

ORIGINAL ARTICLE

Open Access



Crack Fault Classification for Planetary Gearbox Based on Feature Selection Technique and K-means Clustering Method

Li-Ming Wang and Yi-Min Shao*

Abstract

During the condition monitoring of a planetary gearbox, features are extracted from raw data for a fault diagnosis. However, different features have different sensitivity for identifying different fault types, and thus, the selection of a sensitive feature subset from an entire feature set and retaining as much of the class discriminatory information as possible has a directly effect on the accuracy of the classification results. In this paper, an improved hybrid feature selection technique (IHFST) that combines a distance evaluation technique (DET), Pearson's correlation analysis, and an ad hoc technique is proposed. In IHFST, a temporary feature subset without irrelevant features is first selected according to the distance evaluation criterion of DET, and the Pearson's correlation analysis and ad hoc technique are then employed to find and remove redundant features in the temporary feature subset, respectively, and hence, a sensitive feature subset without irrelevant or redundant features is selected from the entire feature set. Further, the k-means clustering method is applied to classify the different kinds of health conditions. The effectiveness of the proposed method was validated through several experiments carried out on a planetary gearbox with incipient cracks seeded in the tooth root of the sun gear, planet gear, and ring gear. The results show that the proposed method can successfully distinguish the different health conditions of a planetary gearbox, and achieves a better classification performance than other methods. This study proposes a sensitive feature subset selection method that achieves an obvious improvement in terms of the accuracy of the fault classification.

Keywords: Planetary gearbox, Gear crack, Feature selection technique, K-means classification

1 Introduction

Owing to its advantages of a compact structure, large transmission ratio, and high load capacity, a planetary gear transmission system is widely used in large-scale and complex mechanical equipment [1, 2], e.g., wind turbines, helicopters, and automobiles.

A planetary gearbox typically consists of some key components: a sun gear, planet gear, ring gear, carrier, and bearing, and faults may occur in these components owing to fatigue or tough working conditions. According to a condition-monitoring report on wind turbines, a gearbox failure is the leading contributor to all wind turbine failures [3]. A vibration-based method was proven

to be one of the most popular techniques in the fault diagnosis of rotating machinery, and it has been determined that certain changes to the vibration signals can be seen when a fault occurs, e.g., crack or spalling [4–6]. The commonly used vibration signal processing methods can be divided into three categories: time domain methods, frequency domain methods, and time–frequency domain methods. Time domain methods refer to the analysis of a signal with respect to time, and are relatively easy and direct compared to both frequency and time–frequency domain methods. Statistical indicators, the time synchronous averaging (TSA) method, and an autoregressive (AR) model are typically used in the fault diagnosis of rotating machinery [4, 7–9]. Frequency domain methods refer to an analysis of the signals with respect to the frequency, and a periodic signal in the time domain can be converted into a frequency component through a Fourier

*Correspondence: ymshao@cqu.edu.cn
State Key Laboratory of Mechanical Transmission, Chongqing University,
Chongqing 400044, China

transformation. In this way, researchers can identify the difference between the spectrum of a normal vibration signal and a fault vibration signal with commonly used methods that include a spectrum-based analysis, resonance demodulation technique, and cepstrum analysis [10–12]. In contrast, time–frequency domain methods are used to study a signal in both the time and frequency domains simultaneously, allowing both the constituent frequency components and their time variation features to be revealed and analyzed. Researchers have developed various time–frequency domain methods including a short-time Fourier transformation, Wigner–Ville distribution, continuous wavelet transform, and Hilbert–Huang transformation [13–16]. Although vibration-based methods have been successfully used in the fault diagnosis and condition monitoring of rotating machinery, the appearance of faults in the analysis results has to be identified artificially, e.g., the identification of a fault characteristic frequency in the spectrum, the determination of a filter sub-band in a demodulation analysis, or the determination of a wavelet type, all of which require considerable of experience and expertise [17–19]. Therefore, it is necessary to develop some intelligent techniques that can automatically determine the health conditions of a planetary gearbox.

Feature extraction is commonly the first important step in an intelligent technique. However, different features display different sensitivity to fault advancements, and some of the features are redundant or irrelevant to the fault diagnosis or classification result [19–21]. Selecting a sensitive feature subset and retaining as much of the class discriminatory information as possible has a direct effect on the accuracy of the results. Kang et al. [20] developed an outlier-insensitive hybrid feature selection methodology to reduce diagnostic performance deterioration caused by outliers in data-driven diagnostics. Peng et al. [22] studied how to select good features according to the maximal statistical dependency criterion based on mutual information. Yang et al. [23] demonstrated an approach to the multi-criteria optimization problem of feature subset selection using a genetic algorithm. Li et al. [24] presented a two-stage feature selection approach combining filter and wrapper techniques to obtain a more compact feature subset for accurate classification of the hybrid faults of a gearbox. Liu et al. [25] introduced a hybrid dimension reduction method that combines kernel feature selection and a kernel Fisher discriminant analysis for a fault-level diagnosis of planetary gearboxes. Lei et al. [19, 26] and Shen et al. [27] proposed a feature subset selection method based on a distance evaluation technique (DET) for fault classification of a roller bearing and gear reducer. Among these feature subset selection methods, DET is relatively simpler and more efficient

than the other methods, and has thus been widely used in fault diagnosis [28–30]. However, DET tends to select redundant features because it does not consider the relationships between features; one relevant feature selected by DET may be redundant in the presence of another relevant feature with which it is strongly correlated. To address this problem, an improved hybrid feature selection technique (IHFST) combining DET, Pearson's correlation analysis, and an ad hoc technique is proposed. Using IHFST, the relevant features are selected according to the distance evaluation criterion of DET, and the redundant features are then further suppressed based on the Pearson's correlation analysis and ad hoc technique. Hence, not only irrelevant features, but also redundant features, are removed from the entire feature set based on the proposed IHFST.

Once a sensitive feature subset is selected, it can be classified into several classifications based on some machine learning (ML) methods. The k-means clustering method is a widely used unsupervised pattern recognition algorithm that aims to partition n observations into k clusters, where each observation belongs to the cluster with the nearest mean [31–34]. Therefore, the k-means clustering method is further employed to validate the effectiveness of the proposed IHFST method.

The rest of this paper is organized as follows: Section 2 briefly introduces the feature-subset selection based on DET. Section 3 presents the fault classification process based on the proposed IHFST and k-means method. Section 4 describes the experiment system and data acquisition. The effectiveness of the proposed method as validated using various datasets is then described in Section 5. Finally, some concluding remarks are provided in Section 6.

