Fast bearing fault diagnosis of rolling element using Lévy Moth-Flame optimization algorithm and Naive Bayes

Indexed by:
Web of Science Group

Shuang Sun [a], Krzysztof Przystupa [b]*, Ming Wei [c], Han Yu [c], Zhiwei Ye [a], Orest Kochan [a,d]

[a]School of Computer Science, Hubei University of Technology, Wuhan, Hubei, 430068, PR China
[b]Lublin University of Technology, ul. Nadbystrzycka 36, 20-618, Lublin, Poland
[c]Wuhan Fiberhome Technical Services Co., Ltd., Wuhan FiberHome Telecommunication Technologies Co., Ltd., Wuhan, Hubei, 430074, PR China
[d]Lviv Polytechnic National University, Karpinskoho str. 1, Lviv, Ukraine

Highlights

- A fault diagnosis method based on LMFO, ensures high classification accuracy and better efficiency.

- EEMD-based feature extraction method effectively removes signal noise.

- The feature selection method based on LMFO effectively removes feature redundancy.

- The NB-based fault diagnosis method ensures accuracy with high efficiency.

Abstract

Fault diagnosis is part of the maintenance system, which can reduce maintenance costs, increase productivity, and ensure the reliability of the machine system. In the fault diagnosis system, the analysis and extraction of fault signal characteristics are very important, which directly affects the accuracy of fault diagnosis. In the paper, a fast bearing fault diagnosis method based on the ensemble empirical mode decomposition (EEMD), the moth-flame optimization algorithm based on Lévy flight (LMFO) and the naive Bayes (NB) is proposed, which combines traditional pattern recognition methods meta-heuristic search can overcome the difficulty of selecting classifier parameters while solving small sample classification under reasonable time cost. The article uses a typical rolling bearing system to test the actual performance of the method. Meanwhile, in comparison with the known algorithms and methods was also displayed in detail. The results manifest the efficiency and accuracy of signal sparse representation and fault type classification has been enhanced.

Keywords

malfunction diagnostics; naive Bayes; moth-flame optimization algorithm; ensemble empirical mode decomposition.

1. Introduction

Mechanical fault diagnosis is a technology that monitors the state of the machine during operation, determines its overall or partial normality or abnormality, determines the fault and its cause, and can predict the development trend of the fault. Due to the improvement of the automation and integration of the current mechanical system, the maintenance time and cost have gradually increased [7, 14, 21, 47], causing greater economic losses [2, 3, 19, 34]. Therefore, the importance of the fault detection system in the daily maintenance system is becoming stronger [4, 15, 24, 50].

The fault diagnosis method can be divided into time-frequency analysis method and pattern recognition method. The traditional time-frequency analysis method identifies the fault based on the energy change of each frequency band of the vibration signal when the fault occurs. Commonly used methods include Wavelet transform (WT) [37], Empirical mode decomposition (EMD) [11], ensemble empirical mode decomposition (EEMD) [23], and etc [1, 48]. WT can effectively extract the characteristics of nonlinear transient vibration time-frequency signals, but when processing complex vibration signals, different basis functions need to be selected to obtain the best results, and there is no uniform standard for parameter selection. EMD has

adaptive signal processing capabilities, but it has disadvantages such as edge effect and modal mixing [14].

The existing pattern recognition methods mainly include ELM [28], SVM [7, 48], ANFIS [22, 46], and etc. ELM has fast learning speed and requires few training samples, and can realize fast main bearing fault diagnosis, but its stability is relatively weak. SVM can efficiently solve high-dimensional nonlinear decision-making problems, but it is difficult to select kernel parameters and sample parameters, and is significantly affected by fault samples. However, for the pattern recognition method, due to the high cost of obtaining mechanical system failure vibration experimental data and the limited degree of failure, which makes the existing methods have limited recognition accuracy and even misrecognition problems.

Machine learning related methods have certain advantages in small sample classification. However, the original signal is distorted and non-linear, and is often buried in a large amount of background noise and other interference. Therefore, higher requirements are put forward for signal preprocessing and feature extraction. Meanwhile, with different varieties of features given specific effectiveness, it is feasible to apply compound features for eliminating information loss. However, the feature combination will introduce new proper information and

will surely bring some redundancy, which may reduce the performance of fault diagnosis. Thus, feature selection is also necessary.

In the paper, the EEMD is utilized for preprocessing and feature extraction. The character of the bearing is reflected precisely and adequately from the IMFs [6, 9, 12, 13, 15, 49]. As for feature selection, according to the organization of the search process, the approached of feature selection can be roughly classified as filters, wrappers, and embedded [39]. The wrapper methods have many successful applications in recent years, such as Wan [42] combined genetic algorithm and modified binary-coded ant colony optimizer for the wrapper. Wang [43] and Liu [27] utilized the particle swarm optimization for searching the optimal feature subset. Rodrigues [36] introduced an approach based on Bat Algorithm. Zawbaa [45] applied Chaotic Ant lion Optimization to solve the problem of feature selection. Li [26] employed the Grey Wolf Optimization. Mafarja [29] and Sharawi [38] presented solutions with the whale optimization algorithm. Therefore, this article conducts a more in-depth study on the wrapper method.

Since, the aim of the wrapper is to search for the optimal feature subset. As it is a non-convex, discrete, non-linear, and not derivable problem, it is challenging for conventional methods to get solutions. In order to solve such problems, the heuristic search is applied. Although this method cannot guarantee to find the optimal solution, it can find an acceptable solution within a reasonable time. In recent years, with the deepening of research, more and more algorithms are proposed, including nature-inspired algorithms, swarm-based, and physics-based approaches. Such as, Particle Swarm Optimization (PSO) [5, 16-18, 30] Differential Evolution (DE) [40], Genetic Algorithm (GA) [8, 10], Ant Colony Optimization (ACO) [4], Gravitational Search Algorithm (GSA) [35], whale optimization (WOA) [29, 38], Lighting Attachment Procedure Optimization (LAPO) [33] and Moth-flame optimization algorithm (MFO) [32], etc. MFO, a new nature-inspired meta-heuristic algorithm with excellent-searching performance has been widely used in diverse practical applications, such as the annual power load forecasting [25], optimal reactive power dispatch [31, 41], and the wireless sensor networks [20]. Although MFO has not been applied to the problem of fault diagnosis, the successful application of the above-mentioned fields shows the powerful optimization ability of the MFO algorithm. Then, in the paper, a wrapper feature selection method based on MFO is proposed.

Meanwhile, according to the No-Free-Lunch (NFL) [44] theorem that none of the algorithms can solve all optimization problems, which clarifies the need for new and specific algorithms in different areas, and the effectiveness of an algorithm for a set of puzzles does not guarantee its success in other fields. Therefore, it is worth studying in depth whether the MFO algorithm also has excellent performance in the problem of fault diagnosis. In order to further improve the stability of the algorithm and the ability to jump out of the local optimum, we applied the Lévy flight to the MFO. The Lévy flight can generate a random walk for the population which can enhance the exploration ability and avoid trapping around the local optimum.

The rest of the paper is organized as follows. Section 2 presents an overview of the classification models. Section 3 illustrates some details of the LMFO and the proposed method. Section 4 describes the experiments and results.

2. Overview of LMFO

As far as the feature subset selection problem is concerned, it is a non-linear, indirect, and strongly constrained optimization problem, and heuristic search algorithms can effectively solve this type of problem. The MFO algorithm used in this article is a new nature-inspired meta-heuristic algorithm with strong search capabilities and the ability to overcome local optima. In order to further improve the stability of the algorithm and avoid falling into local extremes, the Levy Flight optimizer is used. In order to make a one-to-one correspondence between the individuals and feature selection, in the paper, the binary coding is employed for mapping.

