perspective when the time scales are uncertain. Like other entropy measurement methods, multiscale entropy aims to evaluate the complexity of the time series. One of the main reasons for using multiscale entropy is that the relevant time scales in the time series are not known. Therefore, analyzing the problem over multiple time scales will obtain more information. Moreover, bearing fault diagnosis based on multiscale entropy has been widely used in the field of intelligent fault, such as multiscale fuzzy entropy (MFE) [44] and multiscale permutation entropy (MPE) [45]. Rostaghi and Azami have proposed dispersion entropy (DE) [46] and multiscale dispersion entropy (MDE) [47]. MDE does not need to sort the amplitude of each embedded vector nor does it need to calculate the distance between any two compound delay vectors with embedded dimensions  $m$  and  $m + 1$ , which makes DE and MDE faster than PerEn, SampEn, and MperEn,

significantly. Moreover, MDE has obvious advantages in distinguishing different types of dynamic signals, so MDE is more suitable than other methods for the extraction of the bearing vibration signal.

The result of feature extraction of each sample forms a feature vector. How to classify the feature vector accurately is the most critical problem of intelligent fault diagnosis. Three common classification methods based on machine learning are the BP neural network (BP), support vector machine (SVM), and extreme learning machine (ELM). However, in the traditional BP neural network, a large part of the network training parameters need to be set artificially, which is very random. In addition, BP is adjusted by the gradient descent method, which not only is easy to get into the local optimal but also suffers slow convergence [48]. More importantly, this operation is likely to lead to training failure. On the contrary, because the network structure model is not single, the parameters of the neural network need to be constantly adjusted in each iteration, so the diagnosis efficiency is not high, and it will easily fall into the phenomenon of underlearning or overfitting [49]. The SVM has great advantages in solving binary classification problems with small sample size [50], but it is difficult to implement large-scale sample training and solve multiple classifications [51]. However, the fault diagnosis of the rolling bearing is often a multiclassification problem, which means it is difficult for an SVM to implement the fault diagnosis of the rolling bearing and the high-level accuracy rate is difficult to reach. The ELM is a novel algorithm based on the single-hidden-layer feedforward neural network, which has the advantages of the simple mathematical model, global optimal solution, fast learning speed, less parameter selection, and high generalization, which has been applied in many fields [52]. However, the ELM randomly generates input weights and hidden-layer thresholds, resulting in instability of the algorithm. In order to solve this problem, the kernel function is introduced into the basis of the ELM to obtain the kernel extreme learning machine (KELM). The KELM solves the problem of random initialization in the ELM algorithm, and the number of nodes in the hidden layer can be determined accordingly, instead of being specified manually. In addition, in solving some practical engineering problems, the KELM shows high classification accuracy, good promotion ability, and high robustness [53].

In the intelligent fault diagnosis, the parameter optimization of the diagnosis method is also a key point. For example, particle swarm optimization (PSO) was used to optimize the filtering coefficient of deconvolution [54] and ant colony algorithm [55]. Gandomi proposed a krill herd optimization algorithm (KH) to solve the optimization problem in 2012 [56], which is based on the behavior of constantly updating the position of krills because of the predation of krills in nature. This algorithm can adjust the participation amount of exploration and development by observing the progress of solving the problem through various steps dynamically. In addition, Gandomi also compared the KH with eight common optimization methods [56], and the final result showed the KH converges faster than other optimization algorithms and does not easily fall into the local optimal. So, the KH can effectively solve various optimization problems and is superior to other

optimization algorithms. In addition, the online optimal method can also be used to solve such problems, and the research on online optimal methods is endless. For example, Yin et al. proposed the multivariate extremum seeking approach with the Newton method [57] and the multivariate fractional-order gradient-based extremum seeking approach [58], and the authors also verified their feasibility and advantages. Therefore, the online optimal method will be an important direction of parameter optimization in the future.

In the above context, the proposed optimization algorithms belong to the optimization algorithm of the population class but generally belong to the optimal problem [59]. That is to say, when an individual is in an optimal position, but other individuals cannot know whether this position is the local optimal or the global optimal, then all other individuals will move towards this optimal position and the whole population will fall into the local optimal eventually. At present, in order to solve this problem, the inertia weight of some individual movements can be increased so as to jump out of the local optimal through the high weight motion. But increasing the weight of a motion only works in one individual or one position adjustment, and it not only does not work in other positions but also may be the reason for falling into the local optimal. When an individual falls into the optimal position, the gradient of its position function changes greatly. If the individual moves than it normally does at the next position more, it is still near the optimal position after the move. In other words, each individual has an impulse associated with the previous step, which means the impulse is large when the previous step is large or the impulse is small when the previous step is small. However, the individual trapped in the optimal still does not rush out of the current optimal state after a certain number of moves, so it can be considered that this position is the global optimal position. In addition, it is also an effective method to improve the optimization algorithm by improving the initialization of the population. The OBL method proposed by Tizhoosh is one of the most effective methods to improve the initialization of population [60]. This algorithm increases the diversity of the search range and global search ability by adding the opposite population in the initial population, which avoids the whole population falling into the local optimal from the origin. When the whole population is searching for the global optimal, if the current optimal individual falls into the local optimal, then it may mislead other individuals to also enter the local optimal. However, the opposite of the individual does not usually go with it, but away from it. Therefore, the OBL method can effectively prevent the whole population from falling into the local optimum to some extent.

Under the same working conditions and the same bearing type, the vibration signals between different fault types are different because of different bearing fault types and fault diameters [61]. In addition, because the gearbox load is often changing according to the actual work requirements, the bearing of the axial and radial loads is constantly changing in practice. Too large or too small axial and radial loads may aggravate the vibration of the bearing [62] or increase the noise, or even produce screaming [63],

which may lead to large fluctuation of the vibration signal and further affect the classification performance of the algorithm. In other words, an algorithm may only work properly under one load and not under other loads. In the actual gearbox fault diagnosis, in order to detect the health condition of a gearbox bearing, the load is deliberately set to a certain value, that is to say, to stop the normal operation of the gearbox to detect the health condition of the bearing, which will greatly reduce the economic benefit of the gearbox. Therefore, it is of great significance to consider various load conditions when studying intelligent fault diagnosis of bearings.

Based on the above, this paper proposed a novel bearing intelligent fault diagnosis method based on the kernel extreme learning machine (KELM). A novel krill herd algorithm (NKH) is used to optimize kernel function parameters  $\sigma$  and the error penalty factor  $C$  in the KELM. Opposite-based learning (OBL) and impulse operator are introduced in the optimization algorithm to improve the global search ability of the individual krill and prevent the krill group from falling into the local optimal and increase the robustness of the algorithm. Firstly, in order to verify the correctness of the algorithm proposed in this paper, the bearing data set of Case Western Reserve University is used for this experiment, and the influence of different loads on the experimental results is considered in this experiment. Then, in order to test the performance of the intelligent fault diagnosis method proposed in this paper, it is compared with other methods based on machine learning.

## 2. Feature Extraction

**2.1. Multiscale Operation.** In the analysis of fault vibration signals, a more appropriate choice of the time scale usually means that more useful information can be obtained from the original signal. However, in the engineering practice, the best time scale of the original signal is often unknown. In order to solve this problem, the signal of multiple time scales should be considered comprehensively.

Multiscale allows entropy to be extended to multiple time scales to provide an additional perspective when time scales are uncertain. Like other entropy measurement methods, multiscale entropy aims to evaluate the complexity of the time series. Therefore, analyzing the problem through multiple time scales will obtain more information. For a given time series  $x = \{x_1, x_2, x_3, \dots, x_N\}$  whose length is  $N$ , the conversion equation of the multiscale operation is as follows:

$$u_j^{(\tau)} = \frac{1}{\tau} \sum_{b=(j-1)\tau+1}^{j\tau} x_b, \quad 1 \leq j \leq \left\lceil \frac{L}{\tau} \right\rceil = N, \quad (1)$$

where  $u_j^{(\tau)}$  represents the  $j$ th element in the sequence  $u$  with the time scale of  $\tau$ ,  $N$  represents the length of the sequence  $u$ , and  $x_b$  represents the  $b$ th element in the sequence  $x$ .

A multiscale coarse granulation operation with a time scale of 2 and 3, respectively, is shown in Figures 1 and 2.

**2.2. Dispersion Entropy (DE).** Dispersion entropy can be used to express the dispersion degree of a time series. For a given time series  $x = \{x_1, x_2, x_3, \dots, x_N\}$  whose length is  $N$ , the DE consists of the following four steps:

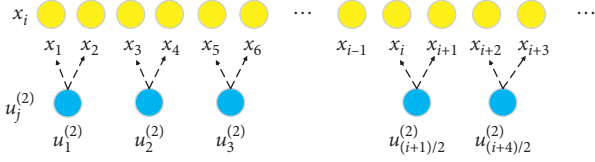


FIGURE 1: Multiscale coarse granulation of the time series with scale 2.

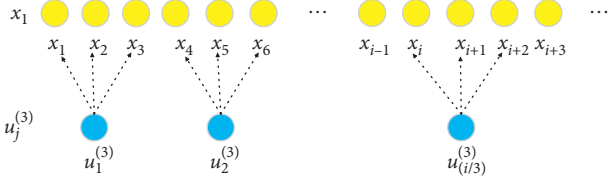


FIGURE 2: Multiscale coarse granulation of the time series with scale 3.

- (1) Firstly,  $x$  is linearly mapped to a sorting sequence of positive integers from 1 to  $c$ . However, considering the irregularity of the signal, that is to say, whether the maximum or minimum value of a signal is much larger or much smaller than the mean or median value of the signal, the signal is mostly concentrated in a subset of the sequence. Therefore,  $x$  is mapped to a sequence  $y = \{y_1, y_2, y_3, \dots, y_N\}$  from 0 to 1 through the normal cumulative distribution function (NCDF). The equation of the normal cumulative distribution function (NCDF) is as follows:

$$y_j = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{-((x-\mu)^2)/(2\sigma^2)} dt, \quad (2)$$

$$p(\pi_{v_0 v_1 v_2 \dots v_{m-1}}) = \frac{\text{number}\{i \mid i \leq N - (m-1)d, z_i^{m,c} \text{ has type } \pi_{v_0 v_1 v_2 \dots v_{m-1}}\}}{N - (m-1)d}. \quad (6)$$

- (4) According to the definition of entropy proposed by Shannon, dispersion entropy can be obtained as follows:

$$\text{DE}(x, m, c, d) = - \sum_{\pi=1}^{c^m} p(\pi_{v_0 v_1 v_2 \dots v_{m-1}}) \times \ln(p(\pi_{v_0 v_1 v_2 \dots v_{m-1}})). \quad (7)$$

- (5) Finally, the standard dispersion entropy is obtained from the dispersion entropy standardization equation. It is expressed as follows:

$$\text{NDE} = \frac{\text{DE}}{\ln(c^m)}. \quad (8)$$

Here is a simple example: supposing there is a time series  $x = \{9, 8, 1, 12, 5, -3, 1.5, 8.01, 2.99, 4, -1, 10\}$  whose length is 12, the embedding dimension  $m$  is 2, the time delay  $d$  is 1, and the types of classifications  $c$  are 3. The sequence  $y = \{0.82, 0.75, 0.21, 0.94, 0.52, 0.05, 0.241, 0.75, 0.35, 0.43, 0.11, 0.87\}$  is obtained by the standard normal cumulative distribution function. And the classification sequence is  $z = \{3, 3, 1, 3, 2, 1, 1, 3, 2, 2, 1, 3\}$ . In this example, the

where  $\sigma$  is the standard deviation of the sequence and  $\mu$  is the mean of the sequence.

