

# A fault diagnosis model based on singular value manifold features, optimized SVMs and multi-sensor information fusion

Zuqiang Su<sup>a,\*</sup>, Fuli Wang<sup>b</sup>, Hong Xiao<sup>a</sup>, Hong Yu<sup>c</sup>, Shaojiang Dong<sup>d</sup>

- a. School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
- b. School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom*
- c. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
- d. School of Mechatronics and Automotive Engineering, Chongqing Jiaotong University, Chongqing 400074, China*

**Abstract**—To achieve better fault diagnosis of rotating machinery, this paper presents a novel intelligent fault diagnosis model based on singular value manifold features (SVMF), optimized support vector machine (SVMs) and multi-sensor information fusion. Firstly, a new fault feature denoted as SVMF is developed to better represent faults. SVMF is acquired by extracting manifold topology features of the singular spectrum. Compared with frequently-used fault features, the feature scale of SVMF is constant under variable rotating speed, and the extraction process of SVMF also has the effect of self-weighting. So SVMF has a better representation of faults. Then, to select optimal parameters for model training of SVMs, an improved fruit fly algorithm is proposed by introducing guidance search mechanism and enhanced local search operation, and as a result, both the convergence speed and accuracy are improved. At last, the Dempster-Shafer evidence theory is introduced to fuse decision-level information of SVMs models of multiple sensors. By information fusion, the conflict of fault diagnosis conclusions from multi-sensor is eliminated, which lead to high robustness and accuracy of fault diagnosis model. As a summary, the proposed method combines the advantages of SVMF in fault representation, SVMs in fault identification and the Dempster-Shafer evidence theory in information fusion, and as a result the proposed method will get better fault diagnosis performance. The proposed intelligent fault diagnosis model is subsequently applied to the fault diagnosis of the gearbox. Experimental results show that the proposed fault diagnostic framework is versatile at detecting faults accurately.

**Keywords**—Gearbox fault, fault diagnosis, singular spectrum, support vector machine, fruit fly algorithm, Dempster-Shafer evidence theory

## 1. INTRODUCTION

Mechanical equipment is fundamental to modern industry. With the growing complexity and scale of industrial equipment, the difficulty of equipment maintenance and repairment increases sharply, and the economic losses caused by mechanical equipment failure become more unbearable. Therefore, the industry has a strong demand for fault diagnosis technology, especially for intelligent fault diagnosis technology. Besides, with the popularity of unmanned factories and the demand of manufacturing industry of transforming to service-oriented manufacturing from traditional manufacturing, it can be expected that intelligent operation and maintenance technology will become one of the key technologies in the future.

Driven by industrial requirements, increasing attention has been paid on the research of fault diagnosis in recent years, including model-based simulation [1], signal processing-based fault diagnosis [2] and data-driven fault diagnosis [3]. Model-based fault diagnosis and signal processing-based fault diagnosis require high professional knowledge, and the results of fault diagnosis largely depend on the experience of experts. So it is difficult to meet the needs of modern industry for intelligent

1 fault diagnosis and intelligent maintenance. Data-driven fault diagnosis combines feature extraction and artificial intelligence  
2 algorithms to overcome the shortcomings of the previous two methods. With the help of artificial intelligence algorithms and  
3 decision fusion mechanisms, fault diagnosis rules are learned from operational data to establish a diagnostic model or expert  
4 system with knowledge inference capacity to achieve an automatic identification of faults. The main advantage of data-driven  
5 fault diagnosis method is that it weakens the requirement of expert knowledge and makes this kind of method more suitable for  
6 industrial practice.  
7  
8  
9

10 Generally, data-driven fault diagnosis approaches can be summarized into two steps: feature extraction and fault classification.  
11 Extracting sensitive features for further fault classification is the key step. Several kinds of features have been introduced or  
12 proposed for fault diagnosis, such as frequency spectrum [4], time-frequency statistical features [5], ARMA model coefficients  
13 [6], information entropy features [7], wavelet packet analysis [8], and so on. These feature extraction methods perform well  
14 under constant operation condition. However, in practice, there are large fluctuations in the operation condition, especially the  
15 rotating speed of machinery, such as wind turbines. In situation of variable rotating speed, the value of the fault feature varies  
16 greatly as the frequently-used feature extraction algorithms are not self-adaptive, and as a result poor fault diagnosis performance  
17 will be derived [9]. In this work, a new kind of fault feature, named as singular value manifold features (SVMF), is proposed.  
18 SVMF is acquired by extracting manifold topology features of the singular spectrum. The singular spectrum is extracted by  
19 Singular value decomposition (SVD), and SVD is essentially a matrix decomposition method, which is often used in signal  
20 processing to decompose signals into multiple subspaces to achieve noise reduction or feature extraction [10]. SVD is  
21 self-adaptive and needs no basis functions [11]. At present, SVD has been widely used in the field of mechanical fault diagnosis,  
22 such as combining SVD with kurtosis index to achieve signal de-noising [12] and using SVD to decompose the coefficient  
23 matrix of continuous wavelet decomposition to enhance the impulse component in fault vibration signal [13], etc. SVD can  
24 self-adaptively separate the components in the signal, and the singular spectrum obtained by SVD reflects the energy distribution  
25 of the signal. Therefore, there are also studies on the combination of the singular spectrum and pattern recognition algorithm for  
26 intelligent fault diagnosis [14-15]. However, the variation range of the scale of singular value is very large. When fault diagnosis  
27 is conducted directly by singular spectrum, the information contained in the small singular values is weakened by large singular  
28 values, so the fault information contained in the singular spectrum cannot be fully utilized. To address the above concerns, this  
29 paper proposes a new kind of fault features denoted as SVMF for fault diagnosis, and the proposed features are acquired by  
30 extracting the manifold topology features of the singular spectrum. The distribution changes of signals collected under different  
31 fault states can be directly reflected as the distribution changes of the singular spectrum, and manifold topology can directly  
32 represent the changes of the singular spectrum, thus reflecting the changes in equipment operation status. At the same time, a  
33 variable scale manner is adopted to carry out feature extraction of SVMF according to the distribution characteristics of the  
34 singular spectrum. Therefore, the feature extraction process of SVMF has the effect of self-weighting, which can increase the  
35 ability of features to characterize faults. Compared with other fault features, the advantages of the proposed SVMF includes three  
36 aspects: 1) The fault information contained in the singular spectrum can be fully utilized to eliminate the influence of the singular  
37 values on the fault feature extraction; 2) The range of SVMF is [0,1], and that is to say, the numerical scale of SVMF extracted  
38 under different rotating speeds is the same, which is more suitable for fault diagnosis under variable speed conditions; 3) The  
39 feature extraction process has the effect of self-weighting, which enhances the ability to characterize faults.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53 After feature extraction, a classification model with high efficiency and robustness is needed to identify fault types. Until now,  
54 several algorithms have been applied for fault identification, such as deep learning method [16], Hidden Markov model (HMM)  
55 [17], Artificial Neural Network (ANN) [18], K-Nearest Neighbor (KNN) [19], Naïve Bayes [20], Support Vector Machine  
56 (SVMs) [21] and so on. As big data analysis method, deep learning needs massive data for model training. hidden Markov  
57  
58  
59  
60

1 model, k-nearest neighbour and the artificial neural network will also suffer over-fitting without enough training samples, and  
2 artificial neural network even easily gets trapped in local optimum owing to empirical risk minimization principle. However, it  
3 would be almost impossible to get a large number of training samples in industrial scenarios and thus fault diagnosis results of  
4 these methods are not satisfactory with small size samples. Compared with formerly mentioned methods, researchers have  
5 reported that SVMs can achieve better correct classification [22-23]. That is because SVMs based on structural risk minimization  
6 principle can minimize an upper bound on the expected risk and implements classification by using a separating hyperplane  
7 determined by a few support vectors. Therefore, SVMs are less prone to the problem of over-fitting, local optimal solution  
8 problems faced than other methods, and have better generalization capability for small sample size problem. However, the  
9 performance of SVMs depend largely on the selection of parameters, but unfortunately, there is still no guiding theory for the  
10 parameter selection until now. That is because the SVMs contain two kinds of parameters with a wide range of values, but these  
11 two kinds of parameters are not differentiable. The forced calculation of gradient is approximate, and the effect of optimization  
12 stability is also not good. Some scholars have tried the trial method, grid method and gradient descent method to do this work  
13 [24][25]. However, the trial method lacks theoretical guidance and relies on experience, which results in that the final parameters  
14 are not necessarily optimal. The basic principles of grid parameter optimization algorithm are as follows: firstly, every point in  
15 the specified grid range is traversed, and then every point is transformed into SVMs parameters for verification, and finally the  
16 grid point with the smallest error is selected as the optimal parameter of SVMs. So grid parameter optimization is very  
17 time-consuming. Gradient descent algorithm is very sensitive to the selection of initial value, and sometimes the error of  
18 experimental results is very large. Therefore, these algorithms have been difficult to meet the needs of the actual application. At  
19 present, intelligent optimization algorithms are the most often applied methods for parameter selection, such as particle swarm  
20 optimization (PSO) [26] and genetic algorithm (GA) [27]. The fruit fly algorithm, proposed by Wen-Tsao Pan [28], is a new  
21 intelligent optimization algorithm. Compared with other algorithms, the fruit fly algorithm is easy to implement and has low  
22 computational complexity. However, the imperfect search mechanism and poor local search ability prevent the further  
23 performance improvement of the fruit fly algorithm. In this work, an improved fruit fly algorithm (IFFA) is proposed by  
24 introducing guidance search mechanism and enhanced local search operation. Guidance search mechanism ensures that the  
25 search direction approximates the maximum gradient direction, and which improves search efficiency. Enhanced local search  
26 operation guarantees optimal search location during search iteration, as a result, the search accuracy is improved. Consequently,  
27 optimal parameters selected by IFFA for SVMs greatly improve the fault diagnosis performance.