2.1. The moth-flame optimization algorithm (MFO)

Mirjalili S. proposed the moth-flame algorithm in 2015 [32], which is inspired by the method of moth navigation. The moth keeps a specific angle to the moon by an efficient, extended, straight flight mechanism at night, but may also be trapped in an ineffective, lethal spiral around artificial light. This behavior is then mathematically modeled and proposed as an optimizer called Moth-Flame optimization.

The MFO is a three-tuple and the global optimal can be defined as follows:

$$LMFO = (I, P, T) \tag{1}$$

where:

I represents a random generation function of moths and their associated parameters, P function is the main function to simulate moth motion in searching space. And the T function controls the termination of the algorithm.

As defined above, after the initialization of the MFO, the P function iterates until the T function returns true. The P function is the primary function of moths' position updating, in which the distance between moths and flame can be formulated by the following:

$$M_i^* = S(M_i, F_j) \tag{2}$$

where the:

i-th moth is represented by M, and F_j stands for the j-th flame.

For the spiral function S, all types of spirals can be used to fit the moth's trajectory. In the paper, the logarithmic spirals are used:

$$S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j \tag{3}$$

$$D_i = |F_j - M_i| \tag{4}$$

The distance from i-th to the j-th flame is represented by D_i, the constant that determines the shape of the spiral is represented by b, and t is the random number in the range [-1,1].

However, when updating the location of n moths in search space, the utilization of the optimal solution will be reduced. Then the algorithm utilized an adaptive mechanism for the flames number, which can be formulated as follows:

$$flame \, numbe = round\left(N - I * \frac{N-1}{T}\right) \tag{5}$$

2.2. Lévy flight distribution

In 1926, Paul Lévy, the French mathematician, proposed Lévy flights which is a kind of random walk and is very common in nature. As in the individual MFO, the population diversity will reduce with the iteration proceeding and the algorithm will easily fall into the local optimum. By adding the Lévy model for updating the position after each iteration, the global search ability and the capacity of jumping out of the local optimum can be improved. The moth position update as follows:

$$x_i^{t+1} = x_i^t + \alpha \oplus lévy(\lambda) \tag{6}$$

where:

x_i^t is the current position of the i-th moth, t is the number of current iterations.

The parameter a is the step size which related to the scale of the problem, here values 0.01, and the mathematical model of Lévy flight is defined as below:

$$lévy(\lambda) = \frac{\mu\sigma}{|v|^{\frac{1}{\lambda}}} \qquad (7)$$

Here the μ, v are two random number in [0,1], λ is a constant of 1.5, and σ is calculated as follows:

$$\sigma = \frac{\Gamma(1+\lambda)\sin\frac{\pi\lambda}{2}}{\Gamma\left(\frac{1+\lambda}{2}\right)\lambda 2^{\frac{\lambda-1}{2}}} \qquad (8)$$

where:
Γ is the normal gamma function.

2.3. The LMFO

Based on the above discussion, the combination of MFO and Lévy-flight can further improve the MFO algorithm based on the excellent search ability of MFO to avoid the problem of local optimization, and improve the stability and optimization ability of the algorithm.

The general steps of the LMFO algorithm are as follows:

Algorithm 1 Procedure of LMFO algorithm

Initialize the population, the moths(the potential variables), $Xi(i = 1, 2,\dots,$ n) considering *ub* and *lb* **while** end condition is not satisfied **do**
    updating the number of flames using Eq.5;
    calculate the fitness of the moths
    if iteration == 1 **then**
        the best flame = sort(moths)
        the best flame fitness = sort (the fitness of the moths)
    else
        the best flame = sort($moths_{t\_1}, moths_t$)
        the best flame fitness = sort(the fitness of $(moths_{t-1}, moths_t)$
    end if
    for i=1:*Npop* (each moth) **do**
        for j = 1:N_{dim}(each number of variable) **do**
        update the convergence constants r and t;
          calculate the distance to the flame using Eq.4 with the respect to the corresponding moth; update the moth(*I, j*) using Eq.2 and Eq.3 with the respect to the corresponding moth check the boundaries
        end for
        update the position of the current moth using Lévy-flight
    end for iteration = iteration + 1
end while

---

3. The proposed method

As the fault diagnosis of rolling bearing is a multi-kind classification, for improving the accuracy of fault type recognition, one of the most important processes is to select the optimal feature subset from the extracted features. In the paper, a wrapper method based on LMFO and NB is proposed. The main purpose is to integrate the classifier with the feature selection strategy based on LMFO after features have been extracted, then use the classification performance of the classifier as an evaluation criterion updating the population, the core is the encoding schema and the fitness function.

3.1. Feature extraction

Feature extraction is an analysis step that affects the accuracy and efficiency of the diagnosis model. Since the data used in this article was collected by an accelerometer mounted on the drive side. The data itself is unstable, non-linear, and contains a lot of noise. As an adaptive decomposition method of nonlinear non-stationary signals, EMD can process vibration signals well, decompose the vibration signals into multiple IMFs, and constantly change from high frequency to low frequency, as shown in Figure 1. Due to modal mixing, the
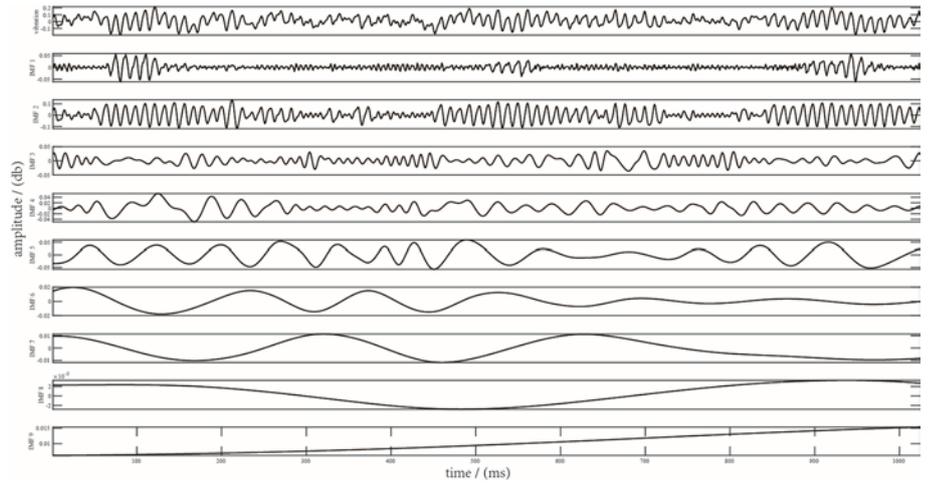

Fig. 1. The IMF component diagram based on the EEMD algorithm

extraction of erroneous feature information has a certain probability, which makes the physical information of the IMF obscure. In order to solve the modal mixing problem, Wu and Huang proposed EEMD to improve the distribution of extreme points by using the statistical properties of uniform distribution of Gaussian white noise to decompose vibration signals of the normal state.

