An intelligent fault identification method of rolling bearings based on SVM optimized by improved GWO

Haisong Huang, Qingsong Fan, Jianan Wei and Dong Huang

Key Laboratory of Advanced Manufacturing Technology, Ministry of Education, Guizhou University, Guiyang, People’s Republic of China

ABSTRACT

To improve the accuracy and the recognition efficiency of a bearing fault diagnosis, a fault diagnosis method based upon the improved grey wolf optimization (IGWO) algorithm and support vector machine (SVM) is proposed in the following manners. First, the data are pre-processed by using the set ensemble empirical mode decomposition (EEMD), Shannon wavelet packet entropy (SWPE), and principal component analysis (PCA). Next, the idea of updating the host bird nest by the cuckoo search (CS) optimization algorithm is introduced into the grey wolf optimization (GWO) algorithm to obtain the IGWO algorithm. Then, the SVM is optimized by the IGWO algorithm to obtain optimal parameters for a new diagnostic model. This model improves the problem where the algorithm easily falls into a local optimum. The learning ability and the generalization ability of the SVM are also enhanced. Finally, the effectiveness of the optimization model is tested by two different bearing data sets. The results show that compared to the genetic algorithm (GA), particle swarm optimization (PSO) and GWO algorithm optimization, the IGWO algorithm can be more accurate and efficient when diagnosing bearings.

ARTICLE HISTORY

Received 1 May 2019
Accepted 29 July 2019

KEYWORDS

Improved grey wolf optimization algorithm; support vector machine; parameter optimization; bearing fault diagnosis

1. Introduction

The rolling bearing is a key component in the operation of mechanical equipment. If this part fails, normal operation of other components will be negatively influenced, and the whole system may be paralyzed. In addition, vibration signals carry a large amount of information that represents the health of mechanical equipment, so vibration signals are widely used in the fields of condition monitoring and the diagnosis of rotating machinery (Lei, He, & Zi, 2011; Žvokelj, Zupan, & Prebil, 2011). The ultimate goal of bearing fault diagnosis is to establish an effective, reliable and fast vibration signal identification system. The performance of this identification system depends on the extraction of fault signal characteristics and the ability of the classifier to correctly distinguish faults (William & Hoffman, 2011).

In recent years, many scholars have performed research that puts forward various effective fault diagnosis methods based on the following: neural networks (Schlechttingen & Santos, 2011), expert systems (Liu & Liu, 2003), clustering algorithms (West, McArthur, & Towl, 2012), support vector machine (SVMs) (Konar & Chattopadhyay, 2011), D-S Evidence theory (Chen, Whitbrook, Aickelin, & Roadknight, 2014), and big data solutions (Wan et al. 2017). Among these methods, the SVM has the best generalization ability compared to any other method. SVM is also more applicable to small-sample, nonlinear and high-dimensional pattern recognition. Therefore, the SVM is considered a promising classification technology and has been widely used in mechanical fault diagnosis (Chen, Tang, Tao, & Li, 2014; Gao, Cecati, & Ding, 2015). In the fault diagnosis of SVM, the diagnostic accuracy has a great relationship with the kernel function parameter $\sigma$ and the penalty coefficient $c$, so the optimization of its parameters is particularly important (Zhang, Liang, Zhou, & Yi, 2015).

In many parameter optimization methods, the most common technique for SVM parameter selection is the grid algorithm. However, the grid algorithm is highly time-consuming and has poor performance in most cases (Fei & Zhang, 2009). Hence, scholars have proposed various intelligent algorithms to deal with this situation. For instance, Jack et al. used the genetic algorithm (GA) to optimize SVM parameters and used the optimized SVM for state detection of rolling bearings. Although he achieved good general performance, the GA easily fell into a local optimum (Jack & Nandi, 2002). Huang et al. suggested a particle swarm optimization (PSO) to
improve the SVM model. This method can select parameters reasonably and achieve higher classification accuracy for the SVM, but the PSO algorithm has a poor local search ability (Huang & Dun, 2008). However, the GWO is advantageous in multiple ways. The GWO has many features such as simple structure, few parameters to be adjusted, and easy implementation. It also has a convergence factor that can be adaptively adjusted and an information feedback mechanism to ensure that it can achieve a balance between local optimization and global search. Therefore, it has good performance in solving the problem and the convergence speed. In terms of function optimization, it has been proved in previous literature that the partial optimization results of the GWO are better than both the GA and the PSO (Mirjalili, Mirjalili, & Lewis, 2014). Therefore, the GWO algorithm is widely used in unmanned combat aircraft path planning (Zhang, Zhou, Li, & Pan, 2016), optimal reactive power dispatch (Sulaiman, Mustaffa, Mohamed, & Aliman, 2015), optimal tuning of PID-fuzzy controllers (Noshadi, Shi, Lee, Shi, & Kalam, 2016), SVM parameter optimization (Bian, Zhang, Du, Chen, & Zhang, 2018) and many other fields. However, the GWO algorithm search mechanism leads to the disadvantage that it has a slow convergence rate and easily falls into a local optimum (Chun Yang & Long, 2016; Saremi, Mirjalili, & Mirjalili, 2015; Zhang, Kang, Cheng, & Wang, 2018).