40 For fault diagnosis, more fault information leads to high reliability and accuracy of fault diagnosis. Compared to a single  
41 sensor installed on the shell of the machine, more abundant fault information is collected by multiple sensors installed in several  
42 positions. Therefore, the fusing of multiple sensors information for fault diagnosis can obtain better fault diagnosis results.  
43 However, due to noise and interference during signal transmission, the fault information acquired by sensors is somehow random  
44 and ambiguous, which not only leads to the uncertainty of fault diagnosis results from the single signal source but also leads to  
45 conflicting conclusions from multiple sensors. D-S evidence theory describes uncertainties by synthesizing knowledge from  
46 multiple data sources and has a strong ability to deal with uncertain information. Without prior probability, the D-S theory  
47 obtains the result of multi-source information fusion through certain reasoning. At present, D-S evidence theory has been  
48 introduced into many application fields to realize multi-sensor information fusion for decision-making [29][31]. Therefore, this  
49 paper constructs several optimized SVM fault diagnosis models through multiple sensors information, and then combines D-S  
50 evidence theory to fusing information of multiple sensors, to improve the robustness and identification accuracy of the fault  
51 diagnosis model. In another word, combining the advantages of SVMF in fault representation, SVMs in fault identification and  
52 D-S in information fusion, is the main innovation of this paper, and hence the proposed method will get better fault diagnosis  
53  
54  
55  
56  
57  
58  
59  
60

performance. The contributions of this paper can be summarized as follows:

1) A new kind of fault features denoted as SVMF is proposed for fault diagnosis. SVMF contains more abundant fault information, of which the feature scale does not change with the operation speed and the feature extraction process has the effect of self-weighting. As a result, the new feature characterizes faults well.

2) The IFFA is proposed by introducing guidance search mechanism and enhanced local search operation to train the SVMs fault diagnosis model. The proposed IFFA algorithm improves both search efficiency and search accuracy.

3) D-S evidence theory is used to fuse decision-level information of SVMs fault diagnosis models constructed by multiple sensors, which eliminates the conflict of fault diagnosis conclusions from multiple sources and increases the robustness and accuracy of fault diagnosis model.

4) The effectiveness of the proposed method is verified by the gearbox fault diagnosis experiment. The experimental results show the effectiveness of the proposed method.

The rest of this paper is organized as follows. In section 2, the feature extraction method of SVMF, the SVMs fault diagnosis model based on IFFA algorithm and the related theory of D-S evidence for multi-sensor information fusion are introduced in detail; the flow and framework of the proposed fault diagnosis method are described in the third section; in section 4, the effectiveness of the proposed algorithm will be verified by gearbox fault data; the fifth section concludes the article.

## 2. THE PROPOSED METHOD OF FAULT DIAGNOSIS

### 2.1 Feature Extraction Method of Singular Spectrum Manifolds

#### 2.1.1 Singular Value Decomposition

As a matrix decomposition method, SVD cannot directly deal with one-dimensional vibration signals. Therefore, one-dimensional vibration signals are usually reconstructed into high-dimensional matrices for processing, such as Toeplitz matrix, periodic matrix, Hankel matrix and so on. Among them, the Hankel matrix is the most widely used and the construction method is the simplest [11]. Given a vibration signal  $\mathbf{X} = [x_1, x_2, \dots, x_N]$ , the construction method of the Hankel matrix is as follows:

$$\mathbf{A} = \begin{bmatrix} x_1 & x_2 & \dots & x_q \\ x_2 & x_3 & \dots & x_{q+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_p & x_{p+1} & \dots & x_N \end{bmatrix} \quad (1)$$

Where  $1 < p, q < N$ ,  $p$  is the reconstruction dimension,  $q = N - p + 1$  is the number of samples in the Hankel matrix,  $N$  is the length of the vibration signal. The definition of SVD for the matrix  $\mathbf{A} \in R^{p \times q}$  is as follows:

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad (2)$$

In formula (2),  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p] \in R^{p \times p}$  and  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p] \in R^{q \times q}$  are two orthogonal decomposition matrices,  $\mathbf{\Sigma} = [\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \mathbf{O}] \in R^{p \times q}$  is a singular value matrix,  $n = \min(p, q)$  is the number of singular values,  $\lambda_1 > \lambda_2 > \dots > \lambda_n$  are singular values,  $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  is a diagonal matrix,  $\mathbf{O}$  is a zero value matrix,  $\mathbf{u}_i, i = 1, L, p$  and  $\mathbf{v}_j, j = 1, L, q$  are respectively the left and right singular vectors. Fig. 1 is the singular value extracted from several gearbox fault signals [32]. It can be seen that the singular spectrum obtained by SVD can represent the distribution of the signal. The distribution of the singular spectrum corresponding to different faults and its manifold topological structure is different, so the singular spectrum is usually used as fault feature for mechanical fault diagnosis. However, the variation range of scales of singular values is very large, and therefore the fault information contained in small singular values is easily submerged by large singular values. As a result, only a few relative larger singular values play a role in fault diagnosis, and the fault information contained in the singular

spectrum cannot be fully utilized.

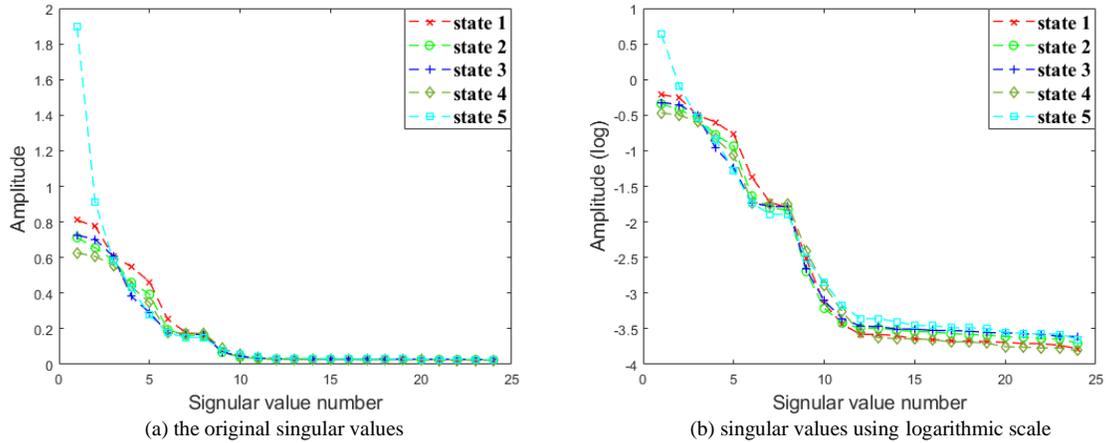


Fig.1 The singular spectrum of gearbox fault signals

### 2.1.2 Analysis of Singular value Manifolds

Fig. 1 shows that the distribution of singular spectrums for different faults is different from each other, which can be directly reflected by the difference of manifold topological structure of singular spectrum. Since the singular value manifolds are one-dimensional, this paper proposes to use the slope of the singular spectrum to extract the topological features of singular spectrum manifolds, namely SVMF. For the singular spectrum  $SVs = [\lambda_1, \lambda_2, \dots, \lambda_n]$ , the extraction method of the proposed SVMF based on the slope of the singular spectrum is as follows:

$$\beta_i = \frac{\lambda_i - \lambda_{i+1}}{k_i}, n > i > 1 \quad (3)$$

According to the definition of formula (3), for a singular spectrum  $SVs, n-1$  slope values,  $SLs = [\beta_1, \beta_2, \dots, \beta_n]$ , of the singular spectrum can be calculated.  $SLs$  is the proposed SVMF and can effectively characterize the manifold topological structure of the singular spectrum. When calculating the slope of the singular spectrum  $\beta_i$ , the value of denominator  $k_i$  of formula (3) has a great influence on the calculation results. As shown in Fig. 2 below, when  $\lambda_i - \lambda_{i+1} = \lambda_j - \lambda_{j+1} = \Delta$ , if  $k_i = k_j$  is taken in calculating the slope, the slope of corresponding  $\lambda_i$  and  $\lambda_j$  is  $\beta_i = \beta_j$ , where  $n > i, j > 1$ . However, the relative variation of the singular spectrum at  $\lambda_j$  is obviously greater than that at  $\lambda_i$ . That is to say, the SVMF should have bigger curvature at  $\lambda_j$ , so  $k_i > k_j$  should be satisfied when calculating the slope of the singular spectrum. In this paper, a variable scale method is proposed to calculate the slope of the singular spectrum. In the calculation process,  $k_i$  is taken as  $k_i = \lambda_i$ . That is to say the proposed SVMF corresponds to the relative differences between consecutive singular values. By introducing  $k_i = \lambda_i$  into formula (3), it can be seen that the range of feature values in SVMF is  $[0, 1]$ . That is to say, the numerical scales of the feature of SVMF extracted under different operational conditions are the same, so SVMF is more suitable for fault diagnosis under variable rotating speeds. At the same time,  $k_i = \lambda_i$  is set in the process of feature extraction, which is similar to self-weighting the feature according to the characteristics of the singular spectrum itself in the process of fault feature extraction, so that the proposed SVMF can more effectively characterize the faults. Fig.3 shows SVMF extracted from gearbox fault signals. Comparing Fig.3 with Fig.1, it can be seen that the fault information in small singular values is also available, and therefore fault information contained in the singular spectrum is fully utilized.