After the signal is decomposed, 12 energy features of the IMFs are obtained which is calculated as follows:

$$E_i = \int_{-\infty}^{+\infty} \left| c_i(t)^2 \right| dt \qquad i = 1, 2, \dots, n \qquad (9)$$

E_i is the energy of the IMFs, the $c_i(t)$ is the IMFs of the original vibration signal, and the energy feature calculated as below:

$$E_i = \frac{E_i}{\sum_{i=1}^{n} E_i} \quad i = 1, 2, \dots, n \qquad (10)$$

Also, in this work, 24 commonly used statistical feature parameters (p1-p24) are employed. Among which there are eleven parameters of time-domain demographic features (the time sequence chart is shown in Figure 2 and thirteen feature parameters for frequency-domain from its FFT spectrum(the frequency spectrum is shown in Figure 3. The definition of twenty-four feature parameters is given in Table 1.

The time domain signal may fluctuate when the machine fails. When the amplitude and distribution may be different from the conventional time domain signal, the spectrum and its distribution may

also change, that is, new frequency portions may appear, and the convergence of the spectrum may also be different. According to these features, corresponding features can be extracted.

While the above features can characterize the failure of rotating machinery from different angles, they differ in their importance for different types of fault identification. Therefore, in order to improve the performance of the classifier and avoid the dimensionality disaster, it is necessary to select the features related to the fault from the feature set, and remove the uncorrelated or redundant features before inputting the feature set into the classifier.

At present, most noise reduction applications take the high-frequency components obtained by EMD as the direct noise removal, and in many cases, it is possible to remove the useful signal components. Especially for rolling bearings, fault-related shock signal components are usually in the high-frequency band. In EMD decomposition, pseudo-components often appear, that is, components independent of the original signal. The frequency components contained in these pseudo-components may exist in the coincidence of characteristic frequency bands, so measures should be taken to identify them and eliminate them.

3.2. LMFO based feature selection

3.2.1. The binary encoding of individuals

In order to apply LMFO for feature selection, the individuals need to be encoded. Which intends a string of characters is needed to stand for the feature is chosen or not. Moreover, it is not difficult to find that the binary string can meet our needs, by using '1' means selected, '0' do not choose. However, LMFO is real-coded which needs to be converted to binary via a transformation function, where the sigmoid function is used.

$$sigmoid = \frac{1}{1 + e^{-x}} \qquad (11)$$

Since the LMFO algorithm is initialized and iterative calculated by real number, so as to handle the problem of feature selection, the upper and lower bounds of the algorithm are specified in [-1,1]. After the initialization and iterative calculation are completed, the real-coded individuals will be converted into binary-coded individuals through the sigmoid function and conversion function to prepare for subsequent feature selection and classification. The transformation process defined as follows:

$$M_{ij} = \begin{cases} 1 & \text{if } M_{ij} > 0.5 \\ 0 & \text{if } M_{ij} \leq 0.5 \end{cases} \qquad (12)$$

where:

M_{ij} . indicates the value of the j-th dimension of the i-th moth

3.2.2. The feature selection scheme

Through the above feature extraction process, 12 energy features representing IMF components and 24 statistical features of the raw signal are extracted, but in order to improve the accuracy and efficiency of subsequent learning classification, the dimensionality reduction and de-redundancy for features becomes very important, here, the LMFO is used to search for optimal feature subsets. The first problem to be solved is the mapping between individuals and actual problems in a population. In the paper, binary strings are used to correspond to the selection of features. The real-coded individuals are converted into binary codes by the conversion function, and then 36 features are corresponding to the binary coding. Each individual corresponds to a feature subset. For example, suppose a feature selection problem with 10 features, in which an individual code is "0100110111", that is, no. 2, 5, 6, 8, 9, and 10 features are selected. The specific step is shown as Algorithm 2.

Table 1.   The statistical features of the raw signal

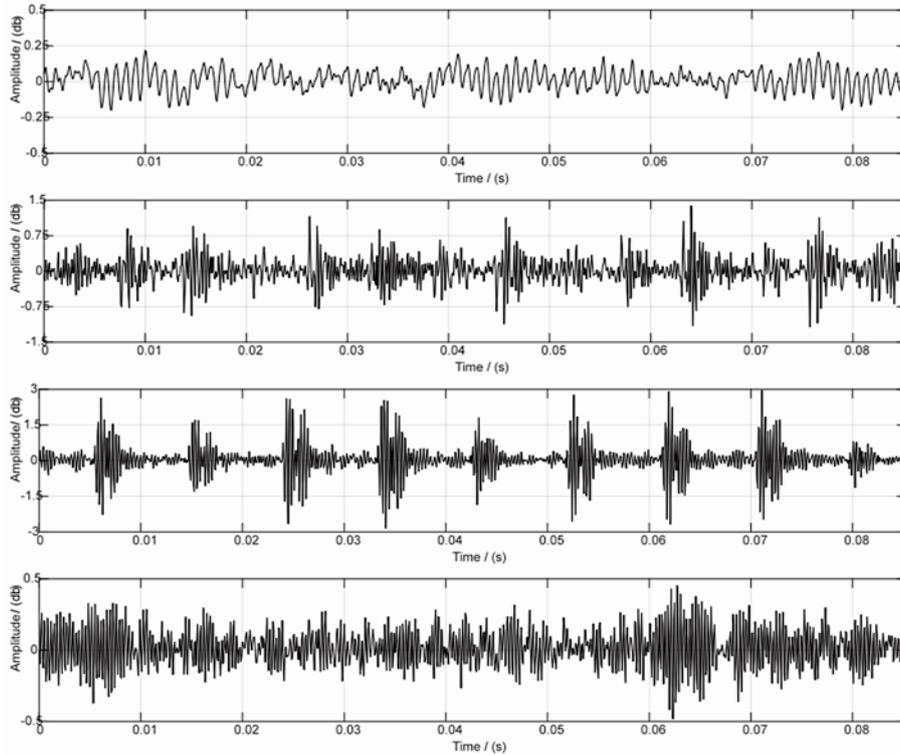| Time-domain feature parameters		Frequency-domain parameters	
$p_1 = \dfrac{\sum_{n=1}^{N} x(n)}{N}$	$p_2 = \sqrt{\dfrac{\sum_{n=1}^{N}(x(n) - p_1)}{N-1}}$	$p_{12} = \dfrac{\sum_{k=1}^{K} s(k)}{K}$	$p_{13} = \dfrac{\sum_{k=1}^{K}(s(k) - p_{13})^2}{K-1}$
$p_3 = \left(\dfrac{\sum_{n=1}^{N}\sqrt{|x(n)|}}{N}\right)^2$	$p_4 = \sqrt{\dfrac{\sum_{n=1}^{N}(x(n))^2}{N}}$	$p_{14} = \dfrac{\sum_{k=1}^{K}(s(k) - p_{12})^3}{K\left(\sqrt{p_{13}}\right)^3}$	$p_{15} = \dfrac{\sum_{k=1}^{K}(s(k) - p_{12})^4}{K p_{13}^2}$
$p_5 = \max|\bar{x}(n)|$	$p_6 = \dfrac{\sum_{n=1}^{N}(x(n) - p_1)^3}{(N-1) p_2^3}$	$p_{16} = \dfrac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$	$p_{17} = \sqrt{\dfrac{\sum_{k=1}^{K}(f_k - p_{16})^2 s(k)}{K}}$
$p_7 = \dfrac{\sum_{n=1}^{N}(x(n) - p_1)^4}{(N-1) p_2^4}$	$p_8 = \dfrac{p_5}{p_4}$	$p_{18} = \sqrt{\dfrac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$	$p_{19} = \sqrt{\dfrac{\sum_{k=1}^{K} f_k^4 s(k)}{\sum_{k=1}^{K} f_k^2 s(k)}}$
$p_9 = \dfrac{p_5}{p_3}$	$p_{10} = \dfrac{p_4}{\frac{1}{N}\sum_{n=1}^{N}|x(n)|}$	$p_{20} = \dfrac{\sum_{k=1}^{K} f_k^2 s(k)}{\sqrt{\sum_{k=1}^{K} s(k) \sum_{k=1}^{K} f_k^4 s(k)}}$	$p_{21} = \dfrac{p_{17}}{p_{16}}$
$p_{11} = \dfrac{p_5}{\frac{1}{N}\sum_{n=1}^{N}|x(n)|}$		$p_{22} = \dfrac{\sum_{k=1}^{K}(f_k - p_{16})^3 s(k)}{K p_{17}^3}$	$p_{23} = \dfrac{\sum_{k=1}^{K}(f_k - p_{16})^4 s(k)}{K p_{17}^4}$
Where $x(n)$ is a signal series for $n=1,2,...,N$ is the number of the data points.		$p_{24} = \dfrac{\sum_{k=1}^{K}\sqrt{f_k - p_{16}} s(k)}{K\sqrt{p_{17}}}$	
		Where $s(k)$ is a spectrum for $k=1,2,...,K$. K is the number of spectrum lines: f_k is the frequency value of the kth spectrum line.	