2 Brief Review of DET

In fault diagnosis, different features can reflect different aspects of the vibration properties of the machinery, and certain features are sensitive to certain changes in machine conditions [19–21, 25]. If all features are applied to a fault diagnosis without a careful selection, the computation complexity of the algorithm will increase with little gain [26]. Therefore, a feature selection process is necessary to reduce the computation complexity and improve the accuracy of the fault diagnosis. The feature selection process based on DET is briefly introduced.

The basic concept of DET is to measure the ratio of the between-class distance to the within-class distance in the feature vector space, as illustrated in Figure 1, where a high DET value indicates high sensitivity or class separability of the feature.

The detailed process of DET can be expressed as follows [19, 26, 27].

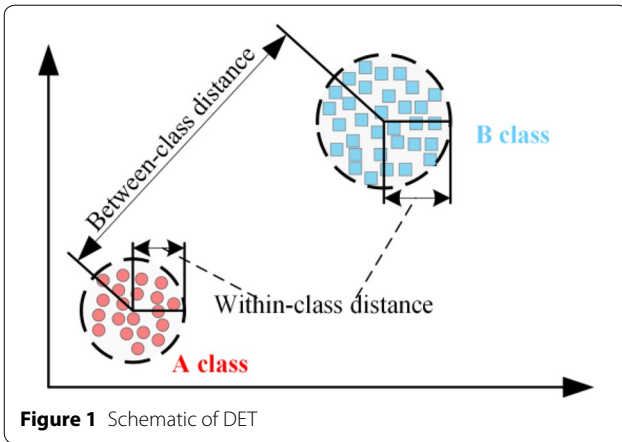


Figure 1 Schematic of DET

Suppose that the entire feature set is obtained by

$$\{q_{i,k,j}, i = 1, 2, \dots, I_k; k = 1, 2, \dots, K; j = 1, 2, \dots, J\}, \quad (1)$$

where $q_{i,k,j}$ is the feature value of the j th feature from the i th sample of the k th health condition, I_k is the total sample number of the k th condition, K represents the condition number, and J is the feature number in the feature vector of each sample. The process of informative feature selection based on DET is as follows:

Step 1: Calculate the within-class average distance of the same condition samples,

$$d_{k,j} = \frac{1}{I_k(I_k - 1)} \sum_{l,j=1}^{I_k} |q_{i,k,j} - q_{l,k,j}|, \quad (2)$$

$l, i = 1, 2, \dots, I_k, l \neq i,$

and then obtain the average distance of K conditions,

$$d_j^{(w)} = \frac{1}{K} \sum_{k=1}^K d_{k,j}. \quad (3)$$

Step 2: Calculate the average feature value of all samples under the same condition,

$$\mu_{k,j} = \frac{1}{I_k} \sum_{i=1}^{I_k} q_{i,j,k}. \quad (4)$$

Then, obtain the average between-class distance between different condition samples,

$$d_j^{(b)} = \frac{1}{K(K - 1)} \sum_{k,e=1}^K |u_{e,j} - u_{k,j}|, \quad (5)$$

$k, e = 1, 2, \dots, K, k \neq e.$

Step 3: Calculate the ratio between $d_j^{(b)}$ and $d_j^{(w)}$, assigning the compensation factor $\varepsilon_j = d_j^{(b)} / d_j^{(w)}$, and then normalize it based on the maximum value and obtain the distance evaluation criterion $\alpha_j = \varepsilon_j / \max(\varepsilon_j)$.

Hence, a feature subset can be selected when the distance evaluation criterion α_j is greater than a given threshold. However, the DET method only removes irrelevant features from the entire feature set. Although the

features selected by DET carry good classification information when treated separately, there may be little gain if they are combined into a feature vector because of a high mutual correlation [21–25].

3 Fault Classification Based on the Proposed IHFST and K-means

Figure 2 shows a flow chart of fault classification based on the proposed IHFST and k-means classification method. The entire feature set is first extracted from time domain signal, frequency domain signal, and difference signal. A sensitive feature subset without irrelevant features or redundant features is then selected based on the proposed IHFST. The feature subset is then classified using the k-means clustering method.

3.1 Feature Extraction

Feature extraction is the first important step of the proposed method. Before calculating the features, the raw vibration signals are divided into several segments by multiplying a sliding Hanning window owing to its good properties, e.g., better side-lobe behavior shown in the spectrum [35], as indicated in Figure 3. The sliding Hanning window equation is given as follows:

$$W(m) = \begin{cases} 0.5 \left[1 - \cos \left(\frac{2\pi(m-\delta)}{M-1} \right) \right], & \delta \leq m \leq M + \delta - 1, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $M = k \times f_s$ is the width of the sliding Hanning window, k is the time length of the window, f_s is the sampling frequency, $\delta = j \times f_s$ is the sliding distance, and j is the time length of the sliding step. Here, M and δ are both

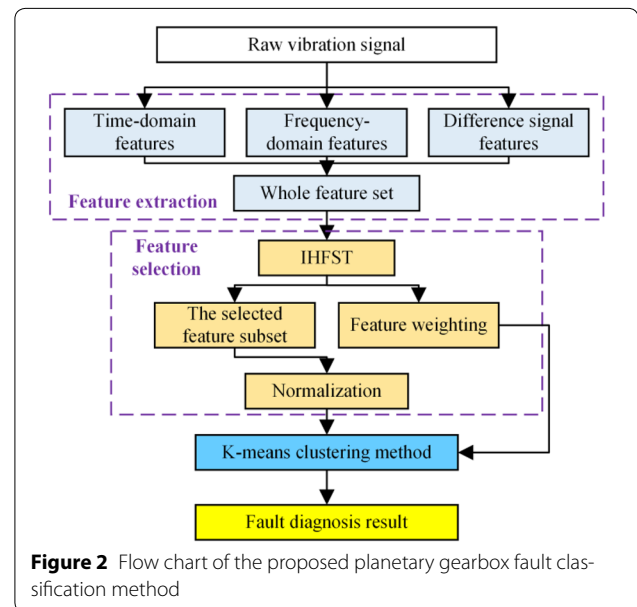
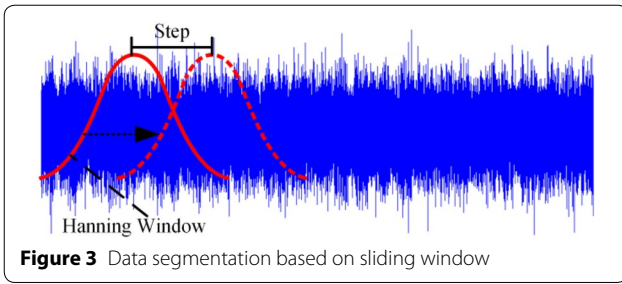


Figure 2 Flow chart of the proposed planetary gearbox fault classification method



integers, and the values of k and j were chosen as 4 and 0.5, respectively.