Secondly, after obtaining the sequence  $y$ , it is linearly mapped to  $z$ , and the equation is as follows:

$$z_j^c = \text{round}(c \times y_j + 0.5), \quad (3)$$

where  $c$  represents the number of types of classifications,  $z_j^c$  represents the  $j$ th element of the sequence, and  $\text{round}(\cdot)$  represents the rounding operation.

- (2) The equation of the embedding vector  $z_i^{m,c}$  is as follows:

$$z_i^{m,c} = \{z_i^c, z_{i+d}^c, z_{i+2d}^c, \dots, z_{i+(m-1)d}^c\}, \quad (4)$$

where  $i = 1, 2, 3, \dots, N - (m-1)d$ ,  $m$  is the embedding dimension, and  $d$  is the time delay.

Then, each embedded vector is mapped to the dispersion pattern; the mapping relationship is as follows:

$$\begin{cases} z_i^c = v_0, z_{i+d}^c = v_1, \dots, z_{i+(m-1)d}^c = v_{m-1}, \\ z_i^{m,c} \longrightarrow \pi_{v_0 v_1 v_2 \dots v_{m-1}}. \end{cases} \quad (5)$$

So there are  $c^m$  different dispersion patterns.

- (3) For each possible dispersion pattern, the equation of its relative probability is

possibilities of the dispersion pattern are  $3^2 = 9$ , that is,  $\pi_{v_0 v_1 v_2 \dots v_{m-1}} = \{\pi_{11}, \pi_{12}, \pi_{13}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{31}, \pi_{32}, \pi_{33}\}$ . Its detailed equation is as follows:

$$\begin{aligned} z_1^{2,3} &= \{3, 3\} \longrightarrow \pi_{33}, \\ z_2^{2,3} &= \{3, 1\} \longrightarrow \pi_{31}, \\ z_3^{2,3} &= \{1, 3\} \longrightarrow \pi_{13}, \\ z_4^{2,3} &= \{3, 2\} \longrightarrow \pi_{32}, \\ z_5^{2,3} &= \{2, 1\} \longrightarrow \pi_{21}, \\ z_6^{2,3} &= \{1, 1\} \longrightarrow \pi_{11}, \\ z_7^{2,3} &= \{1, 3\} \longrightarrow \pi_{13}, \\ z_8^{2,3} &= \{3, 2\} \longrightarrow \pi_{32}, \\ z_9^{2,3} &= \{2, 2\} \longrightarrow \pi_{22}, \\ z_{10}^{2,3} &= \{2, 1\} \longrightarrow \pi_{21}, \\ z_{11}^{2,3} &= \{1, 3\} \longrightarrow \pi_{13}. \end{aligned} \quad (9)$$



So the probability of each dispersion pattern is as follows:

$$\begin{aligned}
p(\pi_{11}) &= \frac{1}{11}, \\
p(\pi_{12}) &= \frac{0}{11}, \\
p(\pi_{13}) &= \frac{3}{11}, \\
p(\pi_{21}) &= \frac{2}{11}, \\
p(\pi_{22}) &= \frac{1}{11}, \\
p(\pi_{23}) &= \frac{0}{11}, \\
p(\pi_{31}) &= \frac{1}{11}, \\
p(\pi_{32}) &= \frac{2}{11}, \\
p(\pi_{33}) &= \frac{1}{11}.
\end{aligned} \tag{10}$$

Its dispersion entropy is calculated as  $DE(x, m, c, d) = DE(x_i, 2, 3, 1) = 1.8642$  and normalized as  $NDE = (DE/\ln(3^2)) = 0.85$ .

**2.3. Multiscale Dispersion Entropy (MDE).** When dealing with the fault of the bearing vibration signal, it is often difficult to know the most suitable time scales, and signal nonstationarity and irregularity tend to be very strong; in order to solve this problem well, Azami et al. proposed multiscale discrete entropy (MDE) in 2017 [47] and proved that the MDE approach based on processing of the non-stationary signal has a strong ability of feature extraction. The flow of MDE is shown in Figure 3.

$$x = \{9, 8, 7.2, 5.3, 6.4, 3.2, 0.89, 10.6, 5.2, 4.3, 6.4, 7.8\}. \tag{11}$$

Here is a simple example: supposing there is a set of time series, the embedding dimension  $m$  is 2, the time delay  $d$  is 1, the types of classifications  $c$  are 3, and the time scale is 3. Firstly, the coarse granulation of the sequence is carried out according to the time scale. And the time scale 2 of the sequence  $y^{(2)} = \{8.5, 6.25, 4.8, 5.745, 4.75, 7.2\}$  and the time scale 3 of the sequence  $y^{(3)} = \{8.07, 4.97, 5.56, 6.17\}$  are obtained. Then, the dispersion entropy of each sequence is calculated to get  $MDE = \{DE^{(1)}, DE^{(2)}, DE^{(3)}\} = \{2.0198, 1.0549, 1.0986\}$ . Finally, it is normalized to  $NMDE = \{0.9193, 0.4801, 0.5\}$ .

### 3. The Proposal of NKH-KELM

**3.1. Novel Krill Herd Algorithm (NKH).** Gandomi proposed the krill herd algorithm (KH) in 2012 by studying krill herd

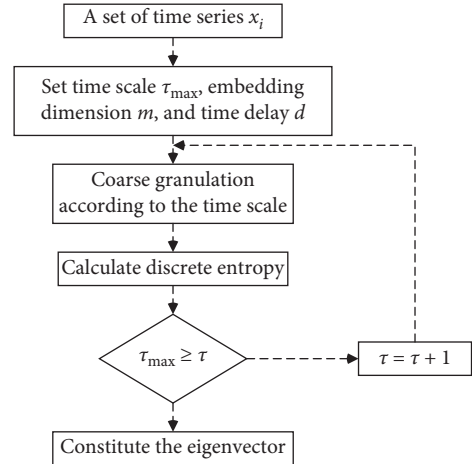


FIGURE 3: Flow chart of MDE.

activity rules [56]. In this algorithm, the solution of the optimization problem is expressed by the position state of the krill, and the optimal solution is constantly sought by the change of the individual position in the process of krill foraging. The Lagrange model is used to describe the  $K$ th location change of the  $i$ th krill in a group of krills including  $L$  krills during the foraging process:

$$\frac{dX_i}{dt} = N_i + F_i + D_i, \tag{12}$$

where  $N_i$  is the induced movement among krills,  $F_i$  is the foraging movement among krills, and  $D_i$  is the random disturbance of krills.

The induced motion velocity  $N_i^{(k)}$  of the  $i$ th krill is expressed as follows:

$$\begin{aligned}
N_i^{(k)} &= N_{\max} \alpha_i^{(k)} + \omega_n N_i^{(k-1)}, \\
\alpha_i^{(k)} &= \alpha_{i\text{local}}^{(k)} + \alpha_{i\text{target}}^{(k)}, \\
\alpha_i^{\text{local}} &= \sum_{j=1}^L K_{ij} X_{ij},
\end{aligned} \tag{13}$$

$$X_{ij} = \frac{X_j - X_i}{\|X_j - X_i\| + \varepsilon},$$

$$K_{ij} = \frac{K_i - K_j}{K^{\text{worst}} - K^{\text{best}}},$$

where  $N_{\max}$  is the maximum induced velocity;  $\omega_n$  is the inertia weight ( $\omega_n \in [0, 1]$ );  $\alpha_i^{(k)}$  is the krill movement induced by the direction;  $\alpha_{i\text{local}}^{(k)}$  is the local impact from the neighbor krill;  $\alpha_{i\text{target}}^{(k)}$  is the current local influence from the optimal krill;  $K_i$  and  $K_j$  are the fitness function values of krills  $i$  and  $j$ , respectively;  $K^{\text{worst}}$  and  $K^{\text{best}}$  are the worst and the best fitness function value of the current krill, respectively;  $K_{ij}$  is the fitness function value of the  $i$ th krill related to the  $j$ th krill;  $X_{ij}$  is the position of the  $i$ th krill related to the  $j$ th krill; and  $\varepsilon$  is a normal number which avoided the singularity of equation (18).

The foraging movement  $F_i^{(k)}$  of the  $i$ th krill is represented by the following equation:

$$F_i^{(k)} = v_f(\beta_{\text{ifood}}^{(k)} + \beta_{\text{ibest}}^{(k)}) + \omega_f F_i^{(k-1)}, \quad (14)$$

where  $v_f$  is the speed of foraging;  $F_i^{(k-1)}$  is the direction of the last foraging;  $\beta_{\text{ifood}}^{(k)}$  is the attraction of food to the  $i$ th krill;  $\beta_{\text{ibest}}^{(k)}$  is the influence of the optimal krill on the  $i$ th krill until the current moment; and  $\omega_f$  is the inertia weight ( $\omega_f \in [0, 1]$ ).

The random disturbance  $D_i^{(k)}$  of the  $i$ th krill is expressed by the following equation:

$$D_i^{(k)} = D_{\text{max}}^{(k)} \delta, \quad (15)$$

where  $D_{\text{max}}^{(k)}$  is the maximum velocity of random diffusion and  $\delta$  is the direction vector of random diffusion disturbance ( $\delta \in [-1, 1]$ ).

In order to improve the global search ability and convergence speed of the krill herd algorithm, the opposite population in the initialization of the krill herd is added. Assuming that the position of the krill is  $X_i = (c_1, c_2, c_3, \dots, c_D)$  in the  $D$ -dimensional space, where  $c_j \in [c_{\text{min}}^{(j)}, c_{\text{max}}^{(j)}]$ ,  $j \in [1, D]$ , the position equation of the opposite krill is

$$\begin{cases} X'_i = (c'_1, c'_2, c'_3, \dots, c'_D), \\ c'_j = c_{\text{min}}^{(j)} + c_{\text{max}}^{(j)} - c_j. \end{cases} \quad (16)$$

Through opposite-based learning (OBL), the diversity of the krill group is increased so as to increase the exploration scope and robustness of the krill group.

There may not only be one optimal solution in the  $D$ -dimensional space that describes the krill herd algorithm; that is to say, there are many local solutions. The impulse operator is introduced to make a krill whose previous movement is very large have a high velocity in the next movement so as to rush out of the local optimal solution. So equation (8) can be improved to get the following equation:

$$\frac{dX_i^{\text{new}}}{dt} = N_i + F_i + D_i + p \frac{d^2 X_i^{\text{old}}}{dt^2}, \quad (17)$$

where  $p$  is the impulse coefficient and  $(d^2 X_i^{\text{old}})/(dt^2)$  is the acceleration of the  $i$ th krill in the last movement. The flow chart of the NKH is shown in Figure 4.

**3.2. The Kernel Extreme Learning Machine Optimized by Novel Krill Herd Algorithm (NKH-KELM).** Consider a sample data set  $(x_i, t_i)$  whose capacity is  $N$ , where  $x_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{iN}]^T$  is the sample input data,  $N$  is the length of each vector  $x_i$ , and  $t_i = [t_1, t_2, t_3, \dots, t_i]^T$  is the sample output value. For a single-layer forward neural network (SLFN) with  $L$  nodes of the hidden layer, its flow is shown in Figure 5.