Fig. 2. The time sequence chart: (a) the normal condition, (b) the inner fault condition, (c) the outer fault condition, (d) the roller fault condition
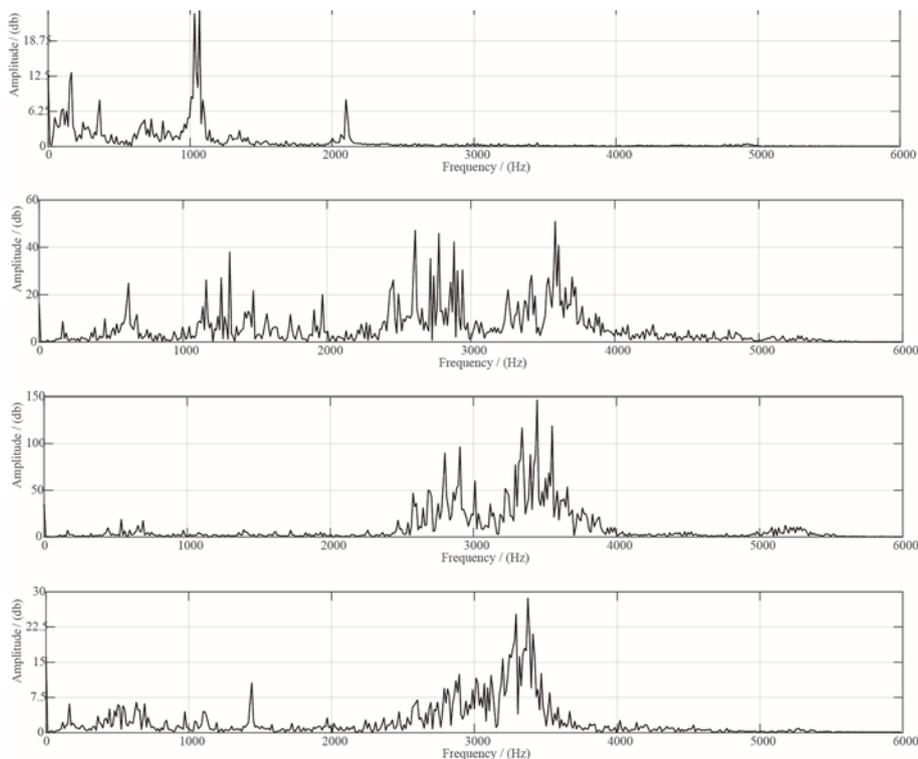

Fig. 3. The frequency spectrum: (a) the normal condition, (b) the inner fault condition, (c) the outer fault condition, (d) the roller fault condition

3.2.3. The fitness function

The definition of the fitness function is one of the cores of the evolutionary algorithm, which determines the actual advantages and disadvantages of each solution in the algorithm. For the feature selection problem, we hope to achieve the optimal classification accuracy while selecting as few features as possible, and the following objective function is defined.

$$F(i) = \frac{Accuracy(i)}{1 + \lambda \cdot n(i)} \qquad (13)$$

where:

Accuracy(i) represents the actual classification accuracy of the feature subset corresponding to the *i*-th individual; *n(i)* represents the number of features selected by the *i*-th individual; λ is a constant value which takes 0.01 in the paper.

Algorithm 2 Procedure of feature selection

Initialize the population, the moths(the potential variables), $Xi(i = 1, 2,…,$ n) considering *ub* and *lb* **while** end condition is not satisfied **do**

updating the number of flames using Eq.5;
convert the individuals into binary encoding using Eq.11 and Eq.12;
calculate the fitness of the moths using Eq.16
if iteration == 1 **then**
the best flame = sort(moths)
the best flame fitness = sort (the fitness of the moths)
else
the best flame = sort(*moths_{t_1}, moths_t*)
the best flame fitness = sort(the fitness of *(moths_{t−1}, moths_t)*)
end if
for i=1:*Npop* (each moth) **do**
for j = 1:*N_{dim}*(each number of variable) **do**
update the convergence constants *r* and *t*;;
calculate the distance to the flame using Eq.4 with the respect to the corresponding moth;
update the moth(*I, j*) using Eq.2 and Eq.3 with the respect to the corresponding moth;
check the boundaries;
end for
update the position of the current moth using Lévy-flight;
end for iteration = iteration + 1
end while

3.3. The classifier

Bayesian decision theory is a primary method for decision making under the probability framework. Assuming all relevant probabilities are known, Bayesian decision theory studies how to choose the best category marker in this ideal situation based on these probabilities and misjudgment losses. Naive Bayes is a classification method based on Bayes' theorem and the independent hypothesis of specific conditions with stable classification efficiency.

In order to minimize the decision risk based on the Bayesian decision criterion, the posterior probability should be obtained first. Based on the Bayesian theorem, it can be written as:

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)} \qquad (14)$$

where:

$P(c)$ is the prior probability, $P(c|x)$ is the likelihood, and $P(x)$ is the normalized constant.

The main difficulty of posterior probability estimation based on the above formula is that the quasi-conditional probability is the joint probability on all attributes, which is difficult to estimate directly from the limited training samples. The joint probability is directly estimated based on the limited training samples, and the combination number explosion will occur in the calculation, and the data will encounter sparse samples. The more attributes, the more serious the problem will be. To avoid this obstacle, researchers adopt the naive Bayes classifier of the 'attribute-conditional independence hypothesis': for a known category, all attributes are assumed to be independent of each other. In other words, all attributes are assumed to influence the classification result independently.

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)} = \frac{P(c)}{P(x)}\prod_{i=1}^{d}P(x_i|c) \qquad (15)$$

For all categories, the $P(x)$ is the same. So that the Naive Bayesian criterion can be written as follows:

$$h_{nb}(x) = \arg \max_{c \in y} P(c)\prod_{i=1}^{d} P(x_i \mid c) \qquad (16)$$

The number of attributes is represented by *d*, and the value of *x* on the *i*-th attribute is represented by x_i.