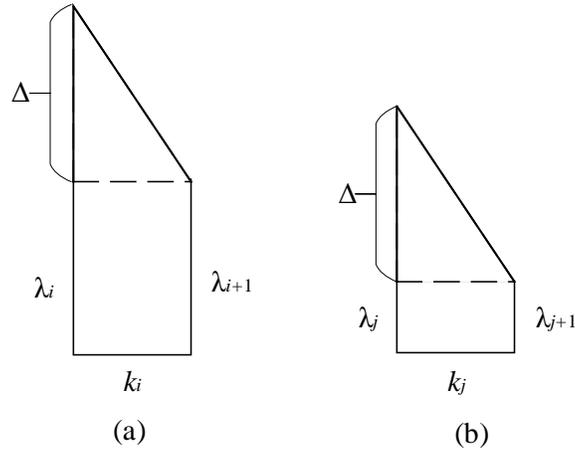


Fig.2 Variable scale extraction of singular value spectrum slope

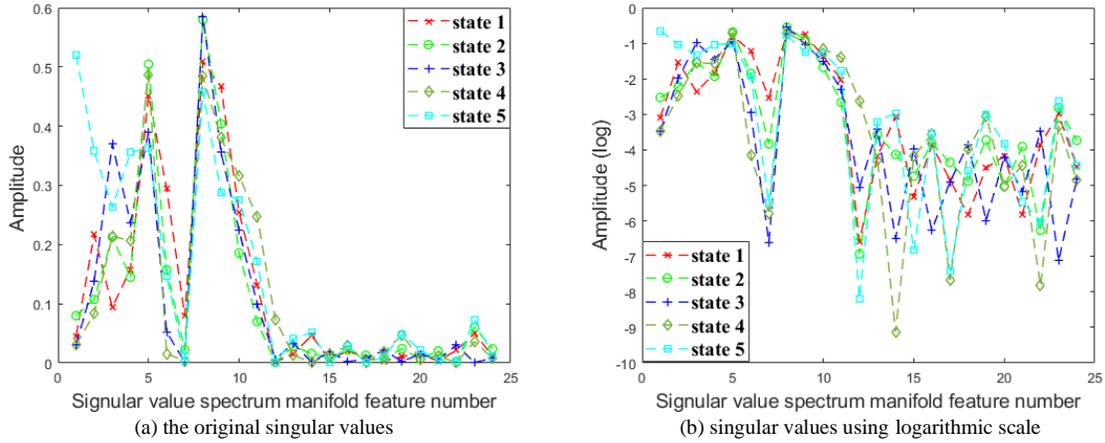


Fig.3 The singular spectrum manifold features of gearbox fault signals

## 2.2 Fruit Fly Algorithm for Parameters Optimization of SVMs

### 2.2.1 Theory of Support vector machine

SVMs is a pattern recognition algorithm based on structural risk minimization. By establishing an optimal decision hyperplane, the distance between inter-class samples nearest to the plane is maximized, thus providing good generalization ability for classification problems. Its model is usually described as follows:

$$\begin{aligned}
 f(\mathbf{x}_i, \mathbf{x}_j) &= \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \\
 \min & \left( \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \varepsilon_i \right) \\
 \text{s.t. } & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) - 1 + \varepsilon_i \geq 0; \\
 & \forall_i = 1, 2, \dots, N
 \end{aligned} \tag{4}$$

Where  $\mathbf{w}$  denotes the normal vector of the hyperplane,  $\varphi(\cdot)$  is the pre-defined function of  $\mathbf{x}_i$ ,  $C$  denotes the penalty factor,  $\varepsilon_i$  denotes the relaxation variable and  $b$  denotes the displacement or the distance to the hyperplane.  $f(\cdot, \cdot)$  is the kernel function which is to map the samples from the original space to a high-dimensional feature space, so that the samples can be linearly separable in the high-dimensional space. At present, the most widely used kernel function is the radial basis function (RBF), which is expressed as follows:

$$f(\mathbf{x}_i, \mathbf{x}_j) = \exp \left( -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2g^2} \right) \tag{5}$$

Where  $g$  denotes the parameter of RBF,  $\mathbf{x}_i$  and  $\mathbf{x}_j$  denote the sample points.

### 2.2.2 Optimized Support vector machine

From the above mathematical model, it can be seen that there are usually two kinds of parameters involved in SVMs, namely penalty factor and kernel parameter. However, there is currently no complete set of theories for the value setting of these two parameters, which are usually tried out through a large number of experiments. In recent years, to improve the efficiency of parameters' selection, many studies use an intelligent algorithm to search for the best value. However, the parameters of SVMs have a wide range of values, which puts forward high requirements for the performance of an intelligent algorithm. This requires that the parameters searched by the algorithm can maximize the performance of SVMs, and also requires that the algorithm has high computational efficiency. To meet the above requirements, an improved fruit fly algorithm (IFFA) is proposed to optimize the parameters of SVMs. Compared with other intelligent algorithms, the advantage of fruit fly algorithm (FFA) is that its algorithm is easy to implement, easy to convert its theoretical ideas into program code and easy to understand. Besides, the time complexity of the algorithm is not high, so it has certain advantages in computational efficiencies.

Although FFA has some advantages, its search mechanism is too simple compared with other algorithms, resulting in poor search ability. In order to improve the search efficiency and accuracy of FFA, guidance search mechanism and enhanced local search ability are introduced into FFA in this paper.

Firstly, this paper introduces guidance search mechanism [33] to improve FFA. To make fruit flies search adaptively, the formula (6) is utilized to update the position of fruit flies so that the search space is adjusted according to the best individual of the last generation:

$$\begin{aligned} X_i &= BestX + RandomValue \times A_1^k \\ Y_i &= BestY + RandomValue \times A_2^k \end{aligned} \quad (6)$$

The position of each fruit fly  $fly_i$  can be represented by a vector  $(X_i, Y_i)$ . In the parameter optimization applied to the SVM,  $X_i$  and  $Y_i$  represent the parameters  $C$  and  $g$  of the SVM respectively. Therefore, in the  $k$ -th iteration,  $A_1^k$  can regulate the search range of penalty factor  $C$ , and  $A_2^k$  can regulate the search range of the parameter  $g$ . The expression of  $A_j^k$  is as follows:

$$A_j^k = \begin{cases} A_j^1, & k=1 \\ C_r A_j^{k-1}, & Smell^k \leq Smell^{k-1} \\ C_a A_j^{k-1}, & Smell^k > Smell^{k-1} \end{cases} \quad (7)$$

In the first generation, the  $A_j^1$  is preset to a constant number which is usually the diameter of the search space.  $Smell^k$  is the current fitness value, and the larger the value is, the better the solution will be. If no better solution is found in iteration, the search space needs to be reduced to improve the possibility of searching for high-quality solutions. If a better solution is found, the search space is increased to explore other locations within the feasible region. Formula (7) reduces the search space by a coefficient  $C_r$  ( $C_r < 1$ ) and enlarges the search space by  $C_a$  ( $C_a > 1$ ).

Furthermore, to make better use of the guidance search mechanism, this paper designs a guiding fruit fly operation, which is described as Algorithm 1.

---

#### Algorithm 1 Generating the guiding fruit fly

---

**Require:**  $(BestX, BestY)$ ,  $fly_i$ ,  $\mu$  and  $\sigma$

Sort the fliers by their fitness values  $Smell(fly_i)$  in ascending order

$$\Delta_k \leftarrow \frac{1}{\sigma\mu} \left( \sum_{i=1}^{\sigma\mu} fly_i - \sum_{i=\mu-\sigma\mu+1}^{\mu} fly_i \right);$$

$$G_k = (BestX, BestY) + \Delta_k;$$

**return**  $G_k$

---

---

**Algorithm 2** Generating the local fruit fly
 

---

**Require:**  $(BestX, BestY)$ ,  $\delta$  and  $r$

```

if rand <  $r$ 
     $local\_X_k = BestX + \delta \times randn$ ;
     $local\_Y_k = BestY + \delta \times randn$ ;
     $L_k = (local\_X_k, local\_Y_k)$ ;
end if
return  $L_k$ 

```

---

In Algorithm 1,  $\sigma$  ranges from 0 to 1, usually 0.2 [33]. The guidance search mechanism is similar to finding the maximum gradient direction ( $\Delta_k$ ). In each generation, the operation yields the current maximum gradient descent direction, which generates a guiding fruit fly  $G_k$ . This operation can make the best use of the search information of fruit flies, thus effectively improving the efficiency of the algorithm.

Besides, to further improve the local search ability and search accuracy of the algorithm, this paper introduces the local search operation to enhance the local search ability of the algorithm [34], which is described in detail as shown in Algorithm 2. In this algorithm,  $\delta$  is a small constant which represents the local search range and  $r$  is the incremental factor. The value of  $r$  increases with the advance of computing time. Its concrete expression is as follows:

$$r^{t+1} = r^0 [1 - \exp(-\gamma t)] \quad (8)$$

Where  $\gamma$  is a constant, usually 0.9. The initial value  $r^0$  of the incremental factor can be selected in the interval  $[0,1]$ . The local search operation is a dynamic adjustment process. In the early stage of the IFFA, the algorithm still needs to expand the search space and carry out global optimization, because the position of the high-quality solution has not been searched. At this stage, the incremental factor is small, and the probability of triggering local search is also small. With the continuous operation of the algorithm, high-quality solutions appear one after another. At this stage, local optimization should be carried out near the high-quality solution to avoid missing the optimal solution. Therefore, the incremental factor also increases, which triggers local search operation and generates the corresponding local fruit fly  $L_k$ .

Finally, according to the above improvements, an improved fruit fly algorithm is proposed as shown in Algorithm 3. The IFFA not only enhances the information utilization ability but also enhances the search ability of the algorithm, which improves the efficiency and accuracy of the algorithm.

---

**Algorithm 3** Improved fruit fly algorithm
 

---

Randomly initialize  $\mu$  fruit flies in the potential space.

Evaluate the flies' fitness and get the current best position:

**repeat**

**for**  $i = 1$  **to**  $\mu$  **do**

    Calculate the position of each fruit fly according to (6) and (7);

    Calculate the fitness value of fruit fly;

**end for**

    Generate the guiding fruit fly according to Alg. 2;

    Generate the local fruit fly according to Alg. 3;

    Evaluate the fitness of all fruit flies;

    Update the current best fitness value and best position;

**until** termination criteria is met

---

---

**return the position and the fitness of the best individual.**

---

### 2.3 Dempster–Shafer evidence theory

For the samples obtained from several measurement points, there are always conflicts between the fault diagnosis results of different measurement points. In order to effectively utilize the information of all measurement points, this paper uses D-S evidence theory to fuse the posterior probability of the output of each measurement point, so as to eliminate the conflict of multi-source signals and obtain a more reliable classification result [29][31]. D-S evidence theory is an effective reasoning method for uncertain information. Compared with traditional fusion decision-making method, it can grasp the unknown and uncertainty of sensor information better. Its basic theory is as follows.