To improve the shortcomings of the local search ability of the GWO algorithm, this paper proposes an IGWO algorithm, which combines the GWO algorithm with the CS algorithm. The CS algorithm will decide the optimal position of a group, and enhance the global search ability of that group, so that individuals can better guide others for decision-making and test the feasibility using the benchmark function. The IGWO algorithm will be combined with SVM and extended to the bearing fault diagnosis problem to verify the reference value.

The bearing data are pre-processed by the ensemble empirical mode decomposition (EEMD), Shannon wavelet packet entropy (SWPE), and principal component analysis (PCA). The SVM parameters $\sigma$ and $c$ are improved by the IGWO algorithm, and the test set classification result is obtained by the SVM ten-fold cross-validation. The steps are as presented in Figure 1.

The subsequent structure of this paper is as follows. Section 2 will provide the theoretical basis for data pre-processing. Sections 3 and 4 will discuss detailed processes of the IGWO-SVM model. Section 5 will verify the feasibility of the proposed method through two different bearing data-sets. Section 6 presents the conclusion based on the corresponding results.

### 2. Feature extraction

#### 2.1. Ensemble empirical mode decomposition (EEMD)

Due to the non-Gaussian and non-stationary characteristics of the bearing vibration signal, the traditional Fourier transform-based noise reduction method has a contradiction between the protection signal edge and the suppression noise, and it is difficult to correctly identify and remove the noise in the signal. The wavelet transform based noise reduction method needs to select the appropriate mother wavelet and set the feasible decomposition layer number. There are problems such as parameter sensitivity and stationarity assumption, which have certain limitations for non-stationary signal processing (Li, Qu, & Liao, 2007). The empirical mode decomposition (EMD) proposed by Huang et al. does not need to select the basis function and is adaptive (Huang et al., 1998). The EMD is suitable for the processing of nonlinear non-stationary signals, but the traditional EMD has a mode mixing problem, which will cause the physical meaning of each Intrinsic Mode Function (IMF) to be unclear and
affect the effect of subsequent feature extraction (Lei, Lin, He, & Zuo, 2013). To reduce the shortcomings of mode mixing, Zhao et al. proposed an effective noise-assisted method EEMD (Zhaohua & Huang, 2009), which can significantly reduce the possibility of excessive mode mixing and preserve the binary nature of any data decomposition. EEMD introduces Gaussian white noise to suppress the mode aliasing problem, which makes the signal continuous at different scales, thus changing the extreme point distribution of the signal to achieve the purpose of suppressing mode aliasing. EEMD is a very mature non-linear tool that has been widely used in a variety of fault diagnosis model signals (Liu, Cui, & Li, 2017; Zhang & Zhou, 2013). Therefore, this study uses EEMD to adaptively decompose the signal to highlight the fault characteristics of each frequency band. The specific calculation steps are as follows.

Step 1: Determine the amount of $N$ and the magnitude of the white noise generated by the increased value $a$.

Step 2: The white noise $n_i(t)$ produced by the value $a$ of the given amplitude is superimposed on the original signal $x(t)$ to generate a new signal:

$$x_i(t) = x(t) + n_i(t)$$  \hspace{1cm} (1)

where $n_i(t)$ represents the $i$th added white noise sequence, $x_i(t)$ is the new signal obtained after the $i$th white noise is superimposed, and $i = 1, 2, \ldots, N$.

Step 3: The newly generated signal $x_i(t)$ is decomposed into IMF using the original EMD algorithm.

$$x_i(t) = \sum_{s=1}^{s} C_{is}(t) + r_i(t)$$  \hspace{1cm} (2)

where $s$ is the number of IMF and $r_i(t)$ is the residual component, which is the average trend of the signal $[C_{i1}(t), C_{i2}(t), \ldots, C_{iS}(t)]$ representing IMF from high frequency to low frequency.

Step 4: According to the number of sets $N$ in step 1, step 2 and step 3 are repeated $N$ times to obtain a set of IMF.

$$[[C_{11}(t), C_{12}(t), \ldots, C_{1S}(t)], [C_{21}(t), C_{22}(t), \ldots, C_{2S}(t)], \ldots, [C_{N1}(t), C_{N2}(t), \ldots, C_{NS}(t)]]$$  \hspace{1cm} (3)

Step 5: Calculate the aggregate mean of the $N$ sets with IMF as the final result.

$$C_i(t) = \frac{1}{N} \sum_{j=1}^{N} C_{ij}(t)$$  \hspace{1cm} (4)

where $C_i(t)$ is the $s_i$th IMF decomposed by EEMD, $i = 1, 2, \ldots, N$ and $s = 1, 2, \ldots, S$.

### 2.2. Shannon wavelet packet entropy (SWPE)

Wavelet packet analysis is like discrete wavelet transforms, which are based on multiresolution analysis frameworks. The difference between them is mainly that the multi-resolution decomposition of wavelet transform only decomposes the scale space. However, the wavelet packet transform simultaneously decomposes the scale space and the wavelet space, so that the spectrum window is further divided and thinned. It has higher time–frequency resolution, and the wavelet packet has sub-bands of the same width in both the high and low-frequency bands. Therefore the wavelet packet is more suitable for processing non-stationary signals (Wu & Liu, 2009; Zhao et al., 2018).