Three types of features are extracted from time domain signal, frequency domain signal, and difference signal, respectively, as given in Table 1. The difference signal is defined as Ref. [4] kept and act as feature weights of the selected

$$d = x - y^d, \tag{7}$$

where x is the raw vibration signal, y^d is the signal containing the mesh frequencies, their harmonics, and their first-order sidebands. Thus, d is composed of higher-order sidebands and Gaussian noise.

Table 1 Calculated features

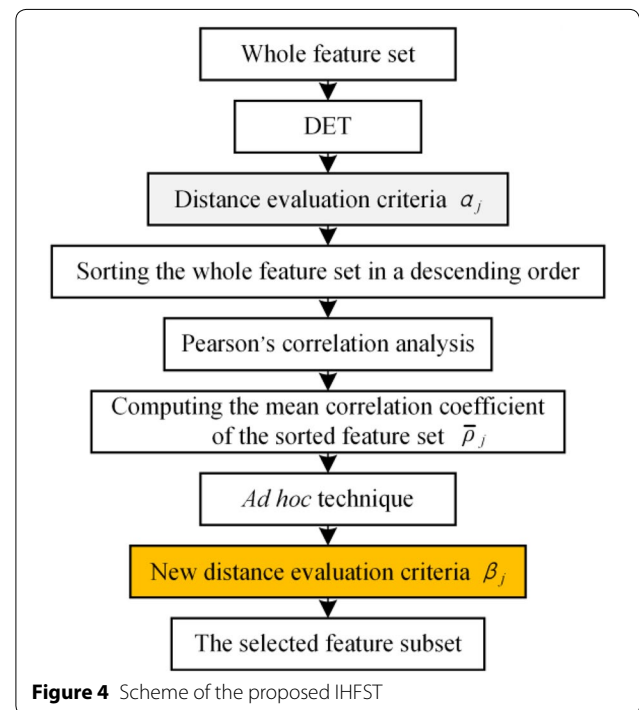
No.	Feature name	Equation
1	Mean	$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$
2	Peak to peak	$x_{p-p} = \max(x) - \min(x)$
3	Variation	$x_{var} = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2$
4	Root mean square	$x_{rms} = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2}$
5	Skewness	$x_{skew} = \frac{1}{N} \sum_{n=1}^N \left(\frac{x_n - \bar{x}}{\sqrt{x_{var}}} \right)^3$
6	Kurtosis	$x_{kurtosis} = \frac{1}{N} \sum_{n=1}^N \left(\frac{x_n - \bar{x}}{\sqrt{x_{var}}} \right)^4$
7	Crestor factor	$x_{cf} = \frac{\max(x)}{x_{rms}}$
8	Impulse factor	$x_{if} = \frac{\max(x)}{1/N \sum_{n=1}^N x_n }$
9	Mean of frequency	$f_{p1} = \frac{\sum_{k=1}^K S_k}{K}$
10	Variation of frequency	$f_{p2} = \frac{\sum_{k=1}^K (S_k - f_{p1})^2}{K-1}$
11	3rd moment of frequency	$f_{p3} = \frac{\sum_{k=1}^K (S_k - f_{p1})^3}{K(\sqrt{f_{p2}})^3}$
12	4th moment of frequency	$f_{p4} = \frac{\sum_{k=1}^K (S_k - f_{p1})^4}{K(f_{p2})^2}$
13	Energy ratio	$ER = \frac{\sqrt{1/N \sum_{n=1}^N (d_n - \bar{d})^2}}{\sqrt{1/N \sum_{n=1}^N (y_n^d - \bar{y}^d)^2}}$
14	FM0	$FM0 = \frac{PP_x}{\sum_{h=0}^H P_h}$
15	FM4	$FM4 = \frac{N \sum_{n=1}^N (d_n - \bar{d})^4}{[\sum_{n=1}^N (d_n - \bar{d})^2]^2}$
16	M6A	$M6A = \frac{N^2 \sum_{n=1}^N (d_n - \bar{d})^6}{[\sum_{n=1}^N (d_n - \bar{d})^2]^3}$

Here, x is the original time domain signal; N is the number of samples; S_k is the spectrum of x for $k = 1, 2, \dots, K$; K is the number of spectrum lines; \bar{y}^d and \bar{d} are the mean values of y^d and d , respectively; PP_x indicates the maximum peak-to-peak amplitude of signal x ; P_h is the amplitude of the h th harmonic; H is the total number of harmonics within the frequency ranges.

In this study, eight features are calculated from time domain signal, four features from frequency domain signal, and four from difference signal. In addition, because the rotating speed is low, the calculated gear mesh frequency [36] and its harmonics are concentrated at a low frequency range, and thus, a low-pass digital filter is used to cover the main gear mesh components of the vibration signals, and cut-off frequencies of 650 and 900 Hz are chosen for the low-pass digital filter at 300 and 400 r/min, respectively. Thus, another sixteen features are calculated from the filtered vibration signals, and a total of thirty-two features are obtained within the entire feature set.