3.4. The implementation of the proposed method

In order to adapt to the fault diagnosis in the environment of small sample data and improve the time efficiency of high-dimensional data while ensuring the accuracy of decision-making, a fault diagnosis method based on LMFO is proposed. In the data preprocessing stage, the article uses the EEMD algorithm to process the original vibration signal to reduce the influence of background noise on the experimental diagnosis. Secondly, 20 common fault detection features are extracted through IMFs. In order to improve the accuracy and efficiency of model training, the classification model is optimized through the wrapper feature selection method based on LMFO, while the proposed features are selected. The specific implementation steps of the method are shown in Figure 4.
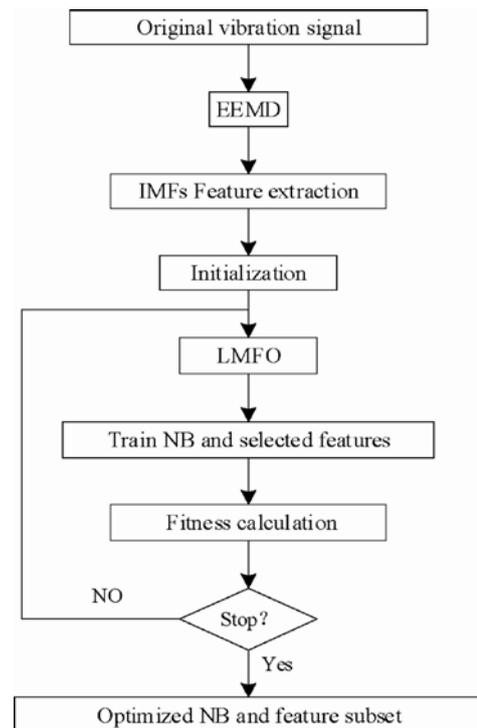

Fig. 4. The scheme of the proposed LMFO-NB

4. Experiment and results

The proposed method is simulated in the MATLAB environment and the experiment programmers are coded by python language on a personal computer with a 3.2GHz CPU and 16.00G RAM under windows 10 operating system.

4.1. Data set description

The test platform consists of a 2hp motor, dynamometer, torque sensor and electronic control unit, as shown in Figure 4. In the experiment, EDM formed a single point fault with a diameter of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils (1 mil = 0.001 inch). The 7,14 and
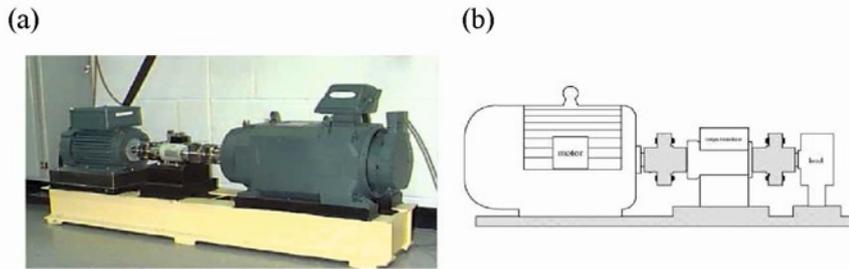

Fig. 5. Schematic diagram of CWRU bearing test platform: (a) the actual photos, (b) simplified schematic diagram

21 mils diameter faults are using the SKF bearing and the 28 mils and 40 mil faults by NTN bearing. The main experimental object of the paper is SKF6205-2rs deep groove ball bearing.

The vibration signal is obtained by an accelerometer, which is typically placed at the fan end or the drive end of the motor housing at the 12 o'clock position. In the paper, the 16-channel DAT recorder is used to collect the data of the driver end, and the sampling frequency is 12 kHz. Figure 5 is the drive end envelope spectrum, and the Figure 6 is the fan end envelope spectrum. As can be seen from the figure below, the characteristic frequency of the internal fault is 162.4 Hz, the characteristic frequency of the external fault is 107.5 Hz, and the characteristic frequency of the roll fault is 141.4 Hz. Bearing failures are handled by electroplating processes and are classified into internal faults, external faults, and roller faults [1]. External faults are located at 3 o'clock (directly in the load zone), 6 o'clock (perpendicular to the load zone) and 12 o'clock. The data in this paper consists of normal bearing data and the above three types of failure data. The detailed description of the experimental data set is shown in Table 2. For each data sets, each data sample consist of 4096 data

points, 100 groups are randomly selected from the above states as training data, and rest of data are testing data.

4.2. Experiment for different algorithms

The experiments in this section are mainly used to assess the exploration performance of the proposed method and other optimization algorithms. In this section, we use seven other different commonly used and latest optimization algorithms (particle swarm optimization, grey wolf optimization, whale optimization algorithm, multiverse optimization algorithm, firefly algorithm, gravity search algorithm, moth flame optimization). In order to ensure that the experimental results only reflect the differences in optimization capabilities between different algorithms, the feature extraction process of different algorithms and the selection of classifiers is the same. The size
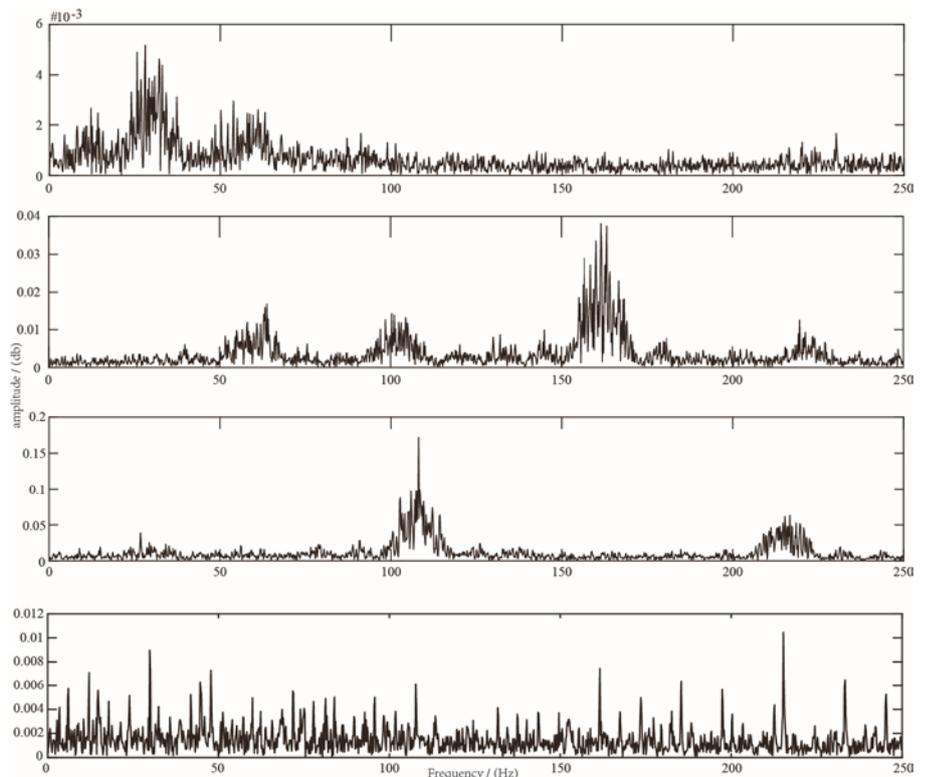

Fig. 6. The envelope spectrum of the drive end: (a) the normal condition, (b) the inner fault condition, (c) the outer fault condition, (d) the roller fault condition