(1) *Basic probability assignment (BPA)*: D-S evidence theory regards the possible solutions to a judgment problem and constitute a set, denoted as  $\Theta$ . Given some fault samples of  $k'(k' > 2)$  classes, the fault class labels constitute a set, namely the framework of identification  $\Theta$ . Furthermore, we can get a belief function of each type of fault  $m(\cdot)$  on the framework. The belief assignment function on  $\Theta$  is defined as  $m: \Omega(\theta) \rightarrow [0, 1]$ , and it satisfies:

$$(i) m(\emptyset) = 0; \quad (ii) \sum_{\theta \in \Theta} m(\theta) = 1 \quad (9)$$

Where  $m$  is the basic probability assignment,  $\theta$  is the proposition of a certain type of fault and  $m(\theta)$  is the basic probability value of  $\theta$ . In the formula (9), condition (i) denotes that there is no belief for empty propositions, and condition (ii) indicates that the sum of belief values for propositions of all faults under the framework of identification is 1. For every SVM classifier, the BPA of the  $s$ -th sensor is  $\{m_s^1, m_s^2, \dots, m_s^k\}$ .

(2) *Belief function*: The belief function of the proposition  $\theta$  is:

$$Bel(\theta) = \sum_{B \subseteq \theta} m(B), \quad \forall \theta \subseteq \Theta \quad (10)$$

It is also named as the lower bound, meaning the least probability that the proposition holds. If some evidence supports one proposition, then it should also support the inference of the proposition similarly.

(3) *Plausibility function*: The plausibility function of a proposition  $\theta$  is defined as follows.

$$Pls(\theta) = 1 - Bel(\bar{\theta}) = \sum_{B \mid \theta \neq \emptyset} m(B) \quad (11)$$

The plausibility function is also named as the upper bound, giving an uncertainty measure for the proposition holding.

(4) *The fusion rule*:

Let  $m_1^k, m_2^k, \dots, m_s^k$  denote different basic probability assignments derived from multiple sensors, then the Dempster fusion rule is as follows:

$$\begin{cases} m(\theta) = K \sum_{\theta_i = \theta} \prod_{j=1}^n m_j(\theta_i) & \theta \neq \emptyset \\ m(\emptyset) = 0 \end{cases} \quad (12)$$

Wherein  $K^{-1} = 1 - \sum_{\theta_i = \emptyset} \prod_{j=1}^n m_j(\theta_i)$ ,  $\theta \neq \emptyset$ . The value of  $K$  reflects the conflict of the evidence: the larger  $K$  means the

more serious conflict and the worse effect of fusion. Since  $D_i = \begin{bmatrix} m_1^1 & m_1^2 & \dots & m_1^k \\ m_2^1 & m_2^2 & \dots & m_2^k \\ \vdots & \vdots & \ddots & \vdots \\ m_n^1 & m_n^2 & \dots & m_n^k \end{bmatrix}$  is the probability assignment of sample  $i$  on

$n$  sensors,  $i = 1, 2, \dots, l$  then for  $l$  test samples, we have  $\{D_1, D_2, \dots, D_l\}$ . For any  $D_i$ , it is integrated based on the equation (12) after which we can obtain the probability of the sample falling into every class. The final result is the label with the highest probability.

### 3. THE FRAMEWORK OF THE PROPOSED FAULT DIAGNOSIS METHOD

The flow chart of the fault diagnosis method presented in this paper is shown in Fig. 4. Firstly, the vibration signals of multiple sensors located in several measuring points are collected. Subsequently, we can see that in the figure, the proposed SVMF is extracted from the collected vibration signals to construct the fault feature set for each measuring point. Then, the acquired fault feature sets are utilized to train and test the optimized SVMs models, while the parameters of SVMs are selected by IFFA, and the posterior probability of the test samples and the diagnosis results of each measuring points are obtained. In this figure, it is shown that the input data of each sensor is processed by SVMF and optimized SVMs, the classification result is output according to the posterior probability of each state. However, the output of each sensor may conflict with each other. Therefore, the posterior probabilities of multi-sensors are fused by using the D-S evidence theory to get the final diagnosis results in the last stage of this method.

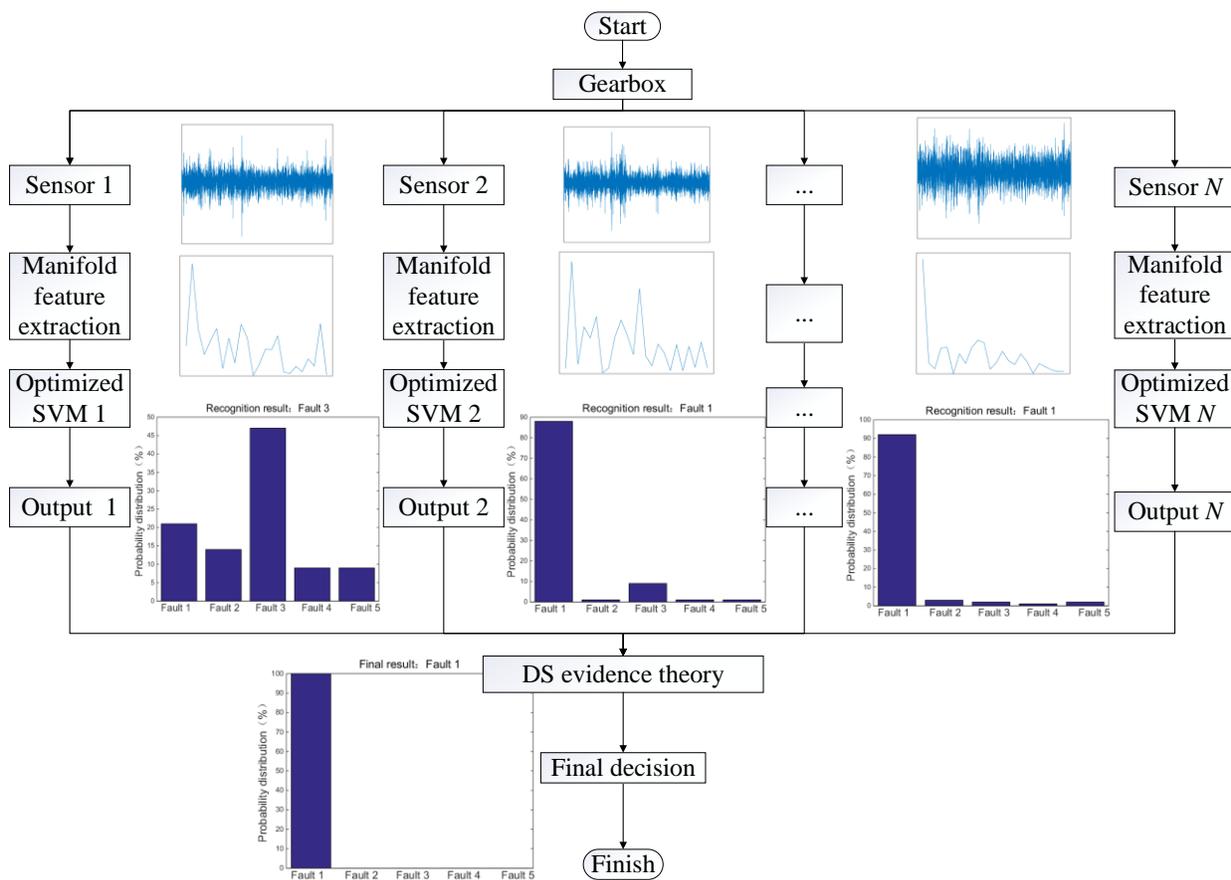


Fig.4 The fault diagnosis process of the proposed method

### 4. EXPERIMENTAL VERIFICATION

#### 4.1 Comparison of feature extraction methods

In this section, the proposed SVMF is compared with several existing fault features to verify the feasibility and superiority of the proposed method. The experimental data come from a gearbox, and Fig.5 shows the gearbox fault diagnosis test rig. The gearbox fault diagnosis test rig mainly consists of a speed control device, variable frequency AC drive motor, planetary gearbox, second stage parallel shaft gearbox, magnetic brake, etc. In this experiment, five kinds of gearbox faults are simulated: tooth root corrosion, tooth root crack, bearing inner ring fault, bearing outer ring fault, and mixed fault. For the above faults, the corresponding data were collected under the conditions of variable rotating speed and constant rotating speed. The rotating speed

of driving motor linearly increases from 20-40Hz under the variable rotating speed conditions, while the rotating speed of driving motor is 20Hz under the condition of constant rotating speed. The vibration data were collected at a sampling frequency of 12.8kHz for the conditions of variable rotating speed and 25.6 kHz for the conditions of constant rotating speed, and three sensors are applied for vibration signals collection. The three sensors are respectively located on the left side of the planetary wheel X (Point1), planetary wheel Y (Point2) and the intermediate shaft of the gearbox (Point3). Each experiment was repeated for four times, and four vibration signals have been collected by each sensor for each fault state under each operating condition. Figure.6 shows the vibration signals collected under a constant rotating speed condition in Point1. Then, the vibration signals are divided into fault samples, and each fault sample contains 4096 sampling points. The fault samples obtained from two of the four vibration signals are used as training samples, and the rest samples are used for model testing. In this experiment, for each sensor, there are 3200 fault samples collected under variable speed condition, and 960 fault samples collected under constant speed condition. That is to say, for variable speed condition, 1600 samples are used for training, and 1600 samples are used for testing. For constant speed condition, 480 samples are used for training, and 480 samples are used for testing.