The wavelet packet function can be defined recursively as follows:

$$u_{2n}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h(k)u_{n}(2t - k)$$

$$u_{2n+1}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} g(k)u_{n}(2t - k)$$  \hspace{1cm} (5)

where $k$ is the translation factor, $n = 0, 1, 2, \ldots, h(k)$ is a high pass filter bank, and $g(k)$ is a low pass filter bank. Shannon information entropy theory points out that for an uncertain system, if a random variable $X$ with a finite value is used to represent its state feature, the probability that $X$ takes the value of $x_j$ is $p_j = P(X = x_j) (j = 1, 2, \ldots, L)$; and if $\sum_{j=1}^{L} p_j = 1$, then the information obtained by a certain result of $X$ can be represented by $l_j = \log(1/p_j)$, so the information entropy of $X$ is defined as

$$H(X) = - \sum_{j=1}^{L} p_j \log p_j$$  \hspace{1cm} (6)

When $p_i = 0$, we can define $p_i \log p_i = 0$, and the information entropy $H$ is an information measure of the positioning system under certain conditions. It is a measure of the unknownness of the sequence and can be used to estimate the complexity of a random signal.

After the $j$-layer wavelet packet decomposition of the signal, the wavelet packet decomposition sequence is $S_{(jk)}(k = 0 \sim 2^j - 1)$. Here, the wavelet packet decomposition of the signal can be regarded as a kind of division of the signal. The measurement of this division is defined as

$$\varepsilon_{(jk)}(i) = \frac{S_{F(j,k)}(i)}{\sum_{i=1}^{N} S_{F(j,k)}(i)}$$  \hspace{1cm} (7)

where $S_{F(j,k)}(i)$ is the $i$th value of the Fourier transform sequence of $S_{(jk)}(k = 0 \sim 2^j - 1)$; $N$ is the original signal length.
According to the basic theory of information entropy, the wavelet packet characteristic entropy can be defined as
\[ H_{j,k} = - \sum_{i=1}^{N} \epsilon_{(j,k)}(i) \log \epsilon_{(j,k)}(i) \] (8)
where \( H_{j,k} \) is the characteristic entropy of the \( k \)th wavelet packet of the \( j \)th layer of the signal and \( k = 0 \sim 2^j - 1 \).

3. Support vector machine (SVM)

SVM is a classification method based on statistical learning theory proposed by Cortes and Vapnik (1995). The training data set is \((x_i, y_i), x_i \in \mathbb{R}^p, y_i \in \{-1, +1\}\) is the category label, and we use SVM to construct an objective function to find the optimal partitioning hyperplane:
\[ y = w^T \phi(x) + b \] (9)
where \( \omega \) represents a hyperplane normal vector and \( b \) represents an offset. If the linearity is inseparable, the following optimal equation needs to be solved:
\[
\min V(w, \xi) = \frac{1}{2} ||w||^2 + c \sum_{i=1}^{n} \xi_i \\
\text{s.t.} \quad \begin{cases} 
    y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i \\
    \xi_i \geq 0, i = 1, 2, \ldots, l
\end{cases}
\] (10)
where \( \xi_i \) is a linearly inseparable time introduced as a slack variable, and \( c \) is a penalty factor used to represent a penalty index for misclassification.

To minimize the objective function, when \( c \) is large, \( \xi_i \) can only approach as close as possible to 0. The sample tolerance between the boundaries is very low, the misclassification is less, the fitting of the sample is better, but the prediction effect is not good. When \( c \) is small, there are more samples between the two boundaries. The possibility of misclassification becomes larger, and the fit to the sample is reduced, but it may be more reasonable because there may be noise between the samples (Bao, Hu, & Xiong, 2013; Phan, Nguyen, & Bui, 2017). \( c \) is a key factor in determining the SVM learning ability and experience risk coordination. If \( c \) is too large, it will cause over-learning, which will reduce the generalization ability of the classifier. In contrast, it will lead to the classification accuracy of the classifier being too low, and even the entire classifier model will be invalid (Wang, Zhang, Song, Liu, & Dong, 2019). Then, the Lagrangian function is introduced to solve Equation (10), and the structural superplane of the optimal classification is transformed into a convex quadratic programming problem:
\[
\max Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (\phi(x_i^T) \cdot \phi(x_j)) \\
s.t. \quad \sum_{i=1}^{l} y_i \alpha_i = 0 \\
0 \leq \alpha_i \leq c, i = 1, 2, \ldots, l
\] (11)
where \( \alpha_i \) is the Lagrangian multiplier, \( c \) is the penalty factor, and \( \phi(x_i)^T \cdot \phi(x_j) \) is the kernel function, which can be represented by \( K(x_i, x_j) \), then the decision function is:
\[
f(x) = \text{sgn} \left( \sum_{i=1}^{l} \alpha_i^* y_i K(x_i, x) + b^* \right)
\] (12)
where \( \alpha_i^* \) and \( b^* \) are the parameters that determine the best classification hyperplane.