Table 2. Description of the data sets

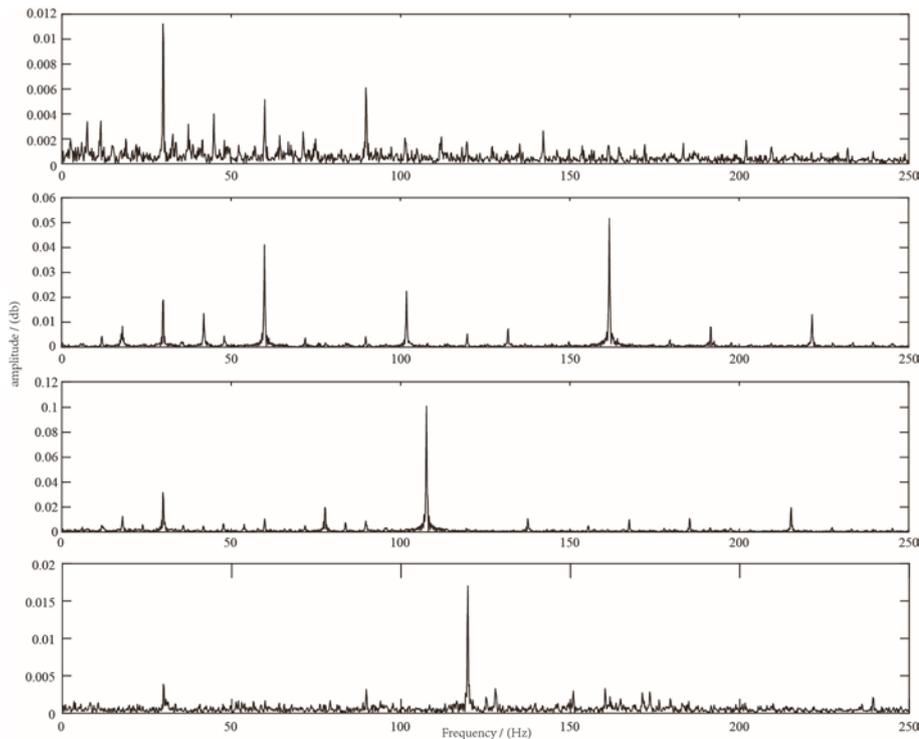| Fault type	Fault diameter	10 fault type				24 fault type			
		0hp	1hp	2hp	3hp	0hp	1hp	2hp	3hp
Normal		✓				✓	✓	✓	✓
BA	0.007	✓				✓	✓	✓	✓
	0.014								
	0.021					✓	✓	✓	✓
	0.028	✓							
IR	0.007	✓				✓	✓	✓	✓
	0.014	✓							
	0.021	✓				✓	✓	✓	✓
	0.028	✓							
OR	0.007	✓				✓	✓	✓	✓
	0.014	✓							
	0.021	✓				✓	✓	✓	✓
	0.028								

Where: BA: ball, IR: inner race, OR: outer race.

Fig. 7. The envelope spectrum of the fan end: (a) the normal condition, (b) the inner fault condition, (c) the outer fault condition, (d) the roller fault condition

of the initial population of each algorithm is 20, the maximum number of iterations is 200, the dimension value range in [-1, 1], and the experiment of each algorithm is repeated of 10 times.

As can be seen from the result data given in the Table 3, the final performance of each algorithm is not much different. Since the pre-processing feature data of the noise reduction method described in the paper is more accurate in characterizing the 10 fault types data set, the selected classifier has appropriate classification performance for the optimization problem. However, it can still be seen from the table that the LMFO, MFO, GSA, FFA and WOA algorithms have excellent search ability with maximum fitness 0.980392. Also, we can find that the number of features selected by LMFO is less, although the classification accuracy is not optimal, it still has a 4% advantage over the worst FFA, so the feature subset selected by LMFO characterized the faults better with less redundancy. At the same time, the LMFO algorithm is more robust through the comparisons of the standard deviation, the minimum fitness value, and the average fitness value. Moreover, although the convergence speed of LMFO is not the fastest from the convergence curve shown in Figure 7, its exploration ability and the ability to jump out of local optimum are better.

Table 3. The results of different optimizer for 10 fault types

Items	Optimizer							
	PSO	GWO	WOA	MVO	FFA	GSA	MFO	LMFO
$Nf.$	7	16	6	7	6	3	2	2
F_{min}	0.841049	0.837021	0.943396	0.890673	0.86478	0.939465	0.953526	0.970874
F_{max}	0.961538	0.970874	0.980392	0.961538	0.980392	0.980392	0.980392	0.980392
F_{avg}	0.911471	0.926884	0.972612	0.931452	0.929094	0.968846	0.972815	0.978488
$Acc.$	97.5274%	100%	100%	99.6654%	95.6967%	99.7911%	99.2271%	99.8058%
$std.$	0.034525	0.037666	0.010887	0.020314	0.034668	0.011704	0.0085	0.003807

Where: $Nf.$: the number of selected features, F_{min}: minimum of the fitness, F_{max}: maximum of the fitness, F_{avg}: the average fitness; $Acc.$: the classification accuracy; $std.$: the standard deviation

Table 4. The results of different optimizer for 24 fault types

Items	Optimizer							
	PSO	GWO	WOA	MVO	FFA	GSA	MFO	LMFO
$Nf.$	10	11	11	7	8	6	6	5
F_{min}	0.768956	0.768956	0.783557	0.774065	0.769747	0.830292	0.809807	0.84474
F_{max}	0.828677	0.847673	0.840486	0.850271	0.862457	0.862457	0.865247	0.865247
F_{avg}	0.798248	0.806911	0.8178736	0.812916	0.817077	0.854611	0.849027	0.861536
$Acc.$	87.8073%	89.56571%	89.9620%	86.9820%	88.2443%	90.5888%	89.9969%	90.4615%
$std.$	0.016776	0.027693	0.018242	0.024699	0.024736	0.009607	0.017167	0.006023

Where: $Nf.$: the number of selected features, F_{min}: minimum of the fitness, F_{max}: maximum of the fitness, F_{avg}: the average fitness; $Acc.$: the classification accuracy; $std.$: the standard deviation
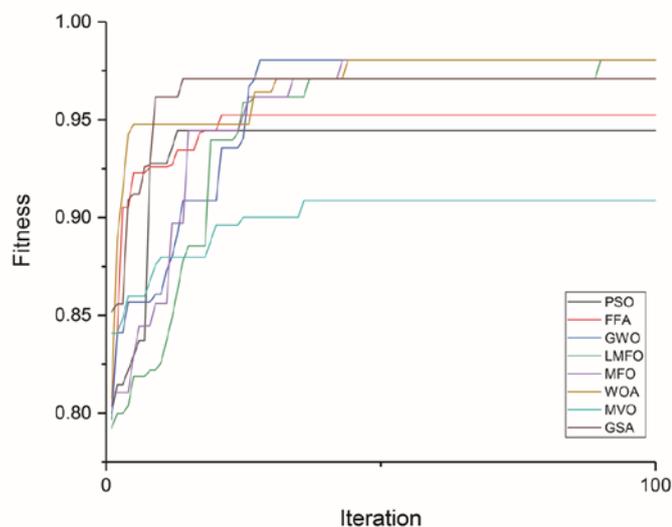
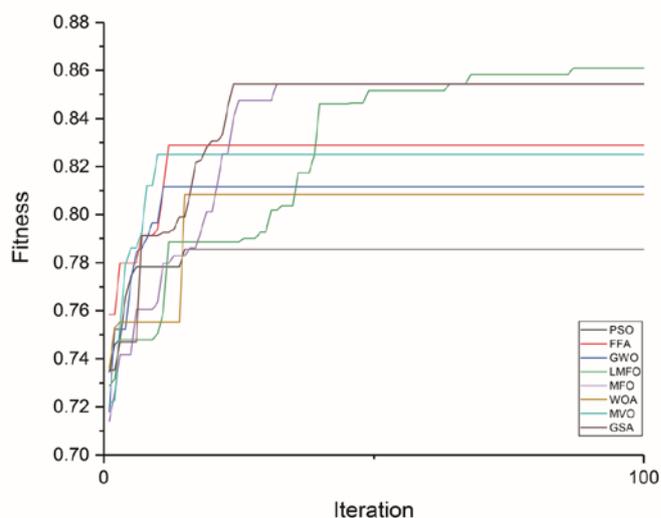Fig. 8. The convergence curve of different algorithms for 10 fault conditions