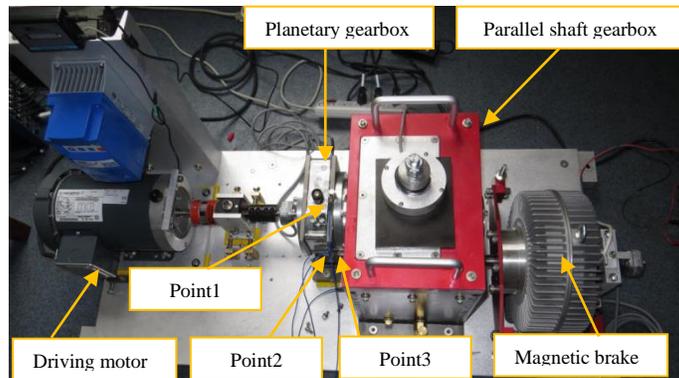


Fig.5. The test point arrangement in gearbox vibration test

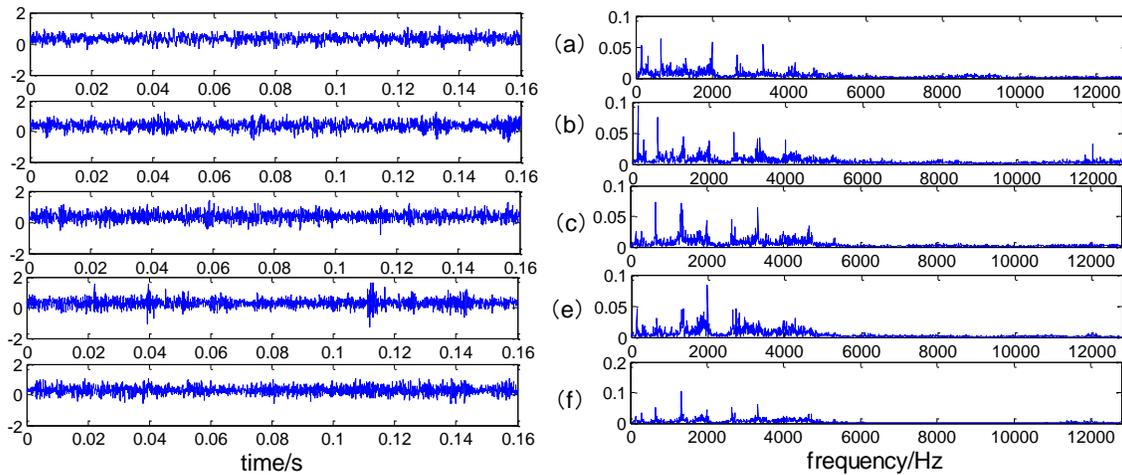


Fig.6. The time and frequency waveforms of the collected vibration signals: (a) tooth root corrosion; (b) tooth root crack; (c) bearing inner ring fault; (d) bearing outer ring fault; (e) mixed bearing fault

The proposed SVMF is compared with some other features extracted from the singular spectrum, including singular spectrum (SVS) [35], singular value ratios (SVR) [36] and normalized singular spectrum (SVE) [11][14]. The definitions of these features are shown in Table I.

TABLE I  
THE DEFINITION OF FAULT FEATURES

Processing Methods	Calculation Formulas
SVS	$\beta_i = \lambda_i$
SVR	$\gamma_i = \lambda_i / \lambda_{i+1}$
SVE	$\alpha_i = \lambda_i^2 / \sum_{j=1}^l \lambda_j^2$

For further validation of SVMF, SVMF is also compared with the time-frequency statistical feature (TF), wavelet packet energy (WPE), autoregressive models (AR) and information entropy (IE). Each kind of fault features is extracted under two conditions (variable speed and constant speed). The data size after processing is shown in Table II.

TABLE II  
THE SIZE OF FEATURE DATA

Feature Extraction Methods	Variable Speed Data (20-40Hz)	Constant Speed Data (20Hz)
SVMF	1600×24	960×24
TF	1600×24	960×24
WPE	1600×24	960×16
AR	1600×6	960×6
AR_10	1600×10	960×10
TF	1600×16	960×16
IE	1600×5	960×5

e.g. 1600×24 means 1600 samples, each with a dimension of 24.

In order to compare the effectiveness of feature extraction methods, SVMs is used to judge fault identification results on the platform of MATLAB R2016b with libsvm toolbox [37], and uniformly use the proposed IFFA to optimize the parameters of SVMs. Fault diagnosis models are performed for several times, and the most stable fault identification results are selected for analyzing, and the fault identification accuracy is shown in Table III. Fig.7 through Fig.10 show the curves of SVS, SVMF, SVE and SVR obtained under variable rotating speed. Several conclusions are derived from Table III: (1) it is obvious that the singular values contain a large amount of fault information. However, as shown in Fig.7, only a few relative larger singular values play a role in fault diagnosis, which as a result lead to poor fault identification; (2) SVMF, SVR and SVE are more effective than SVS, and among which SVMF is superior to other methods, especially for fault diagnosis under variable rotating speed. As shown in Fig.9, the proposed SVMF directly applied the manifold topologic structure information of singular spectrum to fault diagnosis and so overcome influences of the singular values. According to the computation formal of SVE, extraction of SVE is equivalent to the normalization of singular values. Hence, the problem that fault diagnosis accuracy is influenced by the value variation of singular values still exists in SVE. Distribution of SVE in Fig.9 is similar to the SVS in Fig.7, which further verifies the conclusion. Though the fault diagnosis results of SVR is very close to the proposed SVMF, as shown in Fig.10, the value range of SVR is random and non-uniform. So fault diagnosis result using SVR and SVE is still no satisfactory. On the other hand, the normalized value range and self-weighting of SVMF perfectly overcome the shortcoming of SVR and SVE, so fault diagnosis results using SVMF are satisfactory under both constant and variable rotating speed. In Table IV, the results of the SVMs using SVMF of different sensors are also listed out, which further proved the fault clustering effective of these proposed fault features. (3) The frequently-used statistical features perform well under constant rotating speed but unsatisfactory under variable rotating condition. The main reason is that the extraction of these features is not adaptive, so features extracted under different rotating speed are not comparable. In this experiment, it should also be noted that the time-frequency statistical feature contains energy normalization processing.

Table V also gives out the fault diagnosis result of SVS both with and without feature fusion, and the feature fusion is

achieved through dimension reduction by t-Distributed Stochastic Neighbor Embedding t-SNE [38]. After feature fusion, the dimension of SVS is reduced to two, and the range of numerical variation of features are decreased. As shown in Table VI, the testing results of SVS without t-SNE are random, and the fault diagnosis of SVS with t-SNE is much better than the original SVS. As a conclusion, the non-uniform scale of SVD-based features may lead to poor fault diagnosis performance and this also future proves why the proposed SVMF performs better than other SVD-based features.

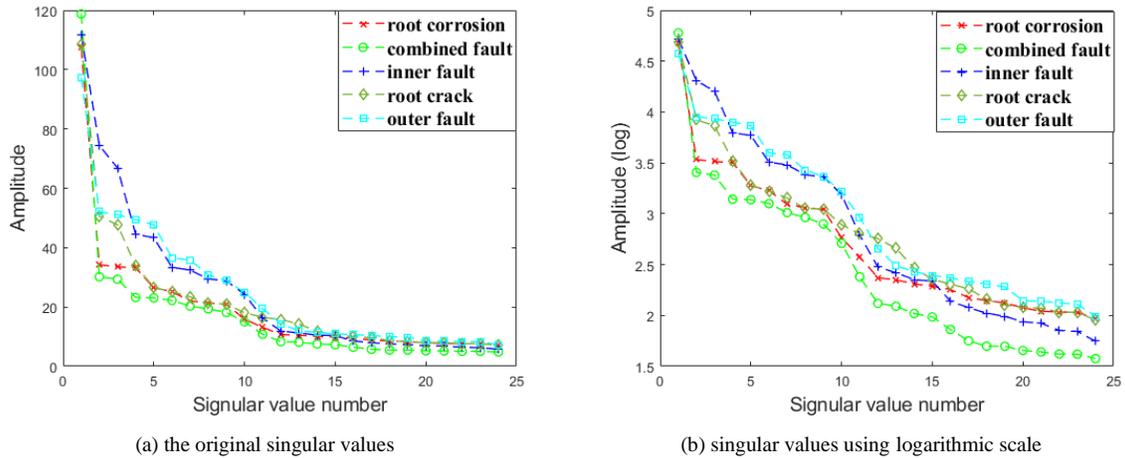


Fig.7 The curve of the singular spectrum obtained from original signals

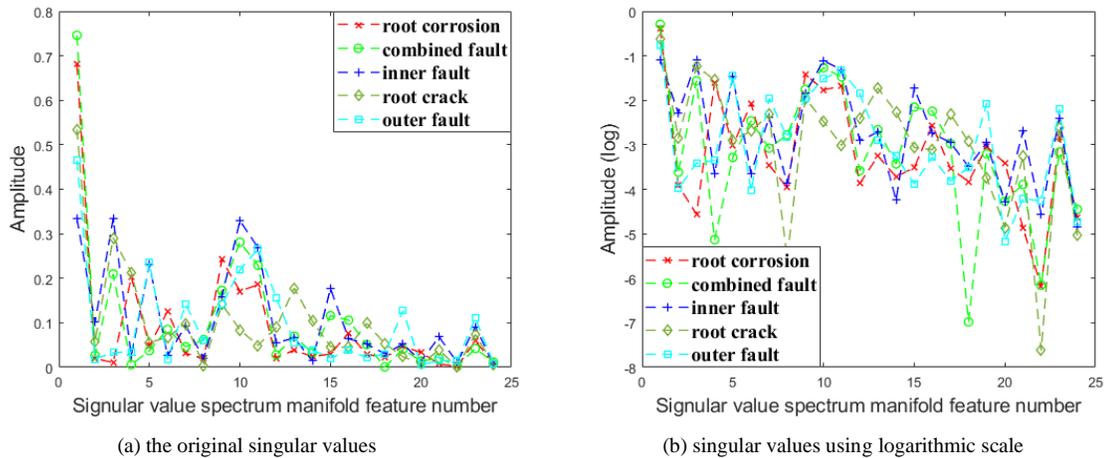


Fig.8 The curve of singular spectrum manifold features obtained from original signals