3.2 IHFST

To remove both irrelevant and redundant features from the entire feature set, the IHFST method combining DET, Pearson's correlation analysis, and the ad hoc technique is proposed, as shown in Figure 4. The distance evaluation criterion α_j of each feature is obtained based on DET, and the entire feature set is then sorted in descending order according to α_j . Next, the mean correlation coefficient of the sorted feature set can be computed based on the Pearson's correlation analysis.



$$\bar{\rho}_j = \frac{1}{M-1} \sum_{m=1}^{M-1} |\rho(q_{k,J_m}, q_{k,j})| \text{ and} \tag{8}$$

$$\rho(q_{k,J_m}, q_{k,j}) = \frac{\sum_{i=1}^{I_k} (q_{i,k,J_m} - \bar{q}_{k,J_m})(q_{i,k,j} - \bar{q}_{k,j})}{\sqrt{\sum_{i=1}^{I_k} (q_{i,k,J_m} - \bar{q}_{k,J_m})^2} \sqrt{\sum_{i=1}^{I_k} (q_{i,k,j} - \bar{q}_{k,j})^2}}, \tag{9}$$

$J_m \neq j$

where, $\bar{\rho}_j$ is the mean correlation coefficient of the j th feature, M is the ranking index of the j th feature according to α_j , J_m is the feature index of the top informative feature for $m = 1, 2, \dots, M - 1$, $q_{k,J_m} = [q_{1,k,J_m}, q_{2,k,J_m}, \dots, q_{I_k,k,J_m}]$ and $q_{k,j} = [q_{1,k,j}, q_{2,k,j}, \dots, q_{I_k,k,j}]$ are two feature vectors, and \bar{q}_{k,J_m} and $\bar{q}_{k,j}$ are the mean values of these two features, respectively. The correlation coefficient ranges from -1 to $+1$, which indicates a high degree of negative or positive linear correlation between q_{k,J_m} and $q_{k,j}$.

Finally, the ad hoc technique associated with the mean correlation coefficient is applied to reevaluate the distance evaluation criteria α_j , and a feature that is highly correlated with the top informative features will be suppressed or removed. The new distance evaluation criterion β_j is obtained, i.e.,

$$\beta_j = \begin{cases} \alpha_j, & \text{when } M = 1, \\ \theta_1 \alpha_j - \theta_2 \bar{\rho}_j, & \text{when } M \geq 2, \end{cases} \tag{10}$$

where, θ_1 and θ_2 are two weighting factors that determine the relative importance of the informative and the correlation terms, respectively. A large θ_1 factor emphasizes the informative term, and a relatively larger θ_2 factor weights the correlation term more heavily, and can thus produce a feature subset with less redundancy [25]. In this paper, when the mean correlation coefficient is higher than 0.7, the feature will be recognized as a severe redundant feature, and will thus be removed. In another situation described herein, the two factors are chosen as $\theta_1 = 1$ and $\theta_2 = 0.2$.

Hence, a feature subset without irrelevant or redundant features can be selected according to the value of the new distance evaluation criteria β_j .

3.3 Normalization and Weighting

Before classification, all features are normalized as follows:

$$nFE_j = \frac{FE_j - \text{mean}(FE_j)}{\text{std}(FE_j)}, \quad j = 1, 2, \dots, J, \tag{11}$$

where FE_j is the feature to be normalized, $\text{mean}()$ is the mean function, and $\text{std}()$ is the standard variance function.

Although the feature subset is selected based on the IHFST, different features in the feature subset have different sensitivities during the fault diagnosis process. Thus, to emphasize the importance of sensitive features,

a feature-weighting step is needed to guarantee a more accurate classification result. Supposing there are L features selected by IHFST, described in Section 3.2, the new distance evaluation criteria β_j of these selected features are also kept and act as feature weights of the selected features.

3.4 K-means Clustering Method

As an unsupervised clustering method, k-means is commonly used to automatically partition samples into k clusters [31]. The purpose of the k-means clustering method is to assign all N samples into k clusters by minimizing the sum of point-to-centroid distances as follows:

$$J = \arg \min_C \sum_{i=1}^k \sum_{x \in C_i} \|\vec{x} - \vec{\mu}_i\|, \tag{12}$$

where $C = \{C_1, C_2, \dots, C_k\}$ indicates k clusters, \vec{x} is an $N \times R$ feature matrix, R represents the dimensions of the matrix, each row is a single observation or sample, and $\vec{\mu}_i$ indicates the cluster centroid of the i th cluster. The detailed process of the k-means clustering method is given as follows.

Step 1: Initialization

Randomly choose k cluster centroids for all feature samples.

Step 2: Assignment

Assign each sample to the nearest cluster centroid by measuring the distance between the sample and each centroid as follows:

$$C_i^t = \{x_p : \|x_p - \mu_i^t\|^2 \leq \|x_p - \mu_j^t\|^2 \quad \forall j, 1 \leq j \leq k\}. \tag{13}$$

Step 3: Update

Find all samples in each cluster, and determine the new cluster centroid using

$$\mu_i^{t+1} = \frac{1}{N_i^t} \sum_{x_j \in C_i^t} x_j, \tag{14}$$

where N_i^t is the sample number of the i th cluster at the t th iteration.

Step 4: Repeat Steps 2 and 3 until the cluster centroid remains unchanged or the function achieves convergence.

4 Experimental System

Figure 5 illustrates the “back-to-back” planetary gearbox experiment setup, which consists of two planetary gearboxes (PGB), two motors, an industrial computer, a control box, and a set of vibration acquisition systems. The test bed is symmetrically arranged, PGB 1 and PGB 2 have the same structure and two motors each, and the

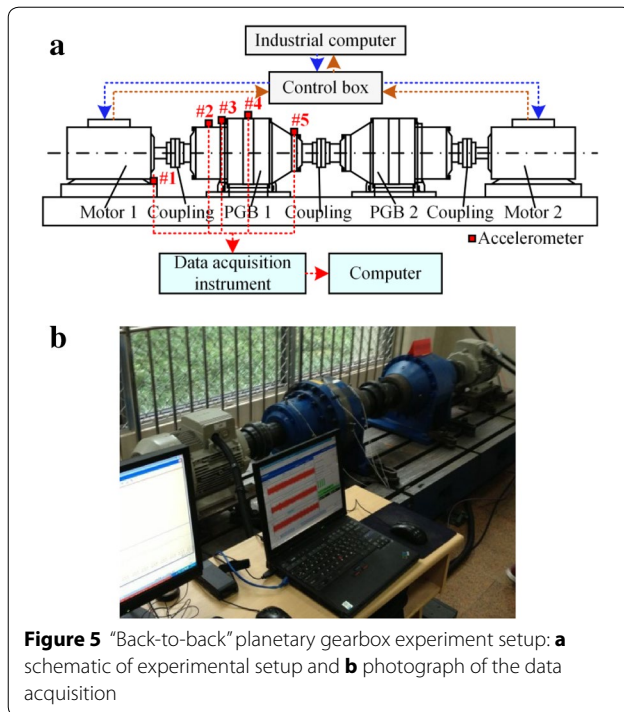


Figure 5 “Back-to-back” planetary gearbox experiment setup: **a** schematic of experimental setup and **b** photograph of the data acquisition

four components are connected using three teeth-shaft couplings.