In Figure 5,  $h(a_i x_i + b_i)$  is the activation function,  $a_i$  is the weight vector between the  $i$ th node of the hidden layer and the node of the input layer,  $b_i$  is the bias of the  $i$ th hidden layer,  $\beta_i$  is the weight vector between the  $i$ th node of the hidden layer and the output layer, and  $y_i$  is the output value. So the output equation is

$$y_i = \sum_{i=1}^L \beta_i h(a_i x_i + b_i), \quad i = 1, 2, 3, \dots, N. \quad (18)$$

When the actual output of the extreme learning machine can approximate the expected output infinitely, in other words,  $\sum_{i=1}^N \|y_i - t_i\| = 0$ ,

$$t_i = \sum_{i=1}^L \beta_i h(a_i x_i + b_i), \quad i = 1, 2, 3, \dots, N. \quad (19)$$

The output of the hidden layer is represented by a matrix  $H = \begin{bmatrix} h_1(a_1 x_1 + b_1) & \dots & h_1(a_1 x_1 + b_1) \\ \vdots & \ddots & \vdots \\ h_L(a_L x_1 + b_1) & \dots & h_L(a_L x_1 + b_1) \end{bmatrix}$ , and  $T$  is the expected output vectors. The least-square method is used to determine the output weight vector  $\beta_i$  of the ELM. Its equation is as follows:

$$\beta = H^+ T = H^T \left( \frac{I}{C} + HH^T \right)^{-1} T, \quad (20)$$

where  $H^+$  is the Moore–Penrose generalized inverse matrix of the hidden-layer output matrix,  $C$  is the penalty factor, and  $I$  is the identity matrix. In order to prevent the roots from deviating from zero, a constant matrix  $I/C$  is added to the matrix  $HH^T$  to improve the stability and generalization ability of the results.

In terms of classification, for any linear indivisible data set  $X$  in a low-dimensional space, there is always a mapping  $k$  that maps the samples to a high-dimensional feature space  $X \rightarrow k(x)$ , which enables the data set to be linearly separated. However, the high-dimensional feature space will greatly increase the calculation amount, and more importantly, the generalization performance of the algorithm will decline with the increase of the dimension, which is very unfavorable to the intelligent fault diagnosis of bearings. In other words, an algorithm is only applicable to intelligent fault diagnosis of bearings under one working condition, but it is often not applicable to intelligent fault diagnosis of bearings under another working condition. The kernel extreme learning machine (KELM) is borrowed from the idea of the support vector machine (SVM) and adopts kernel function to replace the feature mapping of the ELM hidden-layer node, which can avoid the problem of the dimension disaster to a certain extent. Compared with the ELM, the KELM does not need to artificially determine the number of hidden-layer nodes and also avoids the randomization operation of input weight and input bias, thus improving the generalization ability and classification accuracy of the model. In the KELM, only the appropriate kernel function parameters need to be selected to obtain the output weight.

For the KELM, the output vector is expressed as follows:

$$y = \ker(x) H^T \left( \frac{I}{C} + HH^T \right)^{-1} T. \quad (21)$$

The kernel function is expressed as follows:

$$\ker(x_i, x_j) = k(x_i) k(x_j). \quad (22)$$

In the selection of kernel function, since there is no prior knowledge about the classification of fault bearings, and the

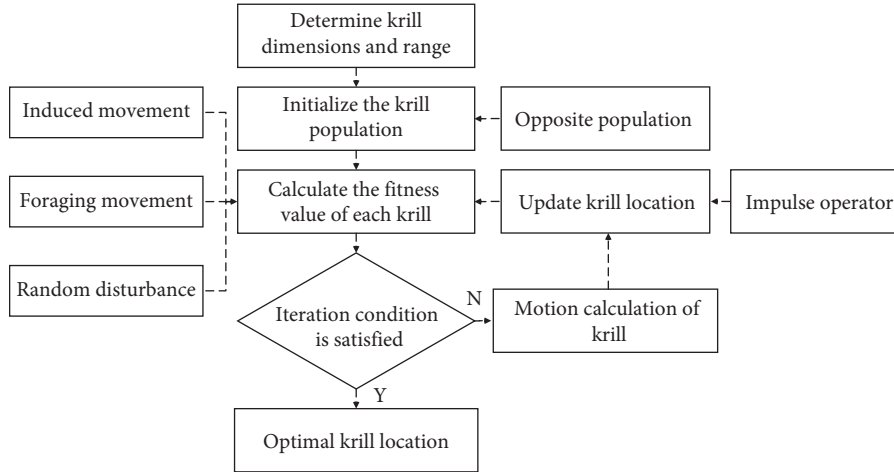


FIGURE 4: Flow chart of the NKH.

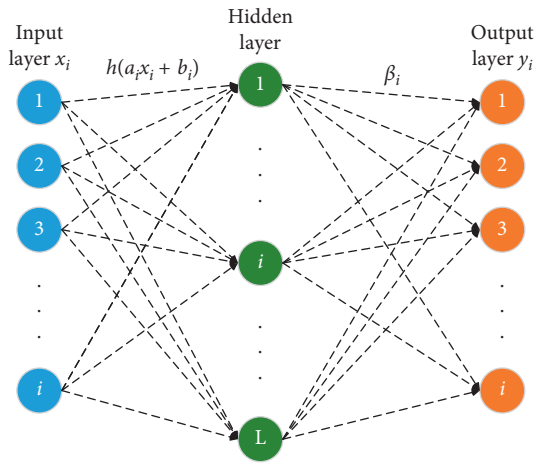


FIGURE 5: Flow chart of the ELM.

KELM is often expected to have a strong ability to explore the local optimum after a certain time of use, radial basis function (RBF) is selected as the kernel function of the KELM [64]. The equation of the Gaussian kernel function is as follows:

$$\ker(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (23)$$

where  $x_i$  is the input vector during training,  $x_j$  is the input vector during testing, and  $\sigma$  is the width parameter of the kernel function.

So the output equation of the KELM is

$$y = \begin{bmatrix} \ker(x, x_1) \\ \vdots \\ \ker(x, x_L) \end{bmatrix}^T H^T \left( \frac{I}{C} + HH^T \right)^{-1} T, \quad (24)$$

where  $L$  is the dimension of the input vector.

Considering that the kernel function parameters  $\sigma$  and the error penalty factor  $C$  in the KELM have a great influence on the results, the novel krill herd algorithm (NKH) is used to optimize the two parameters to improve the classification accuracy and robustness of the KELM. The generalized

mean-square error GMSE can directly reflect the regression performance of the KELM [65]. The equation of the GMSE is

$$\text{GMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (25)$$

where  $n$  is the number of test samples and  $\hat{y}_i$  and  $y_i$  are the estimated and actual values of test samples, respectively.

In this paper, the KELM algorithm is used to solve the classification problem, and the accuracy of the algorithm classification can be verified by 10-fold cross-validation. In other words, the data set is divided into ten parts, and turns are taken to use nine of them as training data and the remaining one as test data, which can be used to evaluate the performance of parameters, reduce overfitting to a certain extent, and obtain as much effective information as possible from limited data to improve model performance [66].

Then, the GMSEs which are obtained ten times are averaged. Finally, the result is used as the fitness function  $K$ :

$$K = \overline{\text{GMSE}} = \frac{1}{k} \sum_{i=1}^{10} \text{GMSE}, \quad (26)$$

where  $k$  is the number of experiments and is taken as 10.

From what has been discussed above, the flow chart of NKH-KELM is shown in Figure 6. As can be seen from Figure 6, the fault diagnosis method proposed in this paper consists of three parts, namely, fault feature extraction, NKH-KELM training, and NKH-KELM testing. In feature extraction, vibration signals are collected by accelerometers, sample design is carried out, and MDE feature extraction is finally carried out so as to extract useful information from the original vibration signals. In NKH-KELM training, the NKH with OBL and impulse operator is used to optimize the width parameters  $\sigma$  and penalty factors  $C$  of kernel function, and then the feature space extracted from the training set is imported into the KELM for training. In the testing of NKH-KELM, the feature space extracted from the feature set is imported into the classifier which is trained in the previous step so as to conduct fault diagnosis and verify the feasibility of the method proposed in this paper.

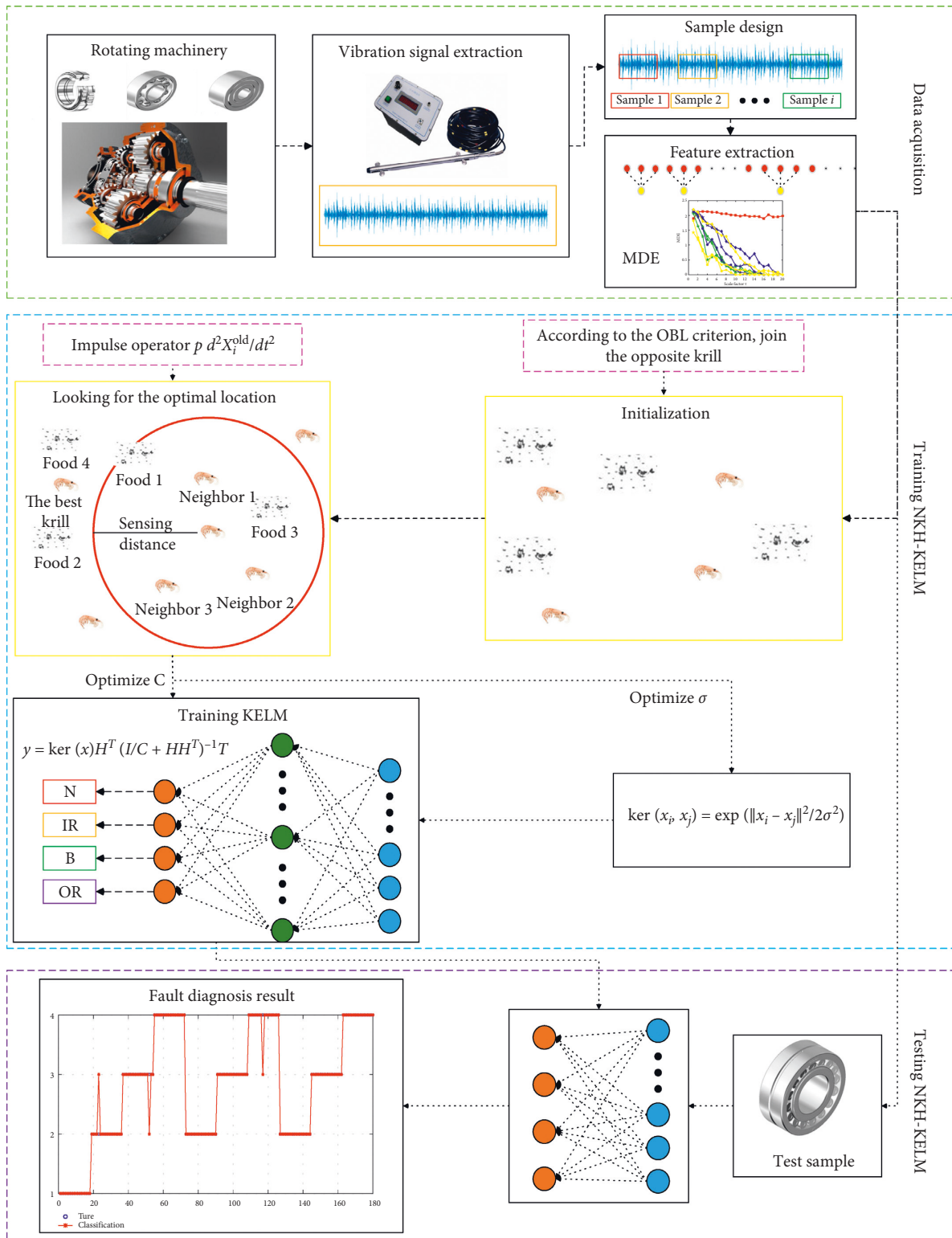


FIGURE 6: Flow chart of NKH-KELM.