In SVM, the choice of kernel function has a large impact on SVM performance. The common kernels are polynomial kernels, radial basis function (RBF), sigmoid kernels, etc. Among these kernels, the RBF kernel function contains only the parameter \( \sigma \), which is simpler for \( \sigma \) optimization and is beneficial to parameter optimization. RBF is the most commonly used kernel function. Therefore, this paper chooses the RBF kernel function, which is defined as:
\[
K(x_i, x_j) = \exp \left( - \frac{||x_i - x_j||^2}{2\sigma^2} \right)
\] (13)
where \( \sigma \) is a kernel function parameter that affects the distribution of sample data complexity in the feature space.

The above analysis shows that reasonable selection of the penalty parameter \( c \) and the RBF kernel function parameter \( \sigma \) can effectively improve the accuracy of the SVM classifier.

4. SVM parameter optimization based on IGWO

This part provides concisely precise descriptions and interpretation for the experimental results and draws experimental conclusions.

4.1. Grey wolf optimization (GWO) algorithm

The GWO Algorithm is a population-based stochastic optimization algorithm proposed by Mirjalili et al. (2014). The algorithm is proposed by simulating the leadership hierarchy and predation behaviour of the natural grey wolf population. By testing 29 continuous function optimization problems, the results have shown that GWO is superior to PSO, GA and other algorithms in solving the accuracy and stability. GWO has been widely used in function optimization and other problems because of its lower complexity, fewer control parameters, higher search efficiency and higher efficiency.
the grey wolf predation behaviour is described as follows:

\[ D = |C \cdot X_p(t) - X(t)| \]  \hspace{1cm} (14)
\[ X(t + 1) = X_p(t) - A \cdot D \]  \hspace{1cm} (15)

where \( t \) represents the current number of iterations, \( A \) and \( C \) are coefficient vectors, \( X_p \) is the position vector of the prey, and \( X \) is the position vector of the grey wolf.

\[ A = 2a \cdot r_1 - a \]  \hspace{1cm} (16)
\[ C = 2 \cdot r_2 \]  \hspace{1cm} (17)

where \( r_1 \) and \( r_2 \) are both random vectors of \([0, 1]\) and \( a \) is a convergence factor, which decreases linearly from 2 to 0 with number of iterations.

The position of the remaining grey wolf \( \omega \) in the population is determined by the positions of \( \alpha \), \( \beta \) and \( \delta \):

\[ D_\alpha = |C_1 \cdot X_\alpha - X|, X_1 = X_\alpha - A_1 \cdot D_\alpha \]
\[ D_\beta = |C_2 \cdot X_\beta - X|, X_2 = X_\beta - A_2 \cdot D_\beta \]
\[ D_\delta = |C_3 \cdot X_\delta - X|, X_3 = X_\delta - A_3 \cdot D_\delta \]  \hspace{1cm} (18)
\[ X(t + 1) = (X_1 + X_2 + X_3)/3 \]  \hspace{1cm} (19)

4.2. Improved grey wolf optimization (IGWO) algorithm

The GWO algorithm selects an individual update position with a high fitness value by using Equations (18) and (19). However, this update method has weak global search capability and high probability of falling into a local optimum. When the CS algorithm updates the position of the host’s nest, it is mainly realized by the two operations of fostering behaviour and Levi’s flight, and the selection of the moving direction is highly random. In addition, the CS algorithm makes it easier to jump from the current area to others to complete the global search. Based on the above advantages of the CS algorithm, this paper introduces the CS algorithm update host nesting idea into the GWO algorithm to improve its algorithm performance and proposes the IGWO algorithm to improve the problem that the GWO algorithm easily falls into a local optimum.

The cuckoo search algorithm simulates the process of cuckoos wonderingly flying to find a nest suitable for spawning (Yang & Deb, 2010). During the search process, the location of \( N \) nests is first randomly tested in the feasible solution space. Then, the fitness of each bird’s nest position is calculated and the optimal fit of the bird’s nest retained for the next generation. The iterative process of the bird’s nest position is entered. The position \( X_i = (x_{i1}, x_{i2}, \ldots, x_{id}), 1 \leq i \leq n \) of the \( i \) bird’s nest is set in the \( d \)-dimensional search space. By evaluating the fitness of each bird’s nest, the optimal position
When the GWO algorithm obtains the new position of the wolves, the IGOW algorithm uses the CS algorithm to update the host nesting idea, and calculates the updated position of the wolf group twice with formula (20). That is, when GWO obtains the position vector $X_1, X_2, X_3$ of each generation of $\alpha, \beta$ and $\delta$ grey wolves, it does not directly enter the next generation but introduces the CS algorithm to continue the search. The position of the position vector $X_1, X_2, X_3$ is updated by the CS algorithm to obtain the position of the new $\alpha, \beta$ and $\delta$ grey wolves, calculate the position of the $\omega$ grey wolf, and then enter the next generation.

### 4.3. IGWO algorithm confirmatory analysis

To test the optimization performance of the proposed algorithm, the GA, PSO, GWO, and IGWO algorithms are selected and compared by four typical benchmark functions of Sphere, Griewank, Rosenbrock and Ackley. These benchmark functions are listed in Table 1 where dim indicates the dimension of the function, range is the boundary of the function’s search space, and the ideal minimum of all functions in the table is 0.