The design parameters of the PGB are given in Table 2, where the PGB is comprised of an inner sun gear surrounded by four planet gears and a standstill ring gear, and the gear ratio of the PGB is 6.25. It should be noted that the simulated faults are located in PGB 1. The two motors are SIEMENS three-phase 15 kW induction motors, where motor 1 is for driving and motor 2 is for loading; in addition, the range of the motor speed is 0–1450 r/min, and the two motors are controlled using an industrial computer through a control box. To measure the vibration signals, five Kistler integrated circuit piezoelectric accelerometers, denoted as #1, #2, #3, #4, and #5, are placed in the test bed. Here, #1 is placed in the motor base, #2 is placed in the housing of the input bearing, #3 is placed in the housing of the planetary gearbox, #4 is placed in the ring gear of the planetary gearbox, and #5 is placed in the housing of the output bearing, as

Table 2 Design parameters of planetary gear

Parameters	Sun	Planet	Ring
Number of teeth	16	33	84
Modulus/mm	4	4	4
Pressure angle/(°)	20	20	20
Tooth thickness/mm	70	70	70
Planet number	4		

illustrated in Figure 5a. An LMS SCADAS system and a computer are used for data acquisition, the sampling frequency of which is 20,480 Hz, and the sampling length is 11 s.

Figure 6a–c illustrate seeded incipient cracks located at the tooth root of the sun gear, planet gear, and ring, respectively. The parameters of the seeded crack are defined as $(q_0, q_1, W_c, \alpha_c)$, as shown in Figure 7. As is well known, it may be more difficult to detect gear faults at low speed because low-frequency characteristics are easily masked by heavy noise [18]. To verify the effectiveness of the proposed method, vibration datasets were collected under two relatively low rotating speeds of 300/400 r/min and four loading conditions of 0, 20, 40, and 60 Nm, as listed in Table 3. Altogether, there were eight vibration datasets collected, each of which includes four types of health conditions: normal, cracked sun gear (CS), cracked planet gear (CP), and cracked ring gear (CR). In addition, the vibration signals of accelerometer #4 mounted on the outer ring gear of PGB 1 were analyzed in this study.

In addition, Figure 8a, b show an example of vibration signals for each health condition in dataset 1 and their spectrum, respectively. It can be observed there are few differences in the raw vibration signals and their spectrum for the four health conditions: normal, CS, CP, and CR, and it is difficult to distinguish the three types of faults from the normal condition.

5 Results and Discussion

The distance evaluation criteria α and β , and the corresponding feature weights of dataset 2, are shown in Figure 9a, b, respectively. It can be observed that three features (Nos. 19, 20, and 26) with higher mean correlation coefficients in the DET method are suppressed according to the new distance evaluation criteria β . Twelve sensitive features are selected from the entire feature set using the DET method with a given threshold, $\text{Thr}(\alpha) = \text{mean}(\alpha_j)$, and eight sensitive features are selected using the IHFST method with a given threshold, $\text{Thr}(\beta) = \text{mean}(\beta_j)$, where $j = 1, 2, \dots, 32$ denotes the feature index. It should be noted that dataset 2 is taken as an example to show the effectiveness of the proposed method, and the other datasets are also analyzed.

In this study, three types of feature selection methods are employed to analyze the same dataset using the proposed method for comparison, which are denoted as Method 1, Method 2, and Method 3. In Method 1, two commonly used features, the root mean square and kurtosis, are selected from the entire feature set. In Method 2, the dimensions of the entire feature set are directly reduced based on a principle component analysis (PCA) [37] without feature selection. In Method 3, features are

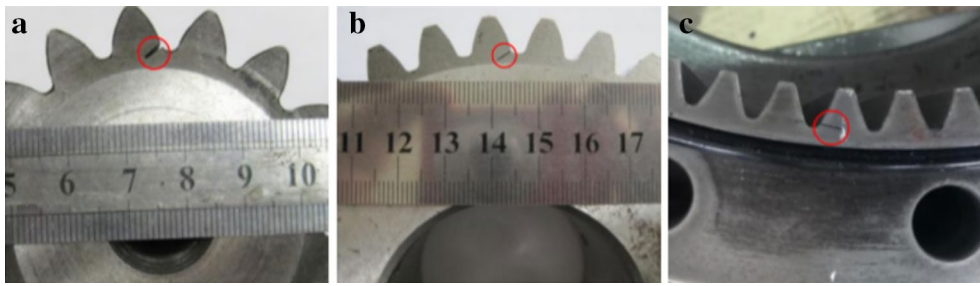


Figure 6 A seeded tooth root crack in **a** a sun gear, **b** planet gear, and **c** ring gear

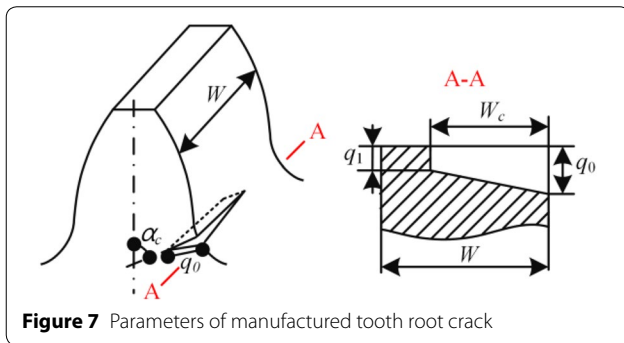


Figure 7 Parameters of manufactured tooth root crack

Table 3 Working conditions of the experiments

Dataset No.	Working conditions	Crack parameters
1	300 r/min, 0 Nm	Each dataset includes four types of health conditions: (1) normal conditions, (2) cracked sun gear (3.5 mm, 1 mm, 3.5 mm, 60°), (3) cracked planet gear (4 mm, 1 mm, 4 mm, 60°), and (4) cracked ring gear (6 mm, 1 mm, 4 mm, 80°)
2	300 r/min, 20 Nm	
3	300 r/min, 40 Nm	
4	300 r/min, 60 Nm	
5	400 r/min, 0 Nm	
6	400 r/min, 20 Nm	
7	400 r/min, 40 Nm	
8	400 r/min, 60 Nm	

selected based on DET. Finally, the proposed method is referred to as Method 4.