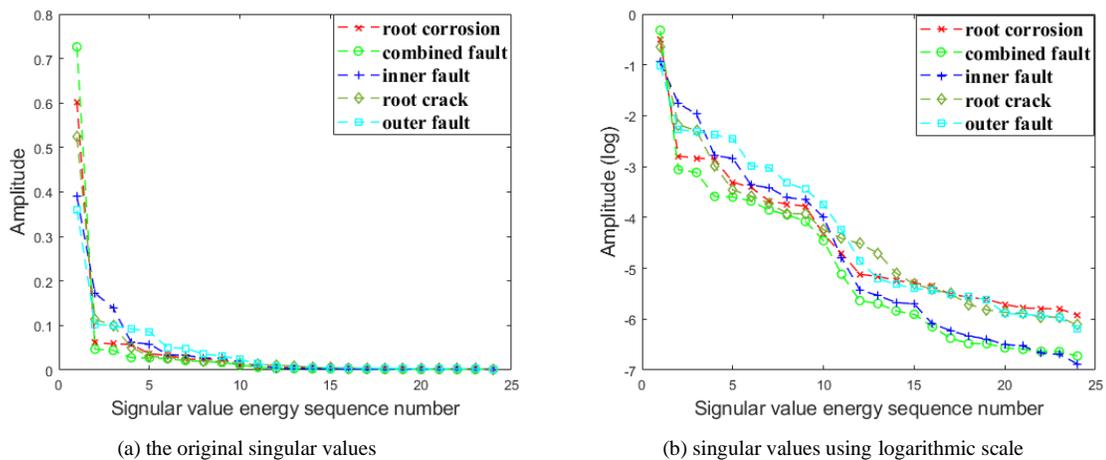


Fig.9 The curve of singular spectrum energy obtained from original signals

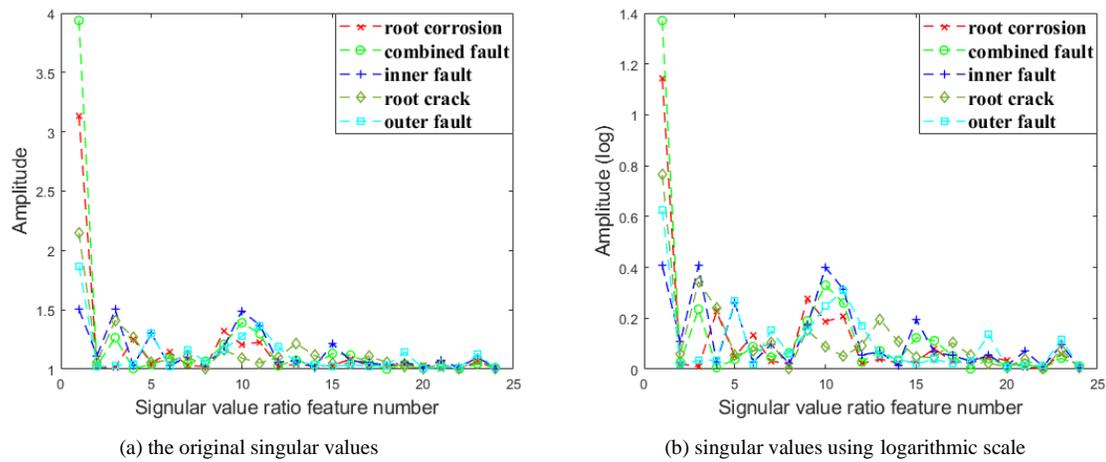


Fig.10 The curve of singular value ratios obtained from original signals

#### 4.2 Comparison of optimized SVMs

In this section, genetic algorithm (GA), adaptive particle swarm optimization (APSO) [39], bat algorithm (BA), classical fruit fly algorithm (FFA), guiding fireworks algorithm (GFWA) and IFFA are selected to optimize the parameters of SVMs to compare the performance of the proposed IFFA. The maximum number of iterations of all algorithms is set to 100 and the population size is set to 20.

In this section, the variable speed data obtained from three measuring points are selected, and the data are processed by time-frequency statistical feature, AR, wavelet packet energy, information entropy, and the proposed SVMF. Then the feature data are input into the support vector machines optimized by intelligent algorithms. The classification accuracy results are shown in Table VI, and the convergence curves of all algorithms are shown in Fig. 11.

From the comparative experiments in this section, we can see that the IFFA has certain advantages in improving the classification accuracy of SVMs. Compared with other algorithms, IFFA has good search performance and information utilization ability, owing to the introduction of guidance search mechanism and enhanced local search operation. Though in some cases IFFA has the same accuracy with other algorithms, IFFA shows better convergence speed. Therefore, the SVMs model based on IFFA has good classification accuracy and computational efficiency in fault diagnosis.

#### 4.3 Comparison of data fusion

In order to reflect the role of multi-sensor information fusion, this section uses D-S evidence theory and voting method to fuse the results of SVMs output from three measurement points. Similarly, the fused results of all feature extraction methods are compared, and the results are recorded in Table VII.

Regarding the voting method, there are two types, hard voting and soft voting, and the fusion rule is that the minority is subordinate to the majority. As shown in Table VII, soft voting can solve the conflict of hard voting, that is because soft voting introduces weight for each information source based on hard voting. In this experiment, the weights of the three measurement points, in turn, are 1, 1, and 1.5. That is to say, the weight of point 3 is 1.5 and the weights of remaining points are 1, because the fault diagnosis accuracy of point3 is better than the other two measurement points. Compared with both types of voting method, D-S evidence theory can integrate the results of each measurement point into a more ideal final fault diagnosis result. For this experiment, except for SVS, the final obtained accuracy has been considerably improved after the fusion by D-S evidence theory. All fault features have better classification accuracy after information fusion, especially in the working condition of

constant speed, the final accuracy of many features corresponds to 100%. While in the case of variable rotating speed, the fusion accuracy is also higher than that of the single measurement point. Note that: the fusion accuracy of the proposed SVMF is 100%, which is superior to other existing feature extraction methods.

TABLE III  
THE ACCURACY COMPARISON OF DIFFERENT FEATURE EXTRACTION METHODS (%)

working Condition	Measuring Points	SVS		SVR		SVE		SVMF		AR		wavelet packet energy		information entropy		time-frequency statistical feature	
		train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test
Variable Speed	Point1	50.88	20	97.63	95.38	95.13	92.19	<b>98.19</b>	<b>96</b>	96	91.31	98.69	95.13	69.38	63.81	88.13	80.81
	Point2	35.38	20	98.31	97.19	96	92.31	<b>99.13</b>	<b>97.75</b>	94.06	90.38	98.81	95.38	79.38	71.25	88.5	84.94
	Point3	33.88	20	99	98.19	96.31	93.94	<b>99.63</b>	<b>99.5</b>	89.88	88.19	97.06	96.19	88.06	82.94	93.88	78.94
Constant Speed	Point1	66.04	20	99.58	99.79	99.58	96.67	<b>100</b>	<b>100</b>	99.38	98.55	99.38	98.33	97.92	91.25	98.54	97.08
	Point2	64.79	20	<b>100</b>	<b>100</b>	<b>100</b>	99.17	<b>100</b>	<b>100</b>	95.63	94.38	98.13	95.66	94.58	93.54	99.38	96.67
	Point3	80.83	20	99.79	98.96	96.67	95.63	<b>99.79</b>	<b>98.96</b>	99.38	98.96	95.66	96.49	96.46	90	99.79	98.33

TABLE IV  
The testing recognition rate of each type of fault with SVMF (%)

Working condition	Measuring point	Root corrosion	Combined fault	Inner fault	Root crack	Outer fault
Variable speed	Point 1	100	96.5625	93.75	100	89.6875
	Point 2	96.875	99.6875	97.1875	97.5	97.5
	Point 3	100	98.75	99.375	100	99.375
Constant speed	Point 1	100	100	100	100	100
	Point 2	100	100	100	100	100
	Point 3	100	97.92	96.875	100	100

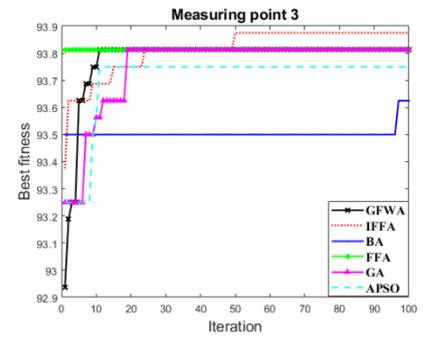
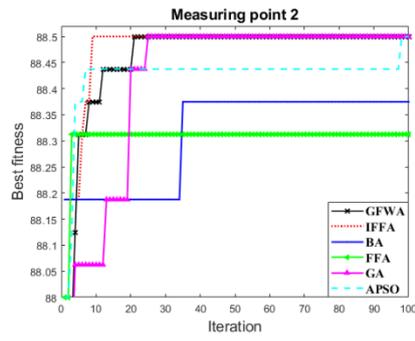
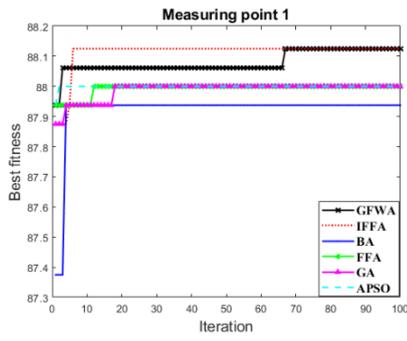
TABLE V  
FAULT DIAGNOSIS OF SVS WITH AND WITHOUT FEATURE FUSION (%)

working Condition	Measuring Points	SVS (without t-SNE)		SVS (with t-SNE)	
		train	test	train	test
Variable Speed	Point 1	50.88	20	80.81	77.31
	Point 2	35.38	20	84.31	81.5
	Point 3	33.88	20	88.63	77.25
Constant Speed	Point 1	66.67	20	98.33	90.625
	Point 2	68.13	20	99.79	98.125
	Point 3	78.54	20	98.125	93.125