![Figure 3. The fitness convergence curves of benchmark functions.](image-url)

| Function name | Function | Dim | Range       |
|---------------|----------|-----|-------------|
| Sphere        | $F(x) = \sum_{i=1}^{n} x_i^2$ | 30  | $[-50,50]$  |
| Griewank      | $F(x) = \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$ | 30  | $[-600,600]$|
| Rosenbrock    | $F(x) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$ | 30  | $[-50,50]$  |
| Ackley        | $F(x) = -20 \exp\left(-0.2 \sum_{i=1}^{n} x_i^2\right) - \exp\left(\frac{1}{n} \sum_{i=1}^{n} \cos(2\pi x_i)\right) + 20 + e$ | 30  | $[-32,32]$  |
intelligent algorithms, the number of iterations $iter$ is sequentially selected as 100, 150, 200, and 250, and the population size $N$ is sequentially selected as 20, 30, 40, and 50. The test results of $iter = 100$ and $N = 20$ are shown in Figure 3.

Figure 3 shows that the accuracy and convergence rate of the proposed IGWO algorithm are better than the accuracy and convergence rates of GWO, PSO and GA. Since IGWO introduced the cuckoo algorithm, it improves the global search ability of the algorithm. The reference value of IGWO is verified, and it has good generalization ability and good robustness. The convergence results of other combined population parameters are the same as this trend and are unlisted in the space, thus determining the superiority of the IGWO algorithm in the optimization process.

4.4. IGWO optimizes the process of SVM fault diagnosis

This paper proposes to extract the sample data features with EEMD-SWPE-PCA. Through the ten-fold cross-validation, the IGWO algorithm is used to optimize the SVM penalty coefficient $c$ and the kernel parameter $\sigma$. The parameter selection steps are as follows.

Step 1: Initialize the IGWO algorithm and SVM-related parameters, including the maximum allowed several iterations, the population size, the penalty coefficient $c$, and the value range of the kernel parameter $\sigma$.

Step 2: SVM performs learning training based on the initial value of $c, \sigma$, calculates the fitness value of each grey wolf under the current $c, \sigma$ and selects the $\alpha, \beta, \delta$ wolf.

![Figure 4. Diagram of the procedure structure for the IGWO-SVM fault diagnosis model.](image-url)
Step 3: Update the current grey wolf position according to Equation (19).

Step 4: Update the IGWO algorithm parameter $A$, $a$, $C$.

Step 5: Calculate the fitness of all grey wolves and update alpha, beta, and delta according to Equations (18) and (20).

Step 6: Compare the current iteration number with the maximum number of iterations. If not, skip Step 3 to continue parameter optimization. Otherwise, end the training, and the $\alpha$ wolf position is the optimal solution, i.e., the optimal solution output of the SVM parameter $c_{best}$, $\sigma_{best}$.

Based on the methods mentioned above, the EEMD-SWPE-PCA-IGWO-SVM model is proposed for fault diagnosis. The flow chart of fault diagnosis model is shown in Figure 4.

5. Experimental verification

5.1. Bearing vibration data from CWRU

In this section, the rolling bearing data from the Electrical Engineering Laboratory of Case Western Reserve University (CWRU) are used to verify the proposed method (2019). The SVM parameters are optimized by different algorithms to emphasize the superiority of the proposed method.

5.1.1. Data description

The bearing vibration signal is collected as the analysis data sample under the speed of 1797 r/min, and the vibration data under the whole working condition are obtained with the sampling frequency of 12 kHz. The experiment uses electro-discharge machining (EDM) technology to arrange a single point of failure on the bearing. The fault width is 0.1778 mm, and the depth is 0.2794 mm. The vibration signals of the four states of normal, inner race fault, outer race fault, and ball fault are collected. The time domain and frequency domain information of the four fault states is shown in Figure 5.

5.1.2. CWRU data pre-processing

The EEMD extraction energy feature is performed on the collected four types of bearing data samples: normal, inner race fault, outer race fault, and ball fault. Since the degree of correlation between the first few modal components and the original signal and the contribution rate of the variance account for the main part, the main original signal characteristic information is also concentrated in the first few modal components. This paper selects the first three modal components after the EEMD decomposition. Taking the rolling bearing with normal and ball fault as an example, Figure 6 shows the first three components of the sampled signal EEMD. After the four types of bearing data samples are decomposed by EEMD, the first three sets of IMFs are respectively subjected to SWPE. The result is shown in Figure 7. Obviously, we can observe from the figure that the four types of samples have a separate trend, showing that the IMF-SWPE value of the bearing vibration signal can be used as the fault detection standard to detect bearing faults.

The PCA is used to reduce dimensionality to better construct and select the characteristic parameters of bearing fault nature. Moreover, the new feature set after PCA dimension reduction is divided into 200/200/200/200 training set and test set samples. These sample points are respectively projected onto a two-dimensional plane, and the effect diagram is as shown in Figure 8.

Figure 5. Time domain and frequency domain diagram of four states of bearings.
Figure 6. EEMD results for normal and ball fault.