## 4. Bearing Intelligent Fault Diagnosis Based on NKH-KELM

### 4.1. Fault Diagnosis Experiment Based on Case Western Reserve University Bearing Data Set

**4.1.1. Description of Data.** In order to verify the feasibility of the proposed method, this paper adopts the test bearing data set provided by Case Western Reserve University to carry out the experiment. The program code of all algorithms was written in MATLAB R2016a, and then it was run on a computer with CPU i7-5500U@2.40 GHz, RAM 12.00 GB, and 64 bit Win10 operating system.

As shown in Figure 7, the bench consists of a 2 horsepower motor (left), a torque sensor/encoder (center), a dynamometer (right), and other control equipment (not shown in Figure 7). The motor shaft is supported by test bearings. The single point failure of all test bearings was caused by Electric Discharge Machining (EDM), and the fault diameters were 0.007 inch, 0.014 inch, and 0.021 inch, respectively. The driving-end bearing (this paper only considers the driving-end bearing data) adopts the 6205-2RS bearing of SKF company. All vibration data were collected by an accelerometer, which was placed at the drive end of the motor housing with the sampling frequency of 12 kHz.

The larger the fault size, the greater the bearing fault degree. 0.007 inch, 0.014 inch, and 0.021 inch are considered in this paper. In addition, for the motor load, the working conditions with no load (0 hp), small load (1 hp), and large load (3 hp) are considered. For the convenience of the following instructions, a notation is used to represent the tested data. For example, IR014\_3 represents the bearing of an inner ring failure type with a load of 3 hp and a failure diameter of 0.014 inch. Table 1 shows the specifications of test bearings.

As can be seen from Table 1, there are 30 kinds of test bearing data sets. Since each fault type corresponds to 3 fault diameters besides a healthy bearing, there are 10 kinds of bearings. For each test bearing data set, every 5000 sampling points are selected as one sample, so each data set has 24 samples and the whole data set can be divided into 720 samples. The time-domain graphs of a sample in the bearing data of each test under three loading conditions are shown in Figures 8–10, respectively.

From the above time-domain diagram, it can be clearly seen that, under the same load, the same fault type, and different fault diameters, the peak size and peak occurrence time of the time-domain diagram are significantly different. For example, when the load is 3 hp, there are three different time-domain diagrams corresponding to the ball fault. In addition, in the case of the same fault type and fault diameter under different loads, there are obvious differences in the peak size and peak occurrence time of time-domain graphs. Moreover, as can be seen from Table 2, the maximum value of the signal may increase or decrease with the increase of load. For example, in the rolling body fault with the fault diameter of 0.007 inch, with the increase of load, the maximum value of the signal is 3.4254, 3.0960, and 3.1301 successively. Moreover, with the increase of load, the

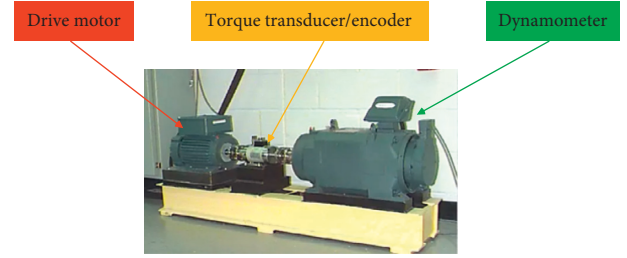


FIGURE 7: Rolling bearing failure simulation test bench of Case Western Reserve University.

TABLE 1: Test bearing specifications.

Health condition	Fault diameter (inch)	Load (hp)	Speed (rpm)	Notation
Healthy (N)	—	0	1797	N000_0
		1	1772	N000_1
		3	1730	N000_3
Inner ring fault (IR)	0.007, 0.014, 0.021	0	1797	IR007_0, IR014_0, IR021_0
		1	1772	IR007_1, IR014_1, IR021_1
		3	1730	IR007_3, IR014_3, IR021_3
Ball fault (B)	0.007, 0.014, 0.021	0	1797	B007_0, B014_0, B021_0
		1	1772	B007_1, B014_1, B021_1
		3	1730	B007_3, B014_3, B021_3
Outer ring fault (OR)	0.007, 0.014, 0.021	0	1797	OR007_0, OR014_0, OR021_0
		1	1772	OR007_1, OR014_1, OR021_1
		3	1730	OR007_3, OR014_3, OR021_3

minimum value, peak value, and root-mean-square value all change to some extent. In addition, with the same load and fault type, but with different fault diameters, the peak values of time-domain graphs also have obvious differences. Therefore, this further proves the necessity of considering load and fault diameter in intelligent fault diagnosis.

### 4.1.2. Fault Bearing Classification in Case Western Reserve University Bearing Data Set Based on NKH-KELM.

According to the intelligent fault diagnosis method proposed in this paper, firstly, MDE is used to extract multiscale features from the original vibration signals. For the choice of the time scale  $\tau$ , if the time scale is too small, that is, the observation angle of the signal is too small, it is impossible to extract more useful information and affect the subsequent operations. Conversely, if the time scale  $\tau$  is too large, it will increase the memory requirements and computing time of

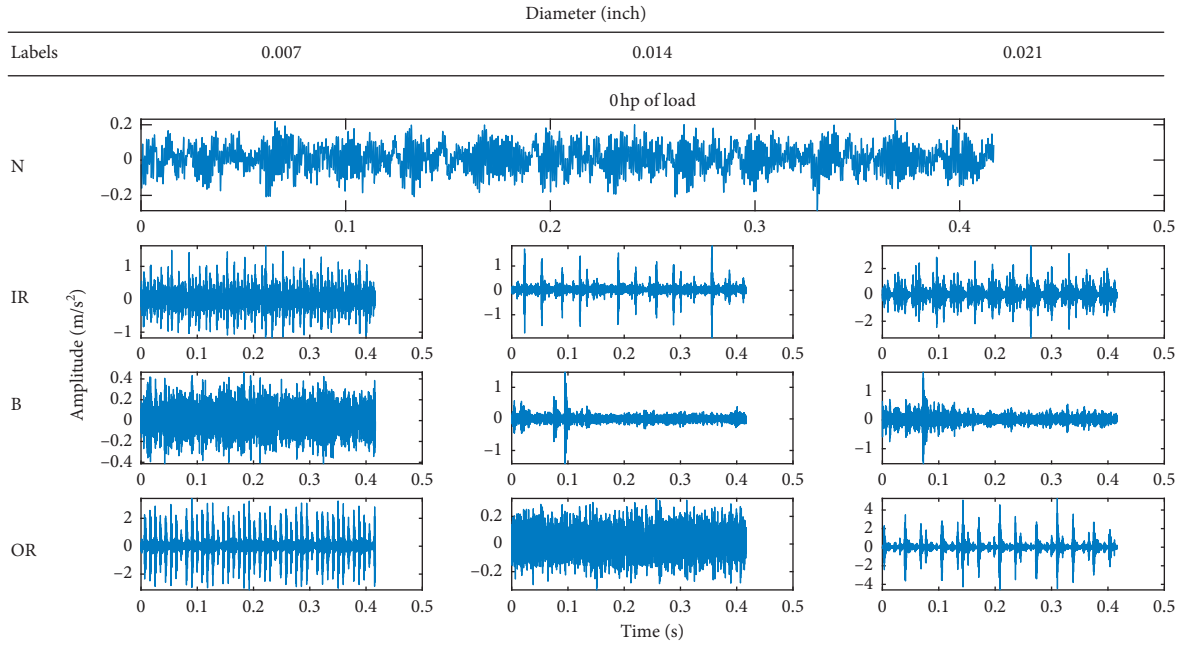


FIGURE 8: Time-domain graph of the original signal when the load is 0 hp.

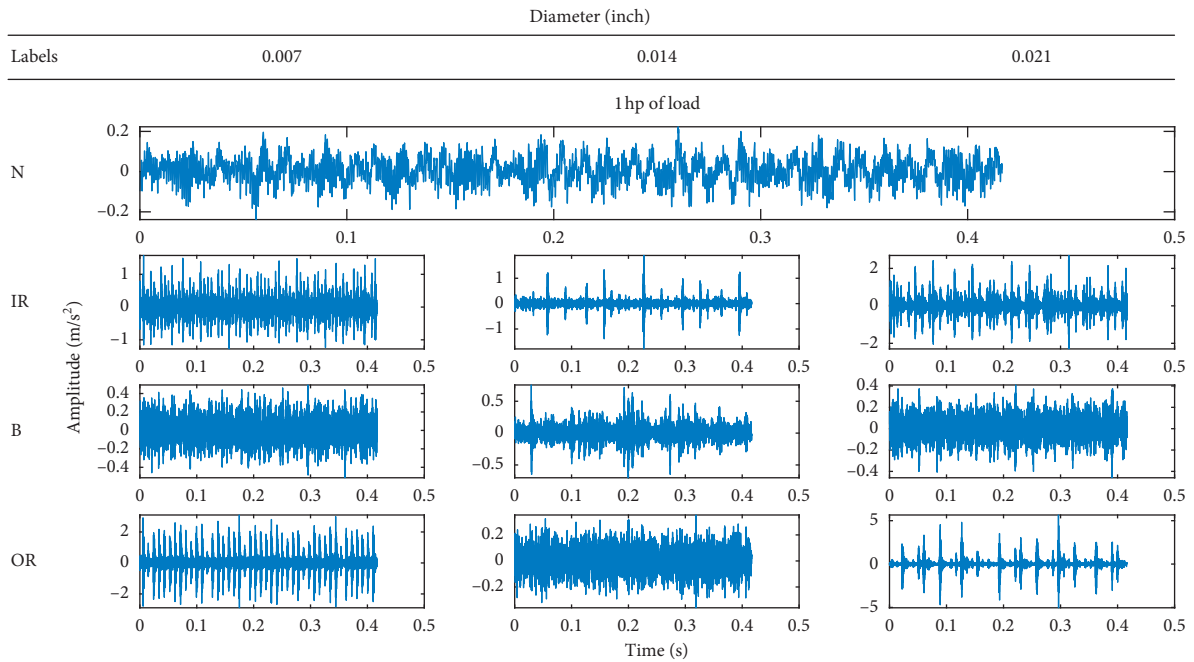


FIGURE 9: Time-domain graph of the original signal when the load is 1 hp.

the computer. Secondly, for the selection of the embedding dimension  $m$ , if  $m$  is too small, it is difficult to detect the tiny dynamic behavior of signals. Moreover, as can be seen from the time-domain graphs in the previous section, the tiny dynamic behavior of signals may bring about changes in the fault type or fault diameter. On the contrary, if the embedding dimension  $m$  is too large, although it can obtain more reliable entropy results, it will greatly increase the memory requirements and computing time of the computer.