Fig. 9. The convergence curve of different algorithms for 24 fault conditions

Table 5. Comparison of different fault diagnosis methods for 10 types fault diagnosis

Methods	Fault types	Accuracy
Our method	10	99.85%
ANFIS-GAs (Lei et al.2007) [22]	10	91.33%
MSE-ANFIS (Zhang et al.2010) [46]	10	99.38%

Table 6. Comparison of different fault diagnosis methods for 24 types fault diagnosis

Methods	Fault types	Accu-racy	execution time(s)
EEMD-ICDSVM (Zhang et al.2013) [48]	24	77.29%	189.39
	24	82.05%	251.99
Our method	24	90.46%	62.99
HGSA-ELM(Luo et al.2016) [28]	24	93.04%	518.98

The results of the 24 fault types experiment are shown in the Table 4. The basic characteristics of the results are similar to those of the 10 fault types. LMFO also exhibits excellent optimization ability and robustness in the experiment. And it can be more clearly seen from the convergence curve shown in the Figure 8 that LMFO has the best performance of jumping out of the local optimum.

4.3. Comparison of commonly used fault diagnosis methods

To illuminate the practical effectiveness, a simple comparison with other methods in the literature is put forward in the Table 5 and table 6, using the same rolling bearing data and evaluated at similar operating conditions. For 10 fault types experiments, an improved distance assessment technique combined with the multi-adaptive neuro-fuzzy inference system method based on EMD is proposed in [22]. The experimental results show that under the ten fault types, the average test accuracy is 91.33%. Similar in [46], the experiment based on multi-scale entropy and multi-adaptive neuro-fuzzy inference system is 99.38%. Compared with this study, our method achieved good per-

formance under most similar conditions, with an average accuracy of 99.85%.

For 24 types fault diagnosis, authors utilized the energy entropy and the IMFs decomposed by the EEMD as the feature of the faults, and employed the SVM optimized inter cluster distance in feature space for classification in [48], which resulted in 77.29% and 82.05% of the classification accuracy. And in [28], ELM is used for feature selection and parameter optimization, which carried out 93.04% classification accuracy.

From the result displayed above, in 10 types experiment, our method has a subtle advantage in classification accuracy. For the 24 types of experiments, the performance of the NB classifier used in the method proposed in this article was slightly lower than that of ELM, resulting in a slight difference of 2.58% in classification accuracy. Meanwhile the proposed scheme could bring a considerable improvement in time efficiency, which is 8 times faster than that of ELM and grey wolf optimization. In both 10 types and 24 types experiment our method has a shorter execution time, that is, the training time to build a classification model for the current bearing. Higher time efficiency can improve the efficiency of fault diagnosis and maintenance, and more efficient model construction time makes the method popularized for other different types or types of bearings in a shorter time. Since the data selected in the paper did not consider the load situation and different motor power was not included in the experiment. In the actual application process, whether the increase of sample data and the expansion of fault types have an impact on the method proposed in the paper, and to what extent, both need to be further explored. Further, the classifier selected in this paper is only used as a part of iterative optimization, and does not consider too much optimization of the classification performance of the classifier itself.

5. Conclusions

In this work, a high-efficiency bearing fault diagnosis method based on LMFO and NB is proposed. By introducing the EEMD method for feature extraction at the prepossessing of the data, the data noise is effectively reduced. In the feature selection process, the proposed LMFO based feature selection method can effectively reduce the feature dimension while ensuring or improving the classification accuracy of the classifier to a certain extent. Through comparative experiments based on the CWRU bearing fault data set, the results show that the proposed method greatly improves the training efficiency of the classification model under the premise of ensuring fault diagnosis accuracy, which makes the proposed method more versatile and practical in solving fault diagnosis problems.

References

1. Antoniadou I, Manson G, Staszewski W, Barszcz T, Worden K. A time-frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions. Mechanical Systems and Signal Processing 2015; 64-65: 188-216, https://doi.org/10.1016/j.ymssp.2015.03.003.

2. Awadallah M, Morcos M. Application of AI tools in fault diagnosis of electrical machines and drives-an overview. IEEE Transactions on Energy Conversion 2003; 18: 245-251, https://doi.org/10.1109/TEC.2003.811739.

3. Chudzik A, Warda B. Fatigue life prediction of a radial cylindrical roller bearing subjected to a combined load using FEM. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2020; 22(2): 212-220, https://doi.org/10.17531/ein.2020.2.4.

4. Dorigo M, Maniezzo V, Colorni A. Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) 1996; 26: 29-41, https://doi.org/10.1109/3477.484436.

5. Eberhart R, Kennedy J. A new optimizer using particle swarm theory. MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science; IEEE: Nagoya, Japan 1995; 39-43, https://doi.org/10.1109/MHS.1995.494215.

6. Feng Z, Liang M, Zhang Y, Hou S. Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation. Renewable Energy 2012; 47: 112-126, https://doi.org/10.1016/j.renene.2012.04.019.

7. Glowacz A. Recognition of acoustic signals of induction motor using fft, smofs-10 and lsvm. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(4): 569-574, https://doi.org/10.17531/ein.2015.4.12.

8. Goldberg D E, Holland J H. Genetic Algorithms and Machine Learning. Machine Learning 1988; 3: 95-99, https://doi.org/10.1023/A:1022602019183.

9. Guo W, Tse P W, Djordjevich A. Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. Measurement 2012; 45: 1308-1322, https://doi.org/10.1016/j.measurement.2012.01.001.

10. Holland J H. Adaptation in Natural and Artificial Systems; MIT press 1992, https://doi.org/10.7551/mitpress/1090.001.0001.

11. Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C, Liu H H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 1998; 454: 903-995, https://doi.org/10.1098/rspa.1998.0193.

12. Jiang F, Zhu Z, Li W, Chen G., Zhou G. Robust condition monitoring and fault diagnosis of rolling element bearings using improved EEMD and statistical features. Measurement Science and Technology 2014; 25: 025003, https://doi.org/10.1088/0957-0233/25/2/025003.

13. Jiang F, Zhu Z, Li W, Zhou G, Chen G. Fault identification of rotor-bearing system based on ensemble empirical mode decomposition and self-zero space projection analysis. Journal of Sound and Vibration 2014; 333: 3321-3331, https://doi.org/10.1016/j.jsv.2014.03.014.

14. Jun S, Kochan O, Kochan R. Thermocouples with Built-In Self-testing. International Journal of Thermophysics 2016; 37: 37, https://doi.org/10.1007/s10765-016-2044-2.

15. Jun S, Kochan O. Common mode noise rejection in measuring channels. Instruments and Experimental Techniques 2015; 58: 86-89, https://doi.org/10.1134/S0020441215010091.

16. Kennedy J. Swarm Intelligence. In Handbook of Nature-Inspired and Innovative Computing; Zomaya, A.Y., Ed.; Kluwer Academic Publishers: Boston 2006; 187-219, https://doi.org/10.1007/0-387-27705-6_6.

17. Kennedy J. The particle swarm: social adaptation of knowledge. Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC '97) 1997; 303-308, https://doi.org/10.1109/ICEC.1997.592326.