Figure 7. EEMD-SWPE results for four types of bearing data samples.
5.1.3. IGWO parameter selection

In this paper, the pre-processed normal, inner race fault, outer race fault, and ball fault types of bearing data samples are separately numbered 0, 1, 2, and 3. The GA, PSO, GWO algorithm, and IGWO algorithm are used to optimize the parameters of the SVM classifier. Each algorithm was tested 100 times. In addition, the initial population $N$ of GA/PSO/GWO/IGWO is the same as the number of iterations $t$, which is 20/100.

In the test, the average accuracy, the highest accuracy, the lowest accuracy, the average running time and the longest and shortest running time difference of the test set were counted as the evaluation criteria to evaluate the performance of the SVM classification model. The classification results of the four algorithms after tenfold cross-validation are shown in Table 2. For the CWRU dataset, the SVM parameter optimization ability of the GA is higher than the SVM parameter optimization ability of the PSO, but the time required is longer, and the optimization time is extremely short due to the unstable optimization period. Although the optimization time of the PSO algorithm is the shortest, the convergence rate of the test set is lowest, and the effect is worst because of the slowest convergence. The GWO algorithm compares the GA and PSO, because the GWO algorithm optimization mechanism is simpler with fewer parameters to be determined. Each result shows better performance, which reflects the superiority of the GWO algorithm. However, the GWO algorithm easily falls into a local optimum, which also produces the defect that the optimization time is very short, and the IGWO algorithm overcomes the shortcoming. Compared with other algorithms, the IGWO algorithm optimization diagnostic model has higher recognition accuracy and higher efficiency, which verifies the feasibility and superiority of the proposed bearing fault diagnosis method. Figure 9 shows the test set recognition results of the SVM classifier with IGWO optimization parameters. Table 2 and Figure 9, show that IGWO-SVM can identify these vibration data with high precision, high efficiency and good diagnostic ability.

5.2. Aerospace bearing test data from IMS

To further verify the algorithm, the classification performance of the proposed model is verified with the second verification of the aerospace bearing fatigue life test data from the University of Cincinnati Intelligent Maintenance Center (IMS) (2019). Various comparisons are also researched to demonstrate the effectiveness and superiority of the proposed method.

5.2.1. Data description

The IMS Aerospace Bearing Fatigue Life Tester consists of a main part, a transmission part, a loading system, a lubrication system and a control circuit. Four RexnordZA-2115 double row bearings are mounted on the test shaft (2019). The AC motor drives the experimental shaft with a test speed of 2000 r/min and radial load of 26.67 kN. The vibration signal is monitored by a high-sensitivity Integrated Circuits piezoelectric (ICP) sensor 353B33 mounted on a Printed Circuit Board (PCB) mounted on the housing, and four thermocouple temperature sensors are used to monitor the bearing temperature and lubrication. The vibration signal was recorded by NI’s DAQCard-6063E digital collector with a sampling frequency 12 kHz.

Table 2. Fault recognition results of four algorithms to optimize the SVM.

| Method | Average test accuracy (%) | Maximum test accuracy (%) | Minimum test accuracy (%) | Average time(s) | $\Delta T$ (s) |
|--------|---------------------------|---------------------------|---------------------------|----------------|--------------|
| GA     | 83.391                    | 99.5                      | 54.375                    | 232.407        | 256.875      |
| PSO    | 73.599                    | 74.125                    | 73.500                    | 478.792        | 42.497       |
| GWO    | 99.771                    | 99.875                    | 99.625                    | 85.245         | 61.492       |
| IGWO   | 99.778                    | 99.875                    | 99.625                    | 66.823         | 52.819       |

$\Delta T =$ Maximum time − Minimum time.
The vibration intensity is defined as the root mean square \textit{RMS} value of the vibration speed, ranging from 10 to 1000 Hz, which is an effective characteristic reflecting the operating state of the equipment. The vibration intensity includes the main parameters and characteristics of the vibration signal. In the mechanical fault diagnosis system, it reflects the overall operating state of the mechanical equipment, so this paper uses the root mean square value $X_{\text{rms}}$ to represent the abnormal situation of bearing vibration:

$$
X_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x(n))}
$$

The inner race fault data of bearing 3 of test 1 are taken as an analysis example. Figure 10 shows the trend of the root mean square value of the bearing during its whole life. The figure shows that for the inner race fault of
bearing 3, the change in root mean square can be divided into two stages. In the first phase, no potential trends were observed during the first 30 days of operation. After 30 days of testing (approximately 86.4 million cycles), \(RMS\) began to increase significantly (Qiu, Lee, Lin, & Yu, 2006).

The normal sample in the picture 38880–41760 min is selected and numbered as ‘−1’, and the 46080–48960 min fault class samples are numbered as ‘+1’. Each type of the 204800 sets of data was randomly selected for data pre-processing. The time domain and frequency domain information for the normal and fault state of the bearing

![Figure 11](image1.png)

**Figure 11.** Time domain and frequency domain diagram of two states of bearings.

![Figure 12](image2.png)

**Figure 12.** EEMD results for normal and ball fault.
is shown in Figure 11. Like the previous article, Figure 12 shows the first three groups of IMFs after two types of samples processed by EEMD. Using the proposed data pre-processing method, the distribution of sample points after seeking SWPE for the IMF can be seen through Figure 13.