The classification performances of these four methods are shown in Figure 10a–d. In addition, it should be noted that, for visualization, the PCA method is implemented based on the clustering results produced through Method 3 and Method 4, and the plots of the first two PCs are shown in Figure 10c, d. Many misclassification samples can be found in Figure 10a, b, which are based on Method 1 and Method 2, and the samples have a loose distribution. In Figure 10c, CS and CP can be clearly discriminated, but it is difficult to discriminate the CR from normal conditions, and there are five misclassification samples. In Figure 10d, the four health conditions can be clearly discriminated, and all

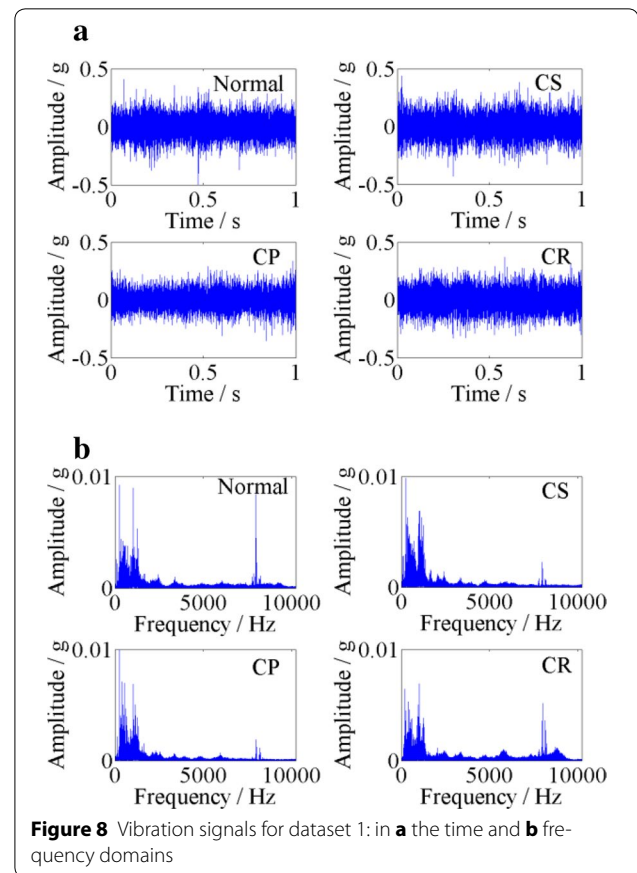
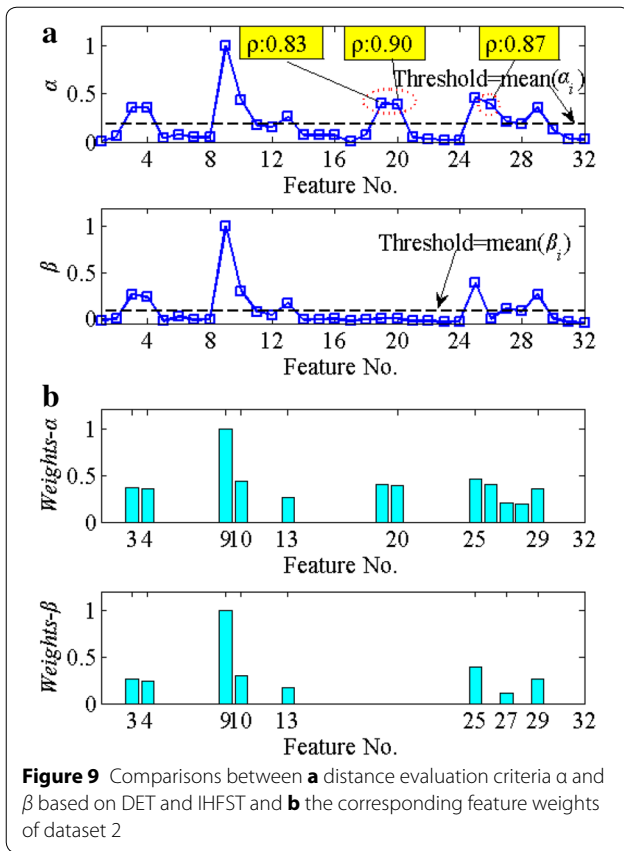


Figure 8 Vibration signals for dataset 1: in **a** the time and **b** frequency domains

four clusters have small within-class distances and large between-class distances.

To compare the classification performance of the four methods, the silhouette value (SV) is adopted in this study. The SV is defined as a measure of how similar a point is to other points in its own cluster compared to points in other clusters, and ranges from -1 to $+1$, where a value close to $+1$ indicates a better classification performance [38]. The SV can therefore be obtained:

$$S(i) = \frac{\min(b(i, :), 2) - a(i)}{\max(a(i), \min(b(i, :)))}, \tag{15}$$



where $a(i)$ is the average distance from the i th point to other points in its cluster, and $b(i, k)$ is the average distance from the i th point to points in another cluster k . The SV of the classification is obtained based on the average of all $S(i)$.

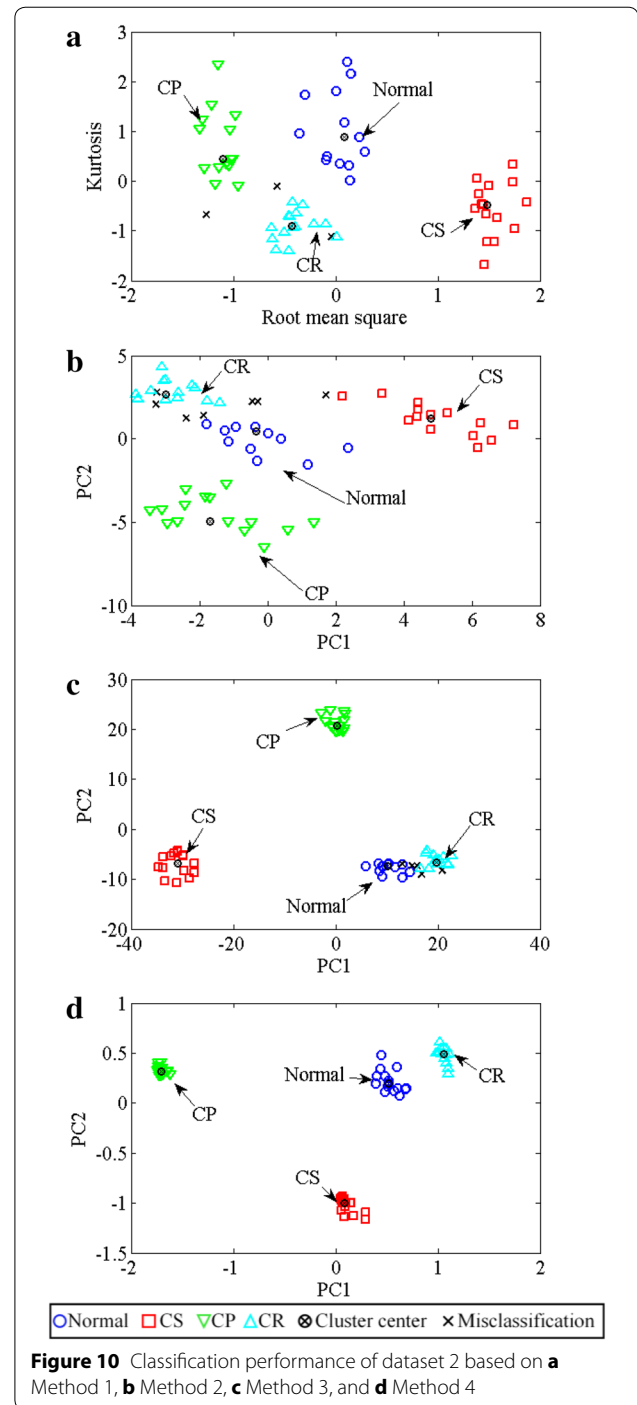
Table 4 presents the classification results of different datasets based on the four methods, along with their misclassification error and SV . There are 36 misclassification errors with Method 1, 44 misclassification errors with Method 2, 7 misclassification errors with Method 3, and zero misclassification errors with Method 4. In addition, the SV of Method 4 is greater than that of the other three methods for all datasets.