Therefore, according to the above contents, in order to extract more useful information on the limited computer memory, the time scale  $\tau$  is assumed to be 20 and the embedding dimension  $m$  to be 2 [67]. In addition, when the time delay  $d$  is greater than 1, aliasing may occur. Therefore, the time delay  $d$  in this paper is taken as 1. Figure 11 shows the MDE extraction results of the health condition of the bearing in the test bearing data set under three different loads.

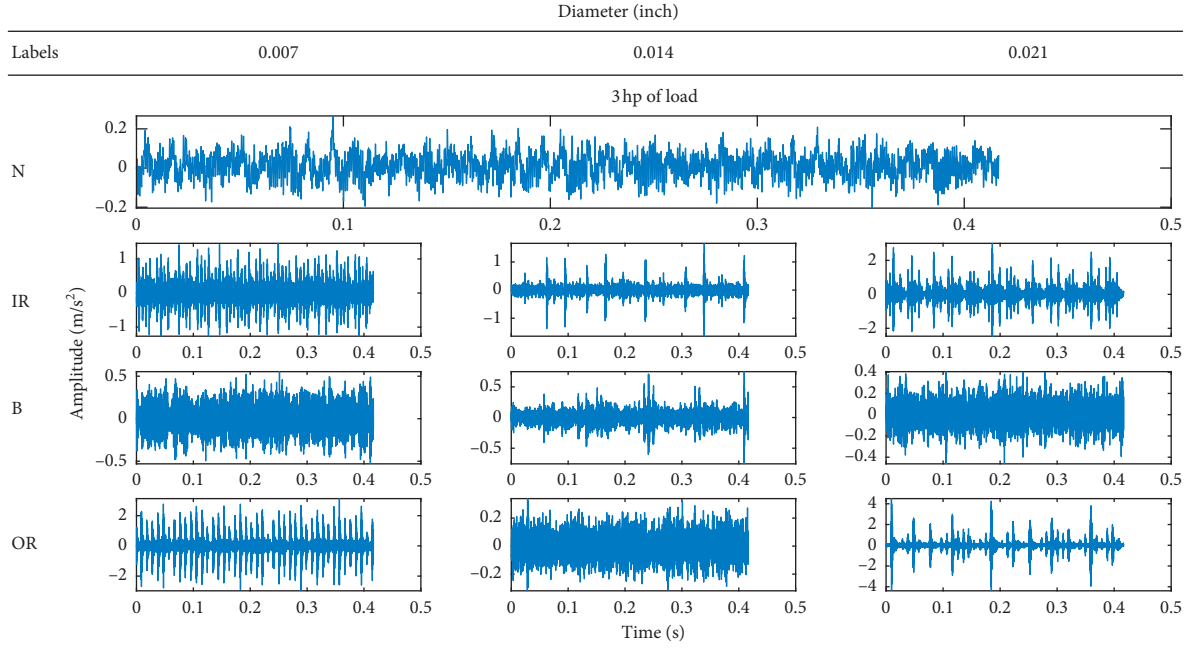


FIGURE 10: Time-domain graph of the original signal when the load is 3 hp.

TABLE 2: Numerical analysis of the healthy bearing and faulty bearing with a fault diameter of 0.007 inch.

Notation	Load (hp)	Max.	Min.	Peak-to-peak value	Root mean square
N000_0	0	0.2316	-0.2866	0.5182	0.0740
N000_1	1	0.2230	-0.2384	0.4615	0.0667
N000_3	3	0.2660	-0.2047	0.4706	0.0654
IR007_0	0	1.6258	-1.1751	2.8009	0.2876
IR007_1	1	1.5768	-1.2784	2.8548	0.2944
IR007_3	3	1.4579	-1.2668	2.7247	0.3178
B007_0	0	0.4644	-0.4136	0.8780	0.1348
B007_1	1	0.4928	-0.5136	1.0064	0.1395
B007_3	3	0.5534	-0.5306	1.0839	0.1585
OR007_0	0	3.4254	-3.1346	6.5599	0.6943
OR007_1	1	3.0960	-2.8946	5.9906	0.5953
OR007_3	3	3.1301	-2.9742	6.1043	0.5818

In this experiment, the extracted feature space is a matrix with a size of  $720 \times 20$ , in which 720 is the number of samples and the dimension of each feature vector is 20. Before using NKH-KELM for classification, the number of training samples is selected as 540, the number of test samples is selected as 180, and the classification label is selected as 4 (N, IR, B, and OR). Firstly, the training sample set and the novel krill herd algorithm are used to optimize the kernel function parameters  $\sigma$  and the error penalty factor  $C$  in the KELM. In the initialization of the krill group, the number of krills is selected to be 25, and the maximum allowable number of krill position updates is 200 [68]. After OBL, 25 opposition krills are produced, so the individual number of whole krill groups is 50. After introducing the impulse operator, the value of the impulse coefficient has a great influence on the robustness of the krill herd algorithm. If the impulse coefficient is too small, the krill may fall into the local optimal and is difficult to rush out of the local optimal because of the fact the krill's impulse is too

small. If the impulse coefficient is too large, then the impulse will be too large and the krill will always be out of the global optimal; that is to say, once the krill enters the global optimal, the krill herd algorithm will never reach the global optimal. When the impulse factor is greater than or equal to 1, the next move of the krill is likely to occur in a situation that will never be global optimal, which is the worst. Therefore, when discussing the selection of the impulse coefficient in this paper, it should first be set to be less than 1 and then its minimum value is set to be 0.1 and the interval between values of two adjacent groups is set to be 0.1. The experiment is repeated five times, and the results are shown in Figure 12.

It can be seen from Figure 12 that when the impulse coefficient  $p$  is 1, the optimal fitness hardly changes after reaching  $10^{-5}$ . That is to say, the phenomenon that the impulse is too large to enter the optimal solution area and the next movement will rush out of the optimal solution area may happen, which indirectly proves the correctness of the

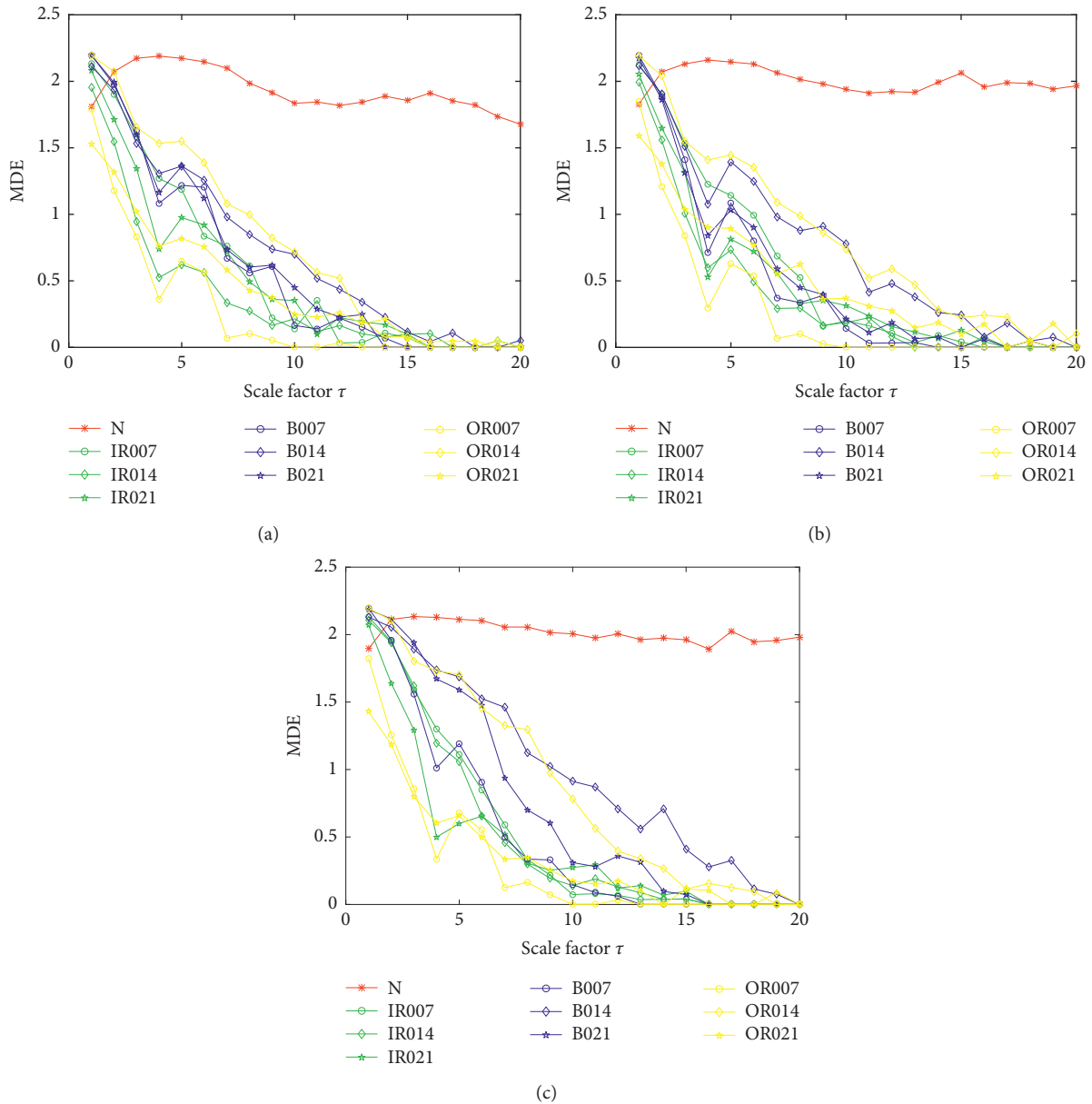


FIGURE 11: Dispersion entropy of each scale of the vibration signal of the test bearing data set under three kinds of loads. (a) Load = 0 hp. (b) Load = 1 hp. (c) Load = 3 hp.

previous hypothesis. As can be seen from Figure 12, when the impulse coefficient  $p$  is 0.6 and krill moves 200 times, the average fitness of the optimal krill individual is  $2.165 \times 10^{-9}$ , the error penalty factor  $C$  after optimization is 15, and the width parameter  $\sigma$  of the kernel function is 0.53. Secondly, the optimized KELM is used to train and test the feature space extracted from MDE. And the input of the KELM is composed of 720 samples of MDE, of which 540 are used for training and the remaining 180 samples are used for testing. Then, the process of training is according to the KELM's kernel mapping rules. Finally, the output of the KELM is obtained, that is, the prediction labels of the input sample. They were compared with the actual labels of samples, and the KELM was trained based on this basis. The experiment is repeated five times. The average fault diagnosis accuracy is

shown in Table 3, and the fault diagnosis of the training and test of a certain experiment is shown in Figure 13. In Figure 13, the abscissa is the number of sample sets, the ordinate is the state of the sample, the circle represents the actual label of the sample, and the asterisk represents the prediction label of the sample. As can be seen from Figure 13, the classification accuracy of the training process and test process reached 99.2593% and 98.3333%, respectively. In addition, the correct rate reached 100% in the diagnosis of the normal bearing set, and the other three kinds of bearings all had a few cases of wrong diagnosis. This also explains the feasibility of the proposed method. Moreover, the sample of error diagnosis mainly focuses on the rolling body, which may be related to the working environment of the rolling body and other factors.