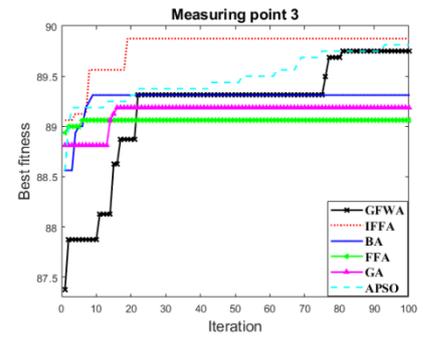
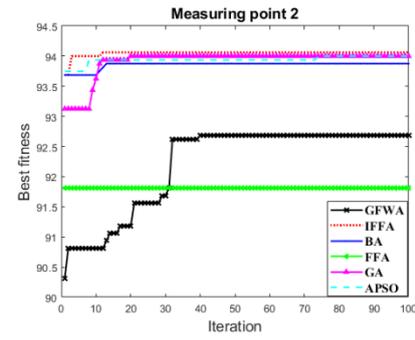
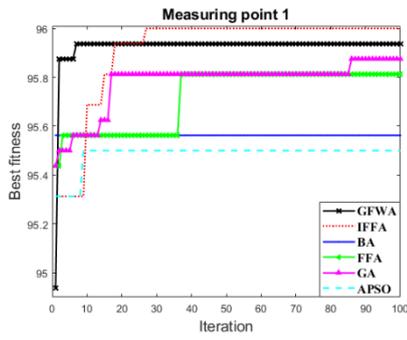
TABLE VI  
THE COMPARISON OF OPTIMIZED SVMs (%)

Feature extraction method	Measuring Points	APSO-SVM		GA-SVM		IFFA-SVM		BA-SVM		FFA-SVM		GFWA-SVM	
		train	test	train	test	train	test	train	test	train	test	train	test
TF	Point1	88	80.63	88	80.75	<b>88.13</b>	<b>80.81</b>	87.94	80.13	88	80.63	88.13	80.75
	Point2	88.5	84.88	88.5	84.69	<b>88.5</b>	<b>84.94</b>	88.38	84.69	88.31	84.19	88.5	84.75
	Point3	93.75	78.56	93.81	78.56	<b>93.88</b>	<b>78.94</b>	93.63	78.69	93.81	78.81	93.81	78.88
AR	Point1	95.5	90.88	95.88	91.19	<b>96</b>	<b>91.31</b>	95.56	91.13	95.81	91.19	95.94	91.19
	Point2	94	89.75	94	90.19	<b>94.06</b>	<b>90.38</b>	93.88	89	91.81	88.69	92.69	88.75
	Point3	89.81	87.88	89.19	87.81	<b>89.88</b>	<b>88.19</b>	89.31	87.69	89.06	87.69	89.75	87.94
WPE	Point1	98.5	95.06	98.63	95.13	<b>98.75</b>	<b>95.13</b>	98.69	94.625	98.31	94.81	98.69	95.13
	Point2	98.81	95.25	98.75	95.19	<b>98.81</b>	<b>95.38</b>	98.81	95.25	98.75	95.25	98.81	95.38
	Point3	97	95.75	97	95.81	<b>97.06</b>	<b>96.19</b>	96.69	95.75	95.38	95.06	97	95.81
IE	Point1	69	64.38	68.88	63	<b>69.38</b>	<b>63.81</b>	68.75	62.38	59.94	57.69	69	63.25
	Point2	78.44	70.94	79.31	70.94	<b>79.38</b>	<b>71.25</b>	75.94	69.06	66.81	62.13	79.25	71.25

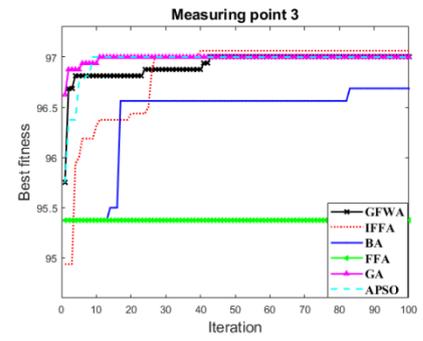
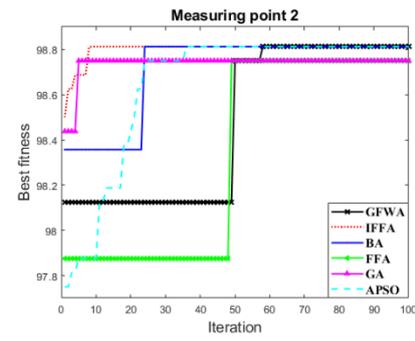
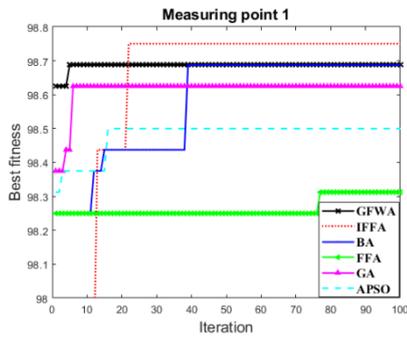




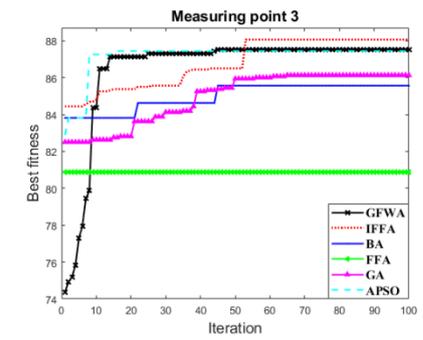
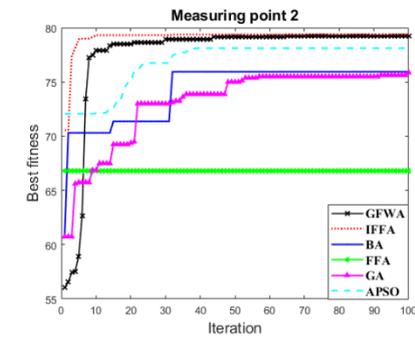
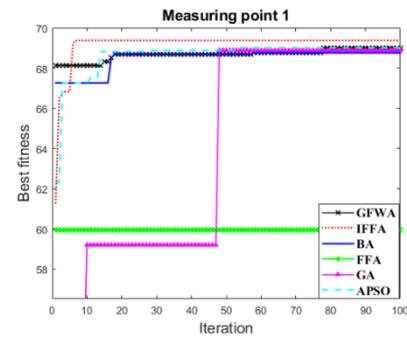
(a)



(b)



(c)



(d)

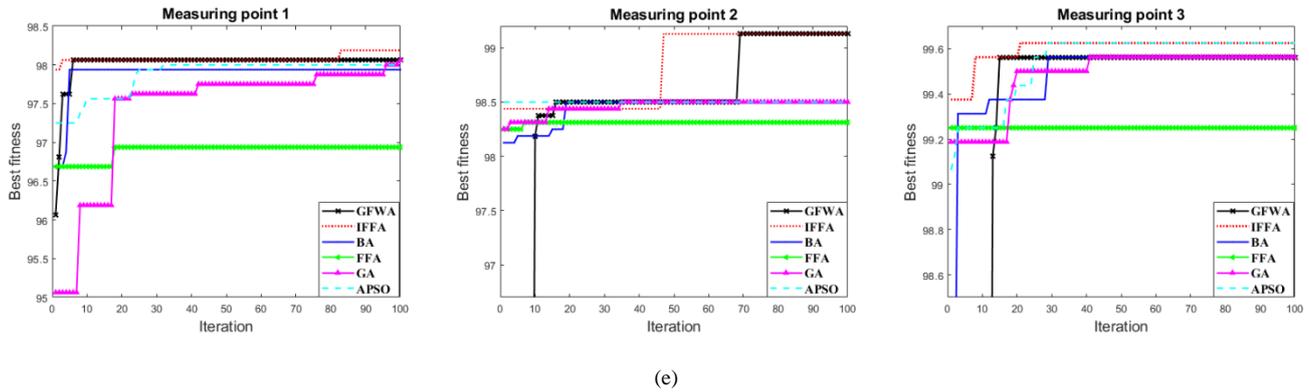


Fig.11 The convergence curves of algorithms. (a) time-frequency statistical feature. (b) AR. (c) wavelet packet energy feature. (d) information entropy. (e)

SVMF.

TABLE VII

THE FUSION RESULTS OF D-S EVIDENCE THEORY (%)

Working condition	Feature extraction method	Point1	Point2	Point3	Fusion result	Voting result	
		testing	testing	testing	fusion	Hard voting	Soft voting
Variable Speed	SVS	20	20	20	20	20	20
	SVR	95.38	97.19	98.19	99.875	Conflict	98.625
	SVE	92.1875	92.3125	93.9375	99.125	Conflict	99.125
	SVMF	96	97.75	99.5	<b>100</b>	<b>99.938</b>	<b>99.938</b>
	AR	91.3125	90.375	88.1875	99.375	Conflict	98.4375
	AR_10	92.375	90.9375	87.375	99.875	Conflict	98.625
	WPE	95.125	95.375	96.1875	99.875	Conflict	99.813
	IE	63.8125	71.25	82.9375	86.188	Conflict	85.313
Constant Speed	TF	80.8125	84.9375	78.9375	91.625	Conflict	90.6875
	SVS	20	20	20	20	20	20
	SVR	99.7917	<b>100</b>	98.9583	<b>100</b>	<b>100</b>	<b>100</b>
	SVE	96.667	99.1667	95.625	<b>100</b>	99.583	99.583
	SVMF	<b>100</b>	<b>100</b>	98.9583	<b>100</b>	<b>100</b>	<b>100</b>
	AR	98.5417	94.375	98.9583	<b>100</b>	99.375	99.375
	AR_10	98.55	97	98.375	<b>100</b>	99.375	100
	WPE	98.3333	95.625	96.4583	99.792	Conflict	99.167
TF	IE	91.25	93.5417	90	99.375	Conflict	97.917
	TF	97.0833	96.6667	98.3333	<b>100</b>	100	100

e.g. the “conflict” in the table indicates that the result cannot be fused by hard voting. AR is 6-dimensional AR data, AR\_10 is 10-dimensional data.