Take 400 sets of sample points for the preprocessed normal and faulty samples. Select 1/2 of the samples as the training set and the remaining 1/2 of the samples as the test set and then project the processed sample points to the 2D plane results as shown in Figure 14.

### 5.2.2. Experimental results and analysis

To further verify the overall performance of the proposed algorithm, complete one hundred SVM parameter selection and sample point classification by GA, PSO, GWO, and IGWO algorithms. The classification results are shown in Table 3, and the IGWO-SVM test set results are shown in Figure 15.

For the IMS dataset, the PSO algorithm is the most stable, and the optimization time is the shortest, but the accuracy of the test set is still the lowest, and the time required is longest, so the effect is worse. The GA has the largest fluctuation, the optimization time is the longest, and the time difference is the most unstable. Compared with the first two algorithms, the GWO algorithm shows better performance without regarding the average accuracy, the optimal averaging time, or the optimization time difference. The GWO algorithm is verified to be feasible for SVM parameter optimization in bearing fault diagnosis feasibility and superiority. Compared with the GWO algorithm, the proposed IGWO algorithm has improved various indicators. The average accuracy rate increased by 0.030%, the average time decreased by 6.022 s, and the optimization time decreased by 11.956 s. By comprehensive performance comparison, the IGWO algorithm converges faster to easily achieve the optimal classification. The ability to judge bearing faults is stronger, and the recognition accuracy is higher.

### Table 3. Fault recognition results of four algorithms to optimize the SVM.

| Method | Average test accuracy (%) | Maximum test accuracy (%) | Minimum test accuracy (%) | Average time (s) | ΔT (s) |
|--------|--------------------------|---------------------------|---------------------------|-----------------|--------|
| GA     | 98.162                   | 99.500                    | 93.250                    | 52.332          | 43.362 |
| PSO    | 87.385                   | 95.750                    | 87.250                    | 74.245          | 6.778  |
| GWO    | 98.715                   | 99.750                    | 97.000                    | 16.535          | 35.195 |
| IGWO   | 98.745                   | 99.750                    | 97.500                    | 10.513          | 23.239 |

ΔT = Maximum time – Minimum time.
6. Conclusions

In terms of the problem that the SVM model parameter selection has a great influence on its classification performance, this paper proposes a new method based on IGWO optimized SVM fault classification for fault diagnosis of rolling bearings. This method has the following advantages: (1) It uses the CS algorithm to update the searchability and randomness of the host’s nest position behaviour, and improve the global searchability of the GWO algorithm. The IGWO algorithm is superior to the GWO algorithm by four commonly used test functions. (2) Using the IGWO algorithm to obtain optimal parameters of the SVM classification model improves the learning ability and generalization ability of the SVM, and obtains the best fault diagnosis model. (3) The analysis of CWRU and IMS bearing fault diagnosis shows that the proposed IGWO-SVM method can accurately and effectively diagnose the bearing. Under the same circumstance, the proposed method has higher fault diagnosis accuracy and recognition efficiency compared to the GA, PSO and GWO algorithms. (4) The EEMD, SWPE and PCA feature extraction methods used in the experiment provide references for the fault feature extraction of mechanical equipment.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the National Natural Science Foundation of China [grant number 51865004], the Education Department Science and Technology Talent Support Project of Guizhou Province [grant number 2018]037], the Major Project of Science and Technology in Guizhou Province [grant number No. 2018]3002], the Key projects of Guizhou’s industrial research [grant number 2016]2332], and the Science and Technology Project of Guizhou Province [grant number 2018]5781].

References

Bao, Y., Hu, Z., & Xiong, T. (2013). A PSO and pattern search based memetic algorithm for SVMs parameters optimization. Neurocomputing, 117, 98–106.
Bian, X., Zhang, L., Du, Z., Chen, J., & Zhang, J. (2018). Prediction of sulfur solubility in supercritical sour gases using grey wolf optimizer-based support vector machine. Journal of Molecular Liquids, 261, 431–438.
Case Western Reserve University Bearing Data Center Web-site. (2019). Retrieved from online:http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website
Center for Intelligent Maintenance Systems. (2019). Retrieved from http://www.imscenter.net/
Chen, F., Tang, B., Tao, S., & Li, L. (2014). Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization. Measurement, 47(1), 576–590.
Chen, Q., Whitbrook, A., Aickelin, U., & Roadknight, C. (2014). Data classification using the Dempster–Shafer method. Journal of Experimental & Theoretical Artificial Intelligence, 26(4), 493–517.
Chun Yang, J., & Long, W. (2016). Improved grey wolf optimization algorithm for constrained mechanical design problems (Vol. 851).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Emary, E., Zawbaa, H. M., Grosan, C., & Hassanien, A. E. (2015). Feature subset selection approach by gray-wolf optimization. In Afro-European Conference for Industrial Advancement (AECIA), Addis Ababa, Ethiopia.

Fei, S. W., & Zhang, X. B. (2009). Fault diagnosis of power transformer based on support vector machine with genetic algorithm. Expert Systems with Applications, 36(8), 11352–11357.