The improvement of Method 4 over the other three methods is considered to be as follows:

$$\Delta S_k = \frac{SV_{M4} - SV_{Mk}}{SV_{M4}} \times 100\%, k = 1, 2, 3, \quad (16)$$

where SV_{M4} denotes the silhouette values of the classification using Method 4, and SV_{Mk} denotes these values for the other three methods, for $k = 1, 2$, and 3 .

Figure 11 shows a comparison of the classification performance between Method 4 and the other three methods. It can be observed that Method 4 consistently yields a better classification performance than the other three

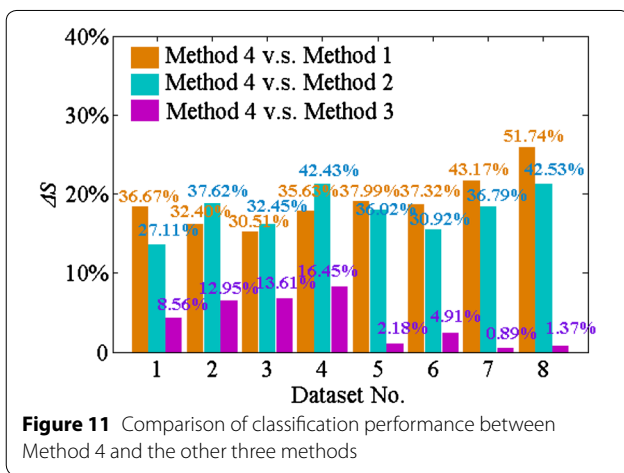


methods for the different datasets. Method 4 achieves a 30.51% to 51.74% improvement over Method 1, 30.92% to 42.43% improvement over Method 2, and 0.89% to 16.45% improvement over Method 3.

As mentioned in Section 3.1, feature extraction is one of the most important steps in the proposed method, and the parameters of the sliding Hanning window have a

Table 4 Classification performance of the four methods

Dataset No.	Method 1		Method 2		Method 3		Method 4	
	Error	SV	Error	SV	Error	SV	Error	SV
1	5	0.5618	0	0.6466	0	0.8112	0	0.8871
2	3	0.5702	7	0.5262	5	0.7343	0	0.8435
3	8	0.5610	7	0.5453	0	0.6974	0	0.8073
4	5	0.5213	14	0.4662	2	0.6766	0	0.8098
5	0	0.5688	7	0.5868	0	0.8972	0	0.9172
6	2	0.5754	1	0.6342	0	0.8729	0	0.9180
7	3	0.5224	2	0.5810	0	0.9110	0	0.9192
8	10	0.4417	6	0.5260	0	0.9028	0	0.9153



direct influence on the classification performance. Therefore, the effects of the Hanning window width and sliding distance on the classification performance of the four methods are illustrated in Figures 12 and 13, respectively.

Figure 12a–h display the effect of the window width on eight datasets, 1 to 8. The window width ranges from 1 to 6 s, and the sliding distance is fixed to 0.5 s. It can be seen that, as the window width increases, the SVs of the four methods show an increasing trend, which may result from too little fault-related information in each segment when the window width is small. In addition, Method 4 shows a higher SV than Method 1 and Method 2 for each dataset, and shows a higher SV than Method 3 for datasets 1 to 4, and dataset 6.

Figures 13a–h display the effect of the sliding distance on the eight datasets, 1 to 8. The sliding distance ranges from 0.1 to 0.9 s, and the window width is fixed to 4 s. It can be seen that, as the sliding distance increases, the SVs of Method 3 and Method 4 remain nearly unchanged, and the SVs of Method 1 and Method 2 have a small range of fluctuations. In addition, Method 4 shows a

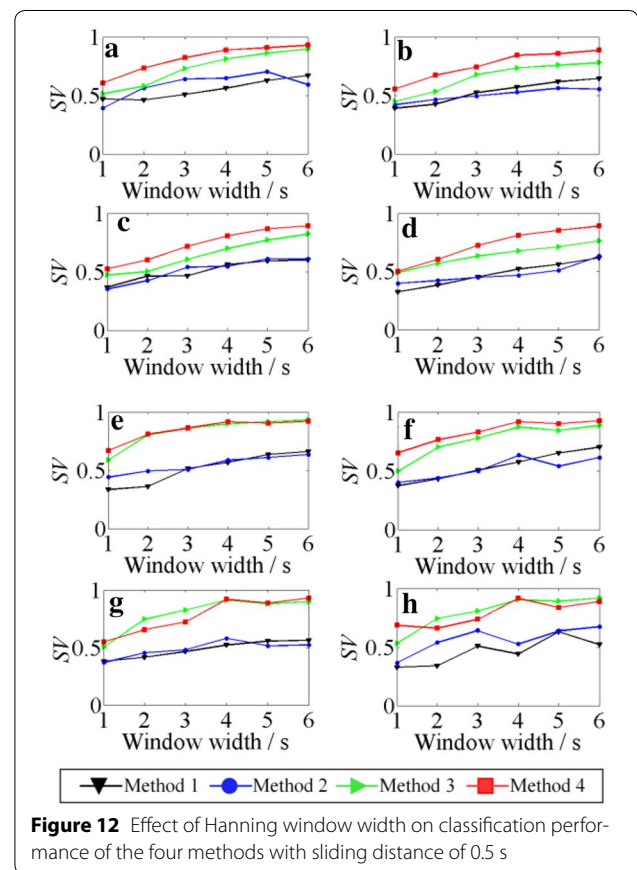


Figure 12 Effect of Hanning window width on classification performance of the four methods with sliding distance of 0.5 s

higher SV than Method 1 and Method 2 for each dataset, and a higher SV than Method 3 for most of the datasets.