TABLE VIII

THE TIME IT TAKES FOR THE ALGORITHM TO ITERATE ONCE (S)

Feature extraction	Measuring point	APSO	GA	IFFA	BA	FFA	GFWA
TF	Point 1	26.814	<b>18.335</b>	20.162	30.215	19.537	28.242
	Point 2	27.697	<b>20.963</b>	26.624	31.354	21.461	32.019

	Point 3	17.916	<b>12.534</b>	16.647	18.367	13.291	18.370
	Point 1	34.032	<b>29.344</b>	30.765	37.130	33.973	35.343
AR	Point 2	31.730	<b>28.108</b>	32.777	39.736	34.359	36.585
	Point 3	34.907	<b>30.253</b>	37.850	39.446	35.258	40.833
	Point 1	33.205	<b>29.708</b>	31.653	40.818	32.187	40.651
WPE	Point 2	39.096	<b>31.844</b>	40.009	48.419	38.619	50.795
	Point 3	28.845	<b>23.228</b>	30.405	33.135	27.748	31.787
	Point 1	41.516	<b>35.695</b>	39.153	49.718	39.780	58.762
IE	Point 2	40.187	<b>33.548</b>	37.633	41.194	36.799	45.562
	Point 3	38.428	<b>32.677</b>	37.453	38.109	35.503	39.452
	Point 1	49.692	<b>41.669</b>	50.863	65.149	50.418	53.595
SVMF	Point 2	53.006	<b>43.625</b>	51.833	57.171	50.831	59.481
	Point 3	52.492	<b>42.614</b>	50.215	59.282	50.772	56.693

Table VIII shows the calculation time of one iteration of all algorithm compared in this experiment. From the calculation time of each algorithm, the genetic algorithm has obvious advantages, which is due to the fact that the genetic algorithm only needs to perform crossover and mutation operations to update the population. However, as can be seen from Figure 11, combined with the convergence of the algorithms, it can be seen that the improved fruit fly algorithm still has dominantly computational advantages.

## 5 CONCLUSION

This study proposes a fault diagnosis model based on singular spectrum manifold features, optimized SVMs and multi-sensor information fusion. First of all, SVMF is proposed which can make the fault information contained in small singular values available. In the fault identification stage, IFFA is proposed to select parameters of SVMs, which has good convergence speed and accuracy. To eliminate conflict of fault diagnosis conclusions of multiple sensors, D-S evidence theory is used to fuse the posterior probabilities of multiple sensors. The proposed fault diagnosis model integrated the advantages of SVMF in fault representation, optimal SVMs in fault identification and D-S evidence theory for information fusion, and as a result, the fault diagnosis accuracy is effectively improved. The fault diagnosis experiment of a gearbox demonstrated and verified the superiority of the proposed method. Finally, only vibration signals are applied for fault diagnosis in this paper, and the information fusion of multiple kinds of signals may future improve fault diagnosis performance. This is a very meaningful potential future work but also a tricky task.

## Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (51705060, 51605064, 51775072) and Chongqing Research Program of Basic Research and Frontier Technology (cstc2017jcyjAX0279).

## REFERENCES

- [1] Park J, Ha JM, Oh H, Youn BD, Choi JH, Kim NH, 2016. Model-based fault diagnosis of a planetary gear: A novel approach using transmission error. *IEEE Trans. Reliab.* **4**, 1830-1841.
- [2] Qin Y, Mao Y, Tang B, Wang Y, Chen H, 2019. M-band flexible wavelet transform and its application into planetary gear transmission fault diagnosis. *Mech. Syst. Signal Process.* **134**: 106298.
- [3] Van M, Kang HJ, 2016. Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization. *IEEE Trans. Ind. Informat.* **1**, 124-135.
- [4] Jianfu C, Ling C, Jialiang Z, Wen C, 2013. Fault diagnosis of complex system based on nonlinear frequency spectrum fusion. *Measurement.* **1**, 125-131.
- [5] Zhu X, Zhang Y, Zhu Y, 2012. Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features. *J. Mech. Sci. Technol.* **9**, 2649-2657.
- [6] Xu HB, Chen GH, Wang XH, Liang J, 2011. The application of EMD and ARMA Bi-cepstrum fault diagnosis method in gearbox of overhead traveling crane. *Adv. Mate. Res.* **308**, 88-91.
- [7] Cui H, Zhang L, Kang R, Lan X, 2009. Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method. *J. Loss. Prevent. Proc. Ind.* **6**, 864-867.
- [8] Wang J, He Q, 2016. Wavelet packet envelope manifold for fault diagnosis of rolling element bearings. *IEEE Trans. Instru. Meas.* **11**, 2515-2526.
- [9] Xing Z, Qu J, Chai Y, Tang Q, Zhou Y, 2017. Gear fault diagnosis under variable conditions with intrinsic time-scale decomposition-singular value decomposition and support vector machine. *J. Mech. Sci. Technol.* **2**, 545-553.
- [10] Zhang S, Lu S, He Q, Kong F, 2016. Time-varying singular value decomposition for periodic transient identification in bearing fault diagnosis. *J. Sound Vib.* **379**, 213-231.
- [11] Zhao X, Ye B, 2009. Similarity of signal processing effect between Hankel matrix-based SVD and wavelet transform and its mechanism analysis. *Mech. Syst. Signal. Pr.* **4**, 1062-1075.
- [12] Golafshan R, Sanliturk KY, 2016. SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults. *Mech. Syst. Signal. Pr.* **70**, 36-50.
- [13] Yang H, Lin H, Ding K, 2018. Sliding window denoising K-Singular Value Decomposition and its application on rolling bearing impact fault diagnosis. *J. Sound. Vib.* **421**, 205-219.
- [14] Muruganatham B, Sanjith MA, Krishnakumar B, Murty SS, 2013. Roller element bearing fault diagnosis using singular spectrum analysis. *Mech. Syst. Signal. Pr.* **35**, 150-166.
- [15] Hernandez-Vargas M, Cabal-Yepez E, Garcia-Perez A, 2014. Real-time SVD-based detection of multiple combined faults in induction motors. *Comput. Electr. Eng.* **7**, 2193-2203.
- [16] Qin Y, Wang X, Zou J, 2019. The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines *IEEE Trans. Ind. Electron.* **66**, 3814-3824.
- [17] Li Z, Fang H, Huang M, Wei Y, Zhang L, 2018. Data-driven bearing fault identification using improved hidden Markov model and self-organizing map. *Comput. Ind. Eng.* **116**, 37-46.
- [18] Liu R, Yang B, Zio E, Chen X, 2018. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal. Pr.* **108**, 33-47.
- [19] Lei Y, Zuo MJ, 2009. Gear crack level identification based on weighted nearest neighbor classification algorithm. *Mech. Syst. Signal. Pr.* **23**, 1535-1547.
- [20] Muralidharan V, Sugumaran V, 2012. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Appl. Soft. Comput.* **8**, 2023-2029.
- [21] Wenyi L, Zhenfeng W, 2013. Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM. *Renew. Energ.* **3**, 1-6.

- 1  
2  
3  
4 [22] Yan X, Jia M, 2018. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing.  
5 *Neurocomputing*. **33**, 47-64.
- 6 [23] Wei Y, Li Y, Xu M, Huang W, 2019. Intelligent fault diagnosis of rotating machinery using ICD and generalized composite multi-scale fuzzy entropy.  
7 *IEEE Access*. **7**, 38983-38995.
- 8 [24] Keerthi S S, 2002. Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms. *IEEE Trans. Neur. Net.* **13**, 1225-1229.
- 9 [25] Chapelle O, Vapnik V, Bousquet O, et al. 2002. Choosing multiple parameters for support vector machines. *Machine learning*. **46**, 131-159.
- 10 [26] Chen F, Tang B, Song T, Li L, 2014. Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle  
11 swarm optimization. *Measurement*. **1**, 576-590.
- 12 [27] Kari T, Gao W, Zhao D, Abiderexiti K, Mo W, Wang Y, Luan L, 2018. Hybrid feature selection approach for power transformer fault diagnosis based on  
13 support vector machine and genetic algorithm. *IET Gener, Transm. Dis.* **21**, 5672-5680.
- 14 [28] Pan WT, 2012. A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example. *Knowledge-Based Syst.* **26**, 69-74.
- 15 [29] Hui KH, Lim MH, Leong MS, Al-Obaidi SM, 2017. Dempster-Shafer evidence theory for multi-bearing faults diagnosis. *Eng. Appl. Artif. Intel.* **57**,  
16 160-170.
- 17 [30] Moosavian A, Khazaei M, Najafi G, Kettner M, Mamat R, 2015. Spark plug fault recognition based on sensor fusion and classifier combination using  
18 Dempster-Shafer evidence theory. *Appl. Acoust.* **93**, 120-129.
- 19 [31] Zhou Q, Zhou H, Zhou Q, Yang F, Luo L, Li T, 2015. The structural damage detection based on posteriori probability support vector machine and  
20 Dempster-Shafer evidence theory. *Appl. Soft. Comput.* **36**, 368-374.
- 21 [32] <http://www.phmsociety.org/references/datasets>. 2009 PHM Challenge Competition Data Set.
- 22 [33] Li J, Zheng S, Tan Y, 2017. The effect of information utilization: Introducing a novel guiding spark in the fireworks algorithm. *IEEE Trans. Evol.*  
23 *Comput.* **1**, 153-166.
- 24 [34] Yang XS, Hossein Gandomi A, 2012. Bat algorithm: A novel approach for global engineering optimization. *Eng. Comput.* **29**, 464-483.
- 25 [35] Tian Y, Ma J, Lu C, Wang Z, 2015. Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mech.*  
26 *Mach.Theory.* **90**, 175-186.
- 27 [36] Jiang H, Chen J, Dong G, Liu T, Chen G, 2015. Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis.  
28 *Mechanical Systems & Signal Processing*, **52**, 338-359.
- 29 [37] Chang C C, Lin C J, 2011. LIBSVM: A library for support vector machines. *ACM. Transon. Intel. Syst. Tec.* **2**, 1-27.
- 30 [38] Zheng JD, Jiang ZW, and Pan HY. Sigmoid-based refined composite multiscale fuzzy entropy and t-SNE based fault diagnosis approach for rolling  
31 bearing. *Measurement* 129(2018):332-342.
- 32 [39] Zhan ZH, Zhang J, Li Y, Chung HS, 2009. Adaptive particle swarm optimization. *IEEE Trans. Syst. Man. Cy-s. Part B (Cybernetics)*. **39**, 1362-1381.
- 33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60