18. Kennedy J, Eberhart R. Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks 1995; vol. 4: 1942-1948, https://doi.org/10.1109/ICNN.1995.488968.

19. Kosicka E, Kozłowski E, Mazurkiewicz D. Intelligent Systems of Forecasting the Failure of Machinery Park and Supporting Fulfilment of Orders of Spare Parts. In International Conference on Intelligent Systems in Production Engineering and Maintenance. Springer, Cham. 2017; 54-63, https://doi.org/10.1007/978-3-319-64465-3_6.

20. Kotary D K, Nanda S J. (2020). Distributed robust data clustering in wireless sensor networks using diffusion moth flame optimization. Engineering Applications of Artificial Intelligence 2020; 87: 103342, https://doi.org/10.1016/j.engappai.2019.103342.

21. Kozłowski E, Mazurkiewicz D, Żabiński T, Prucnal S, Sęp J. Assessment model of cutting tool condition for real-time supervision system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(4): 679-685, https://doi.org/10.17531/ein.2019.4.18.

22. Lei Y, He Z, Zi Y, Hu Q. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. Mechanical Systems and Signal Processing 2007; 21: 2280-2294, https://doi.org/10.1016/j.ymssp.2006.11.003.

23. Lei Y, He Z, Zi Y. Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing 2009; 23: 1327-1338, https://doi.org/10.1016/j.ymssp.2008.11.005.

24. Leturiondo U, Salgado O, Galar D. Multi-body modelling of rolling element bearings and performance evaluation with localised damage. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 638-648, https://doi.org/10.17531/ein.2016.4.20.

25. Li C, Li S, Liu Y. A least squares support vector machine model optimized by moth-flame optimization algorithm for annual power load forecasting. Applied Intelligence 2016; 45: 1166-1178, https://doi.org/10.1007/s10489-016-0810-2.

26. Li Q, Chen H, Huang H, Zhao X, Cai Z, Tong C, Liu W, Tian X. An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis. Computational and Mathematical Methods in Medicine 2017; 1-15, https://doi.org/10.1155/2017/9512741.

27. Liu Y, Wang G, Chen H, Dong H, Zhu X, Wang S. An improved particle swarm optimization for feature selection. Journal of Bionic Engineering 2011; 8: 191-200, https://doi.org/10.1016/S1672-6529(11)60020-6.

28. Luo M, Li C, Zhang X, Li R, An X. Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings. ISA Transactions 2016; 65: 556-566, https://doi.org/10.1016/j.isatra.2016.08.022.

29. Mafarja M M, Mirjalili S. Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. Neurocomputing 2017; 260: 302-312, https://doi.org/10.1016/j.neucom.2017.04.053.

30. Maior C B S, Chagas Moura M, Lins I D. Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(4): 610-619, https://doi.org/10.17531/ein.2019.4.10.

31. Mei R N S, Sulaiman M H, Mustaffa Z, Daniyal H. Optimal reactive power dispatch solution by loss minimization using moth-flame optimization technique. Applied Soft Computing 2017; 59: 210-222, https://doi.org/10.1016/j.asoc.2017.05.057.

32. Mirjalili S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. Knowledge-Based Systems 2015; 89: 228-249, https://doi.org/10.1016/j.knosys.2015.07.006.

33. Nematollahi A F, Rahiminejad A, Vahidi B. A novel physical based meta-heuristic optimization method known as Lightning Attachment Procedure Optimization. Applied Soft Computing 2017; 59: 596-621, https://doi.org/10.1016/j.asoc.2017.06.033.

34. Przystupa K, Ambrożkiewicz B, Litak G. Diagnostics of Transient States in Hydraulic Pump System with Short Time Fourier Transform. Advances in Science and Technology. Research Journal, 2020; 14(1): 178-183, https://doi.org/10.12913/22998624/116971.

35. Rashedi E, Nezamabadi-pour H, Saryazdi S. GSA: A Gravitational Search Algorithm. Information Sciences 2009; 179: 2232-2248, https://doi.org/10.1016/j.ins.2009.03.004.

36. Rodrigues D, Pereira L A, Nakamura R Y, Costa K A, Yang X S, Souza A N, Papa J P. A wrapper approach for feature selection based on Bat Algorithm and Optimum-Path Forest. Expert Systems with Applications 2014; 41: 2250-2258, https://doi.org/10.1016/j.eswa.2013.09.023.

37. Rubini R, Meneghetti U. Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings. Mechanical Systems and Signal Processing 2001; 15: 287-302, https://doi.org/10.1006/mssp.2000.1330.

38. Sharawi M, Zawbaa H M, Emary E, Zawbaa H M, Emary E. Feature selection approach based on whale optimization algorithm. 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI); IEEE: Doha, Qatar 2017; 163-168, https://doi.org/10.1109/ICACI.2017.7974502.

39. Stahczyk U. Feature Evaluation by Filter, Wrapper, and Embedded Approaches. In Feature Selection for Data and Pattern Recognition; Stahczyk, U.; Jain, L.C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg 2015; 584: 29-44, https://doi.org/10.1007/978-3-662-45620-0_3.

40. Storn R, Price K. Minimizing the real functions of the ICEC'96 contest by differential evolution. Proceedings of IEEE International Conference on Evolutionary Computation; IEEE: Nagoya, Japan 1996; 842-844, https://doi.org/10.1109/ICEC.1996.542711.

41. Sulaiman M H, Mustaffa Z, Aliman O, Daniyal H, Mohamed M R. Application of moth-flame optimization algorithm for solving optimal reactive power dispatch problem. 4th IET Clean Energy and Technology Conference (CEAT 2016) 2016, https://doi.org/10.1049/cp.2016.1273.

42. Wan Y, Wang M, Ye Z, Lai X. A feature selection method based on modified binary coded ant colony optimization algorithm. Applied Soft Computing 2016; 49: 248-258, https://doi.org/10.1016/j.asoc.2016.08.011.

43. Wang X, Yang J, Teng X, Xia W, Jensen R. Feature selection based on rough sets and particle swarm optimization. Pattern Recognition Letters 2007; 28: 459-471, https://doi.org/10.1016/j.patrec.2006.09.003.

44. Wolpert D H. The lack of a priori distinctions between learning algorithms. Neural Computation 1996; 8: 1341-1390, https://doi.org/10.1162/neco.1996.8.7.1341.

45. Zawbaa H M, Emary E, Grosan C, Feature Selection via Chaotic Antlion Optimization. PLOS ONE 2016; 11: e0150652, https://doi.org/10.1371/journal.pone.0150652.

46. Zhang L, Xiong G, Liu H, Zou H, Guo W. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. Expert Systems with Applications 2010; 37: 6077-6085, https://doi.org/10.1016/j.eswa.2010.02.118.

47. Zhang W, Jia M P, Zhu L, Yan X A. Comprehensive Overview on Computational Intelligence Techniques for Machinery Condition Monitoring and Fault Diagnosis. Chinese Journal of Mechanical Engineering 2017; 30: 782-795, https://doi.org/10.1007/s10033-017-0150-0.

48. Zhang X, Zhou J. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. Mechanical Systems and Signal Processing 2013; 41: 127-140, https://doi.org/10.1016/j.ymssp.2013.07.006.

49. Zhang Y, Zuo H, Bai F. Classification of fault location and performance degradation of a roller bearing. Measurement 2013; 46: 1178-1189, https://doi.org/10.1016/j.measurement.2012.11.025.

50. Zio E. Reliability engineering: Old problems and new challenges. Reliability Engineering & System Safety 2009; 94: 125-141, https://doi.org/10.1016/j.ress.2008.06.002.