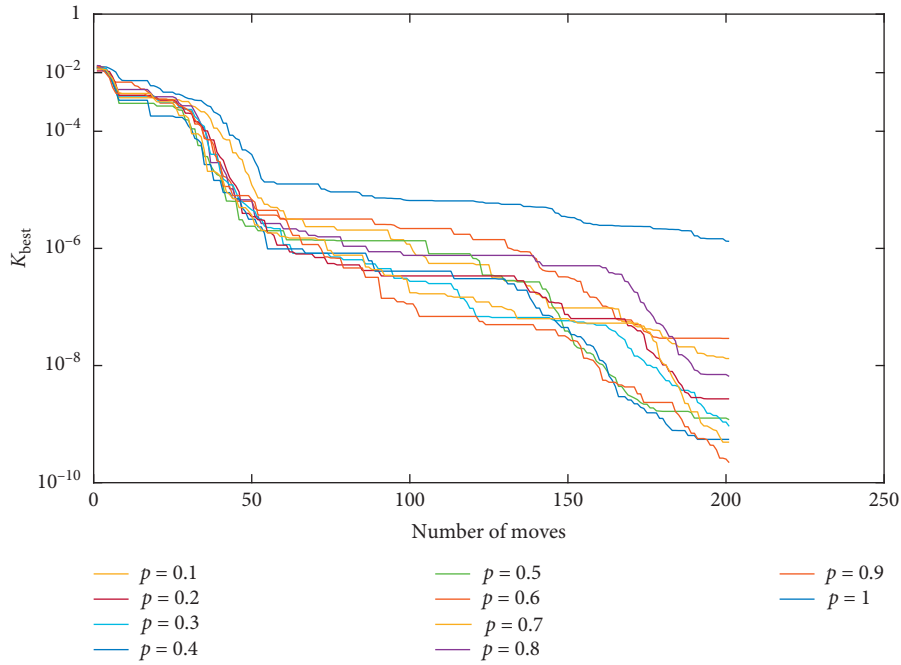


FIGURE 12: Optimal fitness value corresponding to the krill group movement under different impulse coefficients.

TABLE 3: Training accuracy and test accuracy of each type of bearings.

Types of bearings	Average accuracy of the training sample set (%)	Average accuracy of the test sample set (%)
Normal	100	100
Inner ring fault	98.9	97.6
Ball fault	99.8	97.1
Outer ring fault	97.9	98.4
Average	99.1	98.3

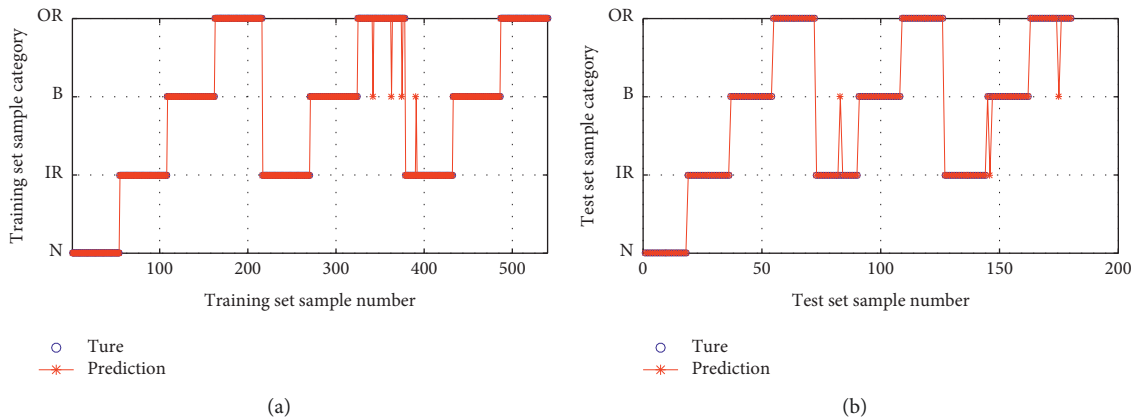


FIGURE 13: Fault diagnosis of a certain experiment. (a) Results of NKH-KELM with classification accuracy = 99.2593%. (b) Results of NKH-KELM with classification accuracy = 98.3333%.

In order to objectively test the performance of the proposed method, the same data set is imported into other algorithms, such as the K-nearest neighbor algorithm (KNN) [69], naive Bayes classification algorithm (NB) [70], symbol dynamic entropy + support vector machine (SVM) [27], radial basis network (RBF) [71], extreme learning machine

(ELM) [31], and multilayer extreme learning machine [32]. The experiment is repeated five times, and the average value of the five times is taken as the evaluation index. The results are shown in Table 4 and Figures 14 and 15.

It can be seen from Table 4 and Figure 14 that the performance of the kernel extreme learning machine



TABLE 4: Average classification accuracy of different algorithms.

Feature extraction	Classification method	Average classification accuracy (%)	Average computing time (s)	Standard deviation
MDE	The proposed method	98.7	0.6189	0.2226
	KELM + standard KH	88.3	0.6398	0.1059
	KNN	96.5	0.5398	0.1447
	NB	77.8	1.1771	0.5369
	RBF	94.8	3.1687	0.3681
SDE	SVM	95.5	3.3622	0.4369
mWAE	ELM	95.2	0.7469	0.2173
MPE	MELM	97.4	0.8312	0.0871

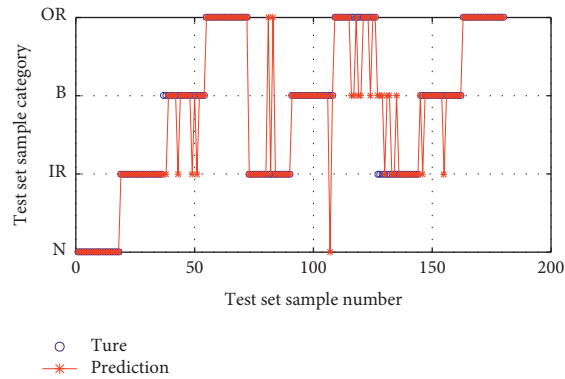


FIGURE 14: Bearing diagnosis of the KELM optimized by the standard krill herd algorithm: results of NKH-KELM with classification accuracy = 88.3333%.

optimized by the krill herd algorithm which is improved by OBL and impulse operator is better than that of the KELM optimized by the standard krill herd algorithm, which also proves the superiority of the novel krill herd algorithm proposed in this paper and the correctness of the algorithm introduced by OBL and impulse operator. Figure 15 shows the average value of the confounding matrix of the six algorithms in the five experiments, where the horizontal coordinate is a certain prediction label, the vertical coordinate is the actual label of the sample, and the number in the grid represents the proportion of the prediction label in the classification and the actual label. It can be seen from Figure 15 and Table 4 that the seven methods used for comparison and the method proposed in this paper reach 100% in the classification of healthy bearings. The phenomenon may be that vibration signals of the health bearing and the vibration signal of the fault bearing are different greatly, the extraction of the vibration signal level of Case Western Reserve University is high, or the feature extraction method (MDE) is very superior, which leads to all kinds of algorithms on the healthy bearing reach the classification accuracy of 100%. Among the three fault types, NKH-KELM is superior to the other seven algorithms, which proves that NKH-KELM has certain advantages of fault diagnosis based on machine learning. In addition, in the fault diagnosis of the ball, the six different algorithms used for comparison have not more than 95%, among which NB has only 49%. It can also be seen from Figure 15 that the six algorithms overlap greatly in the classification of ball faults and outer ring faults. In other words, the ability of these six algorithms

to distinguish the ball fault and outer ring fault is not high. However, the ball fault diagnosis accuracy of NKH-KELM reaches 97.1%, and the outer fault diagnosis accuracy reaches 98.4%, which indicates that NKH-KELM has certain advantages over other algorithms in distinguishing ball faults and outer ring faults.

#### 4.2. Fault Diagnosis Experiment Based on XJTU-SY Bearing Data Set

**4.2.1. Description of Data.** As shown in Figure 16, the bearing test is composed of an AC motor, a motor speed controller, a supporting shaft, a hydraulic loading system, etc. [72]. The radial force is generated by the hydraulic loading system and applied to the bearing housing under the test. The speed is set and maintained by the AC induction motor speed controller. In this experiment, the bearing model is LDK UER204, and its specifications are shown in Table 5.

**4.2.2. Fault Bearing Classification in XJTU-SY Bearing Data Set Based on NKH-KELM.** In order to test the generalization of the method proposed in this paper, the XJTU-SY bearing data set with the sampling frequency of 37.5 Hz was selected. The last 5000 points in the first 108 files of Bearing2\_1 (inner race fault), Bearing2\_2 (outer race fault), and Bearing2\_3 (cage fault) are selected as failure samples, and the first 5000 points in the first 36 samples of Bearing3\_4 are taken as healthy samples. In other words, the input of this experiment

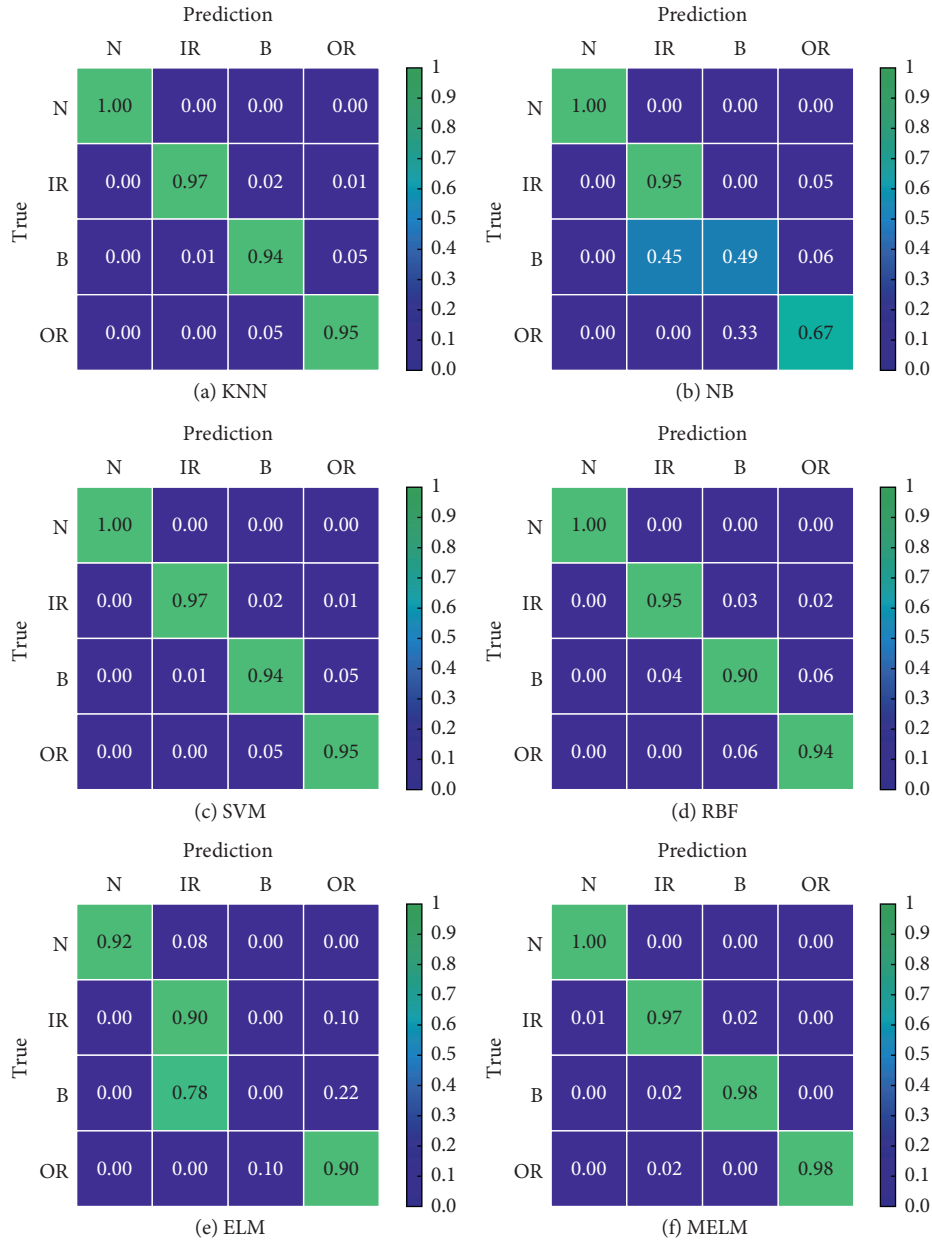


FIGURE 15: Bearing fault diagnosis of six different algorithms.