Gao, Z., Cecati, C., & Ding, S. X. (2015). A survey of fault diagnosis and fault-tolerant techniques—part II: Fault diagnosis with knowledge-based and hybrid/active approaches. IEEE Transactions on Industrial Electronics, 62(6), 3768–3774.

Huang, C. L., & Dun, J. F. (2008). A distributed PSO–SVM hybrid system with feature selection and parameter optimization. Applied Soft Computing Journal, 8(4), 1381–1391.

Huang, N. E., Zheng, S., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., … Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings Mathematical Physical & Engineering Sciences, 454(1971), 903–995.

Jack, L. B., & Nandi, A. K. (2002). Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. Mechanical Systems & Signal Processing, 16(2–3), 373–390.

Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and support vector machines (SVMs). Applied Soft Computing, 11(6), 4203–4211.

Lei, Y., He, Z., & Zi, Y. (2011). EEMD method and WNN for fault diagnosis of locomotive roller bearings. Expert Systems with Applications, 38(6), 7334–7341.

Lei, Y., Lin, J., He, Z., & Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 35(1), 108–126.

Li, L., Qu, L., & Liao, X. (2007). Haar wavelet for machine fault diagnosis. Mechanical Systems and Signal Processing, 21(4), 1773–1786.

Liu, Z., Cui, Y., & Li, W. (2017). A classification method for complex power quality disturbances using EEMD and rank wavelet SVM. IEEE Transactions on Smart Grid, 6(4), 1678–1685.

Liu, S. C., & Liu, S. Y. (2003). An efficient expert system for machine fault diagnosis. International Journal of Advanced Manufacturing Technology, 21(9), 691–698.

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69(3), 46–61.

Noshadi, A., Shi, J., Lee, W. S., Shi, P., & Kalam, A. (2016). Optimal PID-type fuzzy logic controller for a multi-input multi-output active magnetic bearing system. Neural Computing and Applications, 27(7), 2031–2046.

Phan, A. V., Nguyen, M. L., & Bui, L. T. (2017). Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems. Applied Intelligence, 46(2), 455–469.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing diagnostics. Journal of Sound & Vibration, 289(4), 1066–1090.

Saremi, S., Mirjalili, S. Z., & Mirjalili, S. M. (2015). Evolutionary population dynamics and grey wolf optimizer. Neural Computing and Applications, 26(5), 1257–1263.

Schlechtgen, M., & Santos, I. F. (2011). Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mechanical Systems & Signal Processing, 25(5), 1849–1875.

Sulaiman, M. H., Mustaffa, Z., Mohamed, M. R., & Aliman, O. (2015). Using the gray wolf optimizer for solving optimal reactive power dispatch problem. Applied Soft Computing, 32(C), 286–292.

Wang, J., Tang, S., Di, L., Wang, S., Liu, C., Abbas, H., … Vasilakos, A. V. (2017). A manufacturing big data solution for active preventive maintenance. IEEE Transactions on Industrial Informatics, 13(4), 1.

Wang, C., Zhang, Y., Song, J., Liu, Q., & Dong, H. (2019). A novel optimized SVM algorithm based on PSO with saturation and mixed time-delays for classification of oil pipeline leak detection (Vol. 7).

West, G. M., McArthur, S. D. J., & Towle, D. (2012). Industrial implementation of intelligent system techniques for nuclear power plant condition monitoring. Expert Systems With Applications, 39(8), 7432–7440.

William, P. E., & Hoffman, M. W. (2011). Identification of bearing faults using time domain zero-crossings. Mechanical Systems & Signal Processing, 25(8), 3078–3088.

Wu, J. D., & Liu, C. H. (2009). An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. Expert Systems with Applications, 36(3), 4278–4286.

Yang, X. S., & Deb, S. (2010). Cuckoo search via Lévy flights. In World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India.

Zhang, X., Kang, Q., Cheng, J., & Wang, X. (2018). A novel hybrid algorithm based on biogeography-based optimization and grey wolf optimizer. Applied Soft Computing, 67, 197–214.

Zhang, X., Liang, Y., Zhou, J., & Yi, Z. (2015). A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. Measurement, 69, 164–179.

Zhang, X., & Zhou, J. (2013). Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. Mechanical Systems & Signal Processing, 41(1–2), 127–140.

Zhang, S., Zhou, Y., Li, Z., & Pan, W. (2016). Grey wolf optimizer for unmanned combat aerial vehicle path planning. Advances in Engineering Software, 99, 121–136.

Zhao, L., Ding, J., Huang, D., Mi, B., Ke, L., & Liu, Y. (2018). The incipient fault feature enhancement method of the gear box based on the wavelet packet and the minimum entropy deconvolution. Systems Science & Control Engineering, 6(3), 235–241.

Zhaohua, W. U., & Huang, N. E. (2009). Ensemble empirical mode decomposition: A Noise-assisted data analysis method. Advances in Adaptive Data Analysis, 1(01), 1–41.

Zvokelj, M., Zupan, S., & Prebil, I. (2011). Non-linear multivariate and multiscale monitoring and signal denoising strategy using kernel principal component analysis combined with ensemble empirical mode decomposition method. Mechanical Systems & Signal Processing, 25(7), 2631–2653.