For Method 3 and Method 4, a feature subset is selected from the entire feature set with a given threshold, and it is important to clarify the relationship between the threshold and the classification performance. Hence, the effect of the feature selection threshold on the classification performance for Method 3 and Method 4 is also studied, as shown in Figure 14. The thresholds for Method 3 and

Method 4 range from $0.5 \times \text{Thr}(\alpha)$ to $2.0 \times \text{Thr}(\alpha)$, and from $0.5 \times \text{Thr}(\beta)$ to $2.0 \times \text{Thr}(\beta)$, respectively. It can be seen that, as the threshold increases, the SVs of Method

3 and Method 4 have a small range of fluctuation, and Method 4 shows a higher SV than Method 3 for each dataset.

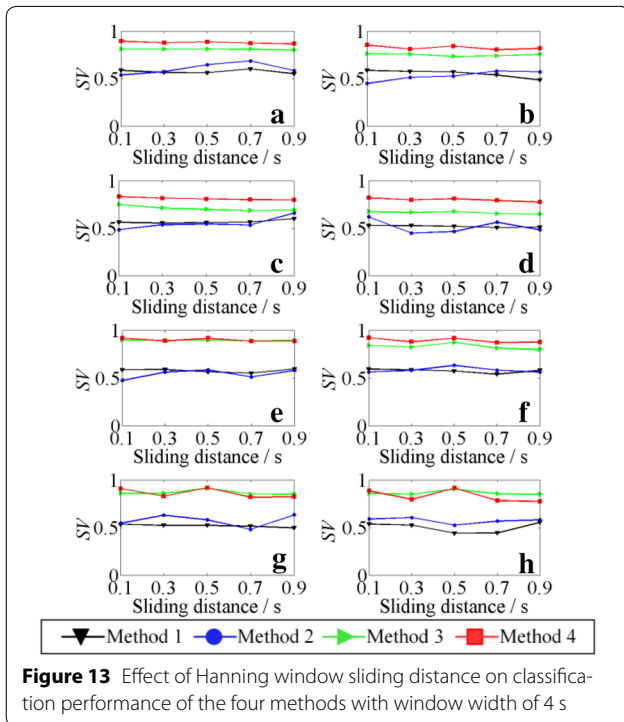


Figure 13 Effect of Hanning window sliding distance on classification performance of the four methods with window width of 4 s

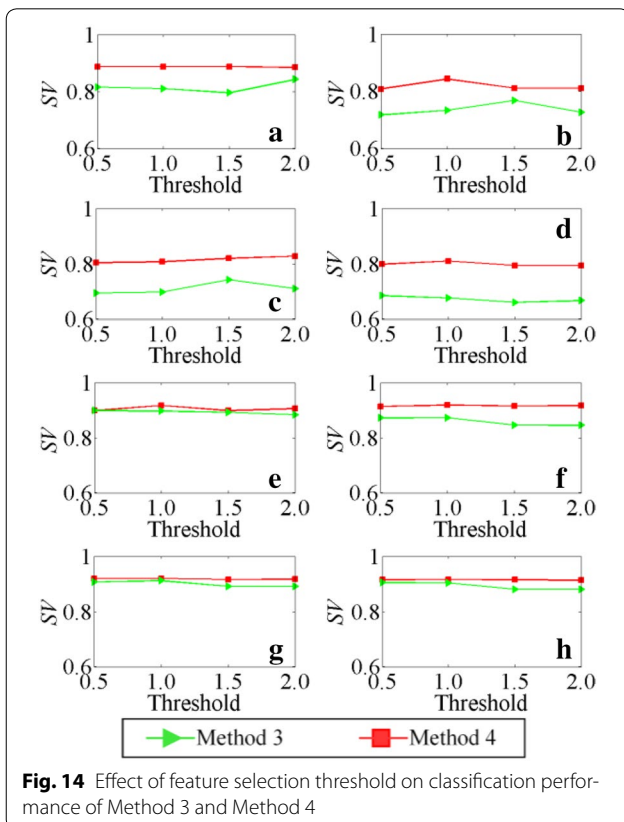


Fig. 14 Effect of feature selection threshold on classification performance of Method 3 and Method 4

6 Conclusions

In this paper, an IHFST method combining DET, Pearson's correlation analysis, and an ad hoc technique was proposed. A sensitive feature subset without irrelevant or redundant features was selected from the entire feature set based on the proposed IHFST method. The k-means clustering method was further employed to automatically partition the vibration data acquired from a planetary gearbox with crack faults into several different classifications. To validate the effectiveness of the proposed method, three types of feature selection methods were employed to analyze the same dataset using the proposed method for comparison. The results indicate that the proposed method can discriminate the four types of health conditions of a planetary gearbox clearly for all the datasets used, and no misclassifications were found. The proposed method achieves a 30.51% to 51.74% improvement over Method 1, 30.92% to 42.43% improvement over Method 2, and 0.89% to 16.45% improvement over Method 3. In addition, the influence of the sliding Hanning window and feature selection threshold on the classification performance was investigated, and the proposed method again achieved a better classification performance than the other three methods.

Authors' contributions

LW carried out the experiment studies and drafted the manuscript. YS carried out the sequence alignment and revision of the manuscript. All authors read and approved the final manuscript.

Authors' Information

Li-Ming, Wang born in 1987, is currently a PhD candidate at *State Key Laboratory of Mechanical Transmission, Chongqing University, China*. He received his bachelor degree from *Chongqing University, China*, in 2011. His main research interests include fault diagnosis of rotating machinery and signal processing.

Yi-Min Shao, born in 1963, is currently a professor at *Chongqing University, China*. He received his PhD degree from *Gunma University, Japan*, in 1997. His research interests include mechanical dynamics, signal processing