is 360 samples, including 36 healthy samples and 324 failure samples. The parameter selection of feature extraction is the same as that of the previous experiment. Then, the feature space after feature extraction is a matrix of  $360 \times 20$ . The ratio of the training sample set and the test sample set is 3 to 1. In other words, the feature space of the training sample set is  $270 \times 20$ , and that of the test sample set is  $90 \times 20$ . Then, the training sample set was imported into NKH-KELM with an impulse coefficient of 0.6 to adjust the model for training. Finally, the test sample set was imported into the trained model to test the performance of the model. The test results are shown in Figure 17.

In Figure 17, the abscissa is the number of sample sets, the ordinate is the state of the sample, the circle represents

the actual label of the sample, and the asterisk represents the prediction label of the sample. N stands for the healthy bearing, IR for the inner race fault, C for the cage fault, and OR for the outer race fault. In this experiment, the fault classification accuracy reaches 95.5556%. The fault classification of the healthy bearing reaches 100%. And the fault diagnosis of the three kinds of fault bearings mainly occurs in the diagnosis of the cage. For example, the fault of the inner race or the outer race is misdiagnosed as the fault of the cage, or the fault of the cage is misdiagnosed as the fault of the inner race or the fault of the outer race. This indicates that the characteristics of the three faults are similar to some extent. However, on the whole, the method proposed in this paper still has certain advantages in fault diagnosis on the

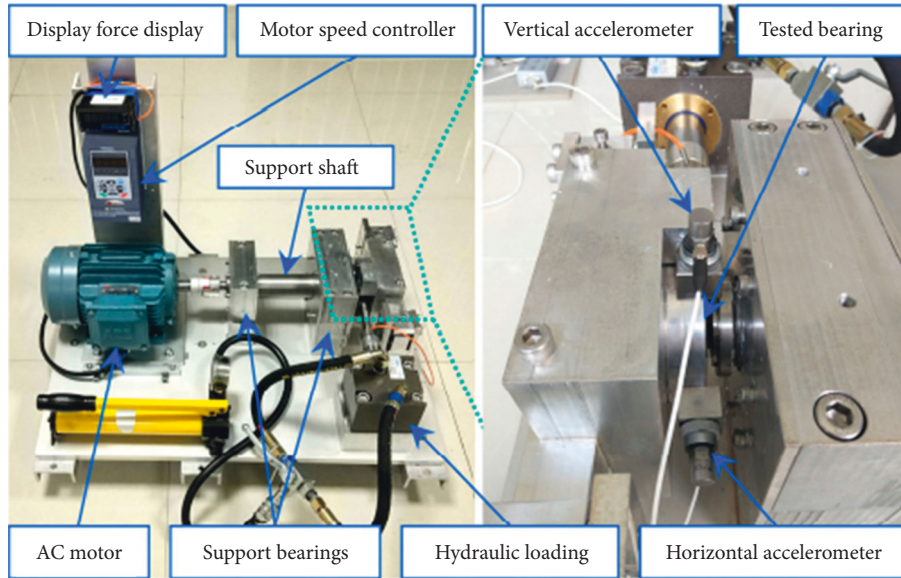


FIGURE 16: XJTU-SY fault diagnosis laboratory.

TABLE 5: Parameters of testing bearing.

Parameter	Value	Parameter	Value
Outer race diameter	39.80 mm	Inner race diameter	29.30 mm
Bearing mean diameter	34.55 mm	Ball diameter	7.92 mm
Number of balls	8	Contact angle	0°
Load rating (static)	6.65 kN	Load rating (dynamic)	12.82 kN

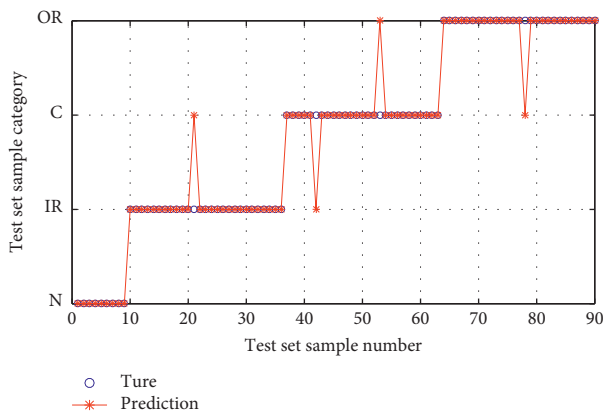


FIGURE 17: Fault diagnosis based on the XJTU-SY bearing data set: results of NKH-KELM with classification accuracy = 95.5556%.

XJTU-SY bearing data set, which also proves the generalization of the method proposed in this paper.

## 5. Conclusion and Future Work

This paper takes the bearing fault diagnosis of the gearbox as the research object. Firstly, the vibration signals are extracted by multiscale dispersion entropy (MDE). Secondly, they are imported into the kernel extreme learning machine (KELM) for fault classification. A novel krill herd algorithm (NKH) with opposite-based learning (OBL) and impulse operators is used to optimize kernel function

parameters  $\sigma$  and the error penalty factor  $C$  in the KELM. Then, the bearing data from Case Western Reserve University are imported into NKH-KELM for training and testing. Finally, the method proposed in this paper is compared with the other seven methods. It can be seen from the experimental results that NKH-KELM is superior to the other seven methods in the diagnosis of three fault types, which also proves the superiority of NKH-KELM in intelligent fault diagnosis of bearings. By comparing NKH-KELM with standard KH + KELM, it can be seen that the optimization ability of the novel krill herd algorithm is better than that of the unimproved krill herd algorithm. However, in the diagnosis of ball faults, although NKH-KELM is better than the other seven algorithms, the diagnostic accuracy of ball faults is significantly lower than that of the other two fault types. This phenomenon may be related to the low signal-to-noise ratio in the acquisition of the vibration signal of the ball, the small proportion of useful information on the ball extracted by the feature extraction method (MDE), or the poor performance of machine learning in the fault diagnosis of the ball. Therefore, this question will be an important research direction in the future. In addition, in the era of big data, deep learning methods are becoming more and more perfect, so not only is there a great demand for intelligent fault diagnosis based on machine learning but also there will be a great demand for intelligent fault diagnosis based on deep learning, which will also be a key research direction.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant no. 51905496, in part by the Shanxi Provincial Natural Science Foundation of China under Grant nos. 201801D221237, 201801D121186, and 201801D121187, and in part by the Science Foundation of the North University of China under Grant no. XJJ201802.

## References

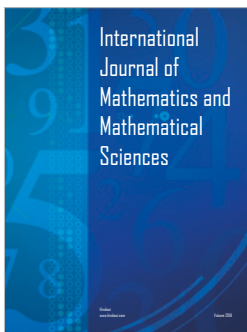
- [1] C. Shen, J. Yang, J. Tang, J. Liu, and H. Cao, "Note: parallel processing algorithm of temperature and noise error for micro-electro-mechanical system gyroscope based on variational mode decomposition and augmented nonlinear differentiator," *Review of Scientific Instruments*, vol. 89, no. 7, Article ID 076107, 2018.
- [2] X. Guo, J. Tang, J. Li, C. Shen, and J. Liu, "Attitude measurement based on imaging ray tracking model and orthographic projection with iteration algorithm," *ISA Transactions*, 2019.
- [3] X. Guo, J. Tang, J. Li, C. Wang, C. Shen, and J. Liu, "Determine turntable coordinate system considering its non-orthogonality," *Review of Scientific Instruments*, vol. 90, no. 3, Article ID 033704, 2019.
- [4] R. Wang, Y. Bai, Z. Zeng, J. Wang, W. Du, and N. Duan, "Opto-electronic platform tracking control of multi-rotor unmanned aerial vehicles based on composite disturbance compensation," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 223, no. 12, pp. 4410–4420, 2019, In press.
- [5] R. Wang, C. Zhao, Y. Bai, W. Du, and J. Wang, "An actuator fault detection and reconstruction scheme for hex-rotor unmanned aerial vehicle," *IEEE Access*, vol. 7, pp. 93937–93951, 2019.
- [6] L. Guo, Y. Huang, H. Gao, and L. Zhang, "Ball screw fault detection and location based on outlier and instantaneous rotational frequency estimation," *Shock and Vibration*, vol. 2019, Article ID 7497363, 12 pages, 2019.
- [7] H. Zhang, Z. Mo, J. Wang, and Q. Miao, "Nonlinear-drifted fractional Brownian motion with multiple hidden state variables for remaining useful life prediction of lithium-ion batteries," *IEEE Transactions on Reliability*, 2019.
- [8] F. Yan, J. Chen, X. Qiu et al., "IoT-IEFR: internet of things based intelligent emergency fire rescue algorithm for WebVR," *IEEE Access*, 2019, In press.
- [9] F. Yan, Y. Hu, J. Jia, Q. Guo, H. Zhu, and Z. Pan, "RFES: a real-time fire evacuation system for mobile Web3D," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 8, pp. 1061–1074, 2019.
- [10] F. Yan, J. Jia, Y. Hu, Q. Guo, and H. Zhu, "Smart fire evacuation service based on internet of things computing for Web3D," *Journal of Internet Technology*, vol. 20, no. 2, pp. 521–532, 2019.
- [11] J. Liu, Z. Xu, L. Zhou, W. Yu, and Y. Shao, "A statistical feature investigation of the spalling propagation assessment for a ball bearing," *Mechanism and Machine Theory*, vol. 131, pp. 336–350, 2019.
- [12] Z. Wang, G. He, W. Du et al., "Application of parameter optimized variational mode decomposition method in fault diagnosis of gearbox," *IEEE Access*, vol. 7, pp. 44871–44882, 2019.
- [13] Z. Wang, J. Wang, W. Cai et al., "Application of an improved ensemble local mean decomposition method for gearbox composite fault diagnosis," *Complexity*, vol. 2019, Article ID 1564243, 17 pages, 2019.
- [14] Z. Wang, J. Zhou, J. Wang et al., "A novel fault diagnosis method of gearbox based on maximum kurtosis spectral entropy deconvolution," *IEEE Access*, vol. 7, pp. 29520–29532, 2019.
- [15] H. Liu and J. Xiang, "Kernel regression residual signal based improved intrinsic time-scale decomposition for mechanical fault detection," *Measurement Science and Technology*, vol. 30, no. 1, Article ID 015107, 2019.
- [16] B. Pang, G. Tang, and T. Tian, "Enhanced singular spectrum decomposition and its application to rolling bearing fault diagnosis," *IEEE Access*, vol. 7, pp. 87769–87782, 2019.
- [17] Z. Mo, J. Wang, H. Zhang, and Q. Miao, "Weighted cyclic harmonic-to-noise ratio for rolling element bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, 2019.
- [18] X. Ding, Q. He, Y. Shao, and W. Huang, "Transient feature extraction based on time-frequency manifold image synthesis for machinery fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, 2019.
- [19] Z. Wang, W. Du, J. Wang et al., "Research and application of improved adaptive MOMEDA fault diagnosis method," *Measurement*, vol. 140, pp. 63–75, 2019.
- [20] J. Liu and Y. Shao, "An improved analytical model for a lubricated roller bearing including a localized defect with different edge shapes," *Journal of Vibration and Control*, vol. 24, no. 17, pp. 3894–3907, 2018.
- [21] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Computers in Industry*, vol. 105, pp. 182–190, 2019.
- [22] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [23] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017.
- [24] W. Gao, R.-J. Wai, S.-P. Qiao, and M.-F. Guo, "Mechanical faults diagnosis of high-voltage circuit breaker via hybrid features and integrated extreme learning machine," *IEEE Access*, vol. 7, no. 1, pp. 60091–60103, 2019.
- [25] Z. Wang, L. Zheng, and W. Du, "A novel method for intelligent fault diagnosis of bearing based on capsule neural network," *Complexity*, vol. 2019, Article ID 6943234, 10 pages, 2019.
- [26] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.



- [27] Y. Li, Y. Yang, X. Wang, B. Liu, and X. Liang, "Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine," *Journal of Sound and Vibration*, vol. 428, pp. 72–86, 2018.
- [28] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mechanical Systems and Signal Processing*, vol. 91, pp. 295–312, 2017.
- [29] Y. Cheng, C. Yin, L. Bai, Q. Bai, X. Huang, and K. Chen, "Fault diagnostics of rolling bearings using feature fusion based BP, RBF and PNN neural networks," *International Journal of Applied Electromagnetics and Mechanics*, vol. 52, no. 1-2, pp. 95–102, 2016.
- [30] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, 2019.
- [31] H. Shao, H. Jiang, X. Li, and S. Wu, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowledge-Based Systems*, vol. 140, pp. 1–14, 2018.
- [32] L. Zheng, Z. Wang, Z. Zhao, J. Wang, and W. Du, "Research of bearing fault diagnosis method based on multi-layer extreme learning machine optimized by novel ant lion algorithm," *IEEE Access*, vol. 7, pp. 89845–89856, 2019.
- [33] C. Yin, T. Xue, X. Huang, Y.-H. Cheng, S. Dadras, and S. Dadras, "Research on damages evaluation method with multi-objective feature extraction optimization scheme for M/OD impact risk assessment," *IEEE Access*, vol. 7, pp. 98530–98545, 2019, In press.
- [34] X. Chou, J. Zhu, S. Qian et al., "All-in-one filler-elastomer-based high-performance stretchable piezoelectric nanogenerator for kinetic energy harvesting and self-powered motion monitoring," *Nano Energy*, vol. 53, pp. 550–558, 2018.
- [35] J. He, T. Wen, S. Qian et al., "Triboelectric-piezoelectric-electromagnetic hybrid nanogenerator for high-efficient vibration energy harvesting and self-powered wireless monitoring system," *Nano Energy*, vol. 43, pp. 326–339, 2018.
- [36] H. Cao, Y. Zhang, C. Shen, Y. Liu, and X. Wang, "Temperature energy influence compensation for MEMS vibration gyroscope based on RBF NN-GA-KF method," *Shock and Vibration*, vol. 2018, Article ID 2830686, 10 pages, 2018.
- [37] H. Cao, Y. Zhang, Z. Han et al., "Pole-zero temperature compensation circuit design and experiment for dual-mass MEMS gyroscope bandwidth expansion," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 2, pp. 677–688, 2019.
- [38] C. Shen, X. Liu, H. Cao, and Y. Zhou, "Brain-like navigation scheme based on MEMS-INS and place recognition," *Applied Sciences*, vol. 9, no. 8, p. 1708, 2019.
- [39] S. Lu, Q. He, and J. Wang, "A review of stochastic resonance in rotating machine fault detection," *Mechanical Systems and Signal Processing*, vol. 116, pp. 230–260, 2019.
- [40] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: a review," *IEEE Access*, vol. 6, pp. 66723–66741, 2018.
- [41] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy based fault classification using the Case Western Reserve University data: a benchmark study," *IEEE Transactions on Reliability*, 2019.
- [42] F. Yang, Z. Kou, J. Wu, and T. Li, "Application of mutual information-sample entropy based MED-ICEEMDAN Denoising scheme for weak fault diagnosis of hoist bearing," *Entropy*, vol. 20, no. 9, p. 667, 2018.
- [43] X. Xue, C. Li, S. Cao, J. Sun, and L. Liu, "Fault diagnosis of rolling element bearings with a two-step scheme based on permutation entropy and random forests," *Entropy*, vol. 21, no. 1, p. 96, 2019.
- [44] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mechanical Systems and Signal Processing*, vol. 85, pp. 746–759, 2017.
- [45] Y. Li, M. Xu, Y. Wei, and W. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80–94, 2016.
- [46] M. Rostaghi and H. Azami, "Dispersion entropy: a measure for time-series analysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016.
- [47] H. Azami, E. Kinney-Lang, A. Ebied, A. Fernandez, and J. Escudero, "Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in Alzheimer's disease," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3182–3185, Jeju Island, Korea, July 2017.
- [48] Z. Zhao, Q. Xu, and M. Jia, "Improved shuffled frog leaping algorithm-based BP neural network and its application in bearing early fault diagnosis," *Neural Computing and Applications*, vol. 27, no. 2, pp. 375–385, 2016.
- [49] M. Liang, D. Su, D. Hu, and M. Ge, "A novel faults diagnosis method for rolling element bearings based on ELCD and extreme learning machine," *Shock and Vibration*, vol. 2018, Article ID 1891453, 85 pages, 2018.
- [50] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, pp. 35042–35056, 2018.
- [51] Z. Yang, Z. Guo, C. Liu, M. Jiang, and X. Wang, "Fast multi-label SVM training based on approximate extreme points," *Intelligent Data Analysis*, vol. 22, no. 5, pp. 1079–1099, 2018.
- [52] W. Zou, F. Yao, B. Zhang, and Z. Guan, "Improved meta-ELM with error feedback incremental ELM as hidden nodes," *Neural Computing and Applications*, vol. 30, no. 11, pp. 3363–3370, 2018.
- [53] C. He, F. Xu, Y. Liu, and J. Zheng, "A fast kernel extreme learning machine based on conjugate gradient," *Network-Computation in Neural System*, vol. 29, no. 1–4, pp. 70–80, 2018.
- [54] Y. Cheng, Z. Wang, W. Zhang, and G. Huang, "Particle swarm optimization algorithm to solve the deconvolution problem for rolling element bearing fault diagnosis," *ISA Transactions*, vol. 90, pp. 244–267, 2019.
- [55] Y. Huang, F. He, and Y. Qiu, "Dynamic strategy based parallel ant colony optimization on GPUs for TSPs," *Science China-Information Sciences*, vol. 60, no. 6, Article ID 068102, 2017.
- [56] A. H. Gandomi and A. H. Alavi, "Krill herd: a new bio-inspired optimization algorithm," *Communications in Non-linear Science and Numerical Simulation*, vol. 17, no. 12, pp. 4832–4845, 2012.
- [57] C. Yin, S. Wu, S. Zhou, J. Cao, X. Huang, and Y. Cheng, "Design and stability analysis of multivariate extremum seeking with Newton method," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1559–1578, 2018.
- [58] C. Yin, S. Dadras, X. Huang, Y. Cheng, and H. Malek, "The design and performance analysis of multivariate fractional-



- order gradient-based extremum seeking approach,” *Applied Mathematical Modelling*, vol. 62, pp. 680–700, 2018.
- [59] L. M. Zheng, Q. Wang, S. X. Zhang, and S. Y. Zheng, “Population recombination strategies for multi-objective particle swarm optimization,” *Soft Computing*, vol. 21, no. 16, pp. 4693–4705, 2017.
- [60] H. R. Tizhoosh, “Opposition-based learning: a new scheme for machine intelligence,” in *Proceedings of the International Conference on Computational Intelligence for Modelling, Control & Automation, & International Conference on Intelligent Agents, Web Technologies & Internet Commerce*, pp. 695–701, IEEE, Vienna, Austria, November 2005.
- [61] Y. Wang, P. W. Tse, B. Tang et al., “Order spectrogram visualization for rolling bearing fault detection under speed variation conditions,” *Mechanical Systems and Signal Processing*, vol. 122, pp. 580–596, 2019.
- [62] B. Zhao, S. Zhang, X. Lu, and Q. Dong, “Cyclic tangential loading of a power-law hardening elastic-plastic spherical contact in pre-sliding stage,” *International Journal of Mechanical Sciences*, vol. 128–129, pp. 652–658, 2017.
- [63] N. Wang, C. Liu, D. Jiang, and K. Behdinan, “Casing vibration response prediction of dual-rotor-blade-casing system with blade-casing rubbing,” *Mechanical Systems and Signal Processing*, vol. 118, pp. 61–77, 2019.
- [64] Y. Behdinan, D. Huang, and J. Yan, “A novel extreme learning machine classification model for e-nose application based on the multiple kernel approach,” *Sensors*, vol. 17, no. 6, p. 1434, 2017.
- [65] H. A. Atee, R. Ahmad, N. M. Noor, A. M. S. Rahma, and Y. Aljeroudi, “Extreme learning machine based optimal embedding location finder for image steganography,” *PLoS One*, vol. 12, no. 2, Article ID e0170329, 2017.
- [66] J.-H. Kim, “Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap,” *Computational Statistics & Data Analysis*, vol. 53, no. 11, pp. 3735–3745, 2009.
- [67] X. Yan and M. Jia, “Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection,” *Knowledge-Based Systems*, vol. 163, pp. 450–471, 2019.
- [68] S. R. Das, Kuhoo, D. Mishra, and M. Rout, “An optimized feature reduction based currency forecasting model exploring the online sequential extreme learning machine and krill herd strategies,” *Physica A: Statistical Mechanics and Its Applications*, vol. 513, pp. 339–370, 2019.
- [69] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, “Efficient kNN classification with different numbers of nearest neighbors,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, 2018.
- [70] T. Li, J. Li, Z. Liu, P. Li, and C. Jia, “Differentially private Naive Bayes learning over multiple data sources,” *Information Sciences*, vol. 444, pp. 89–104, 2018.
- [71] Y. Liu, J.-W. Bi, and Z.-P. Fan, “Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms,” *Expert Systems with Applications*, vol. 80, pp. 323–339, 2017.
- [72] B. Wang, Y. Lei, N. Li, and N. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings,” *IEEE Transactions on Reliability*, 2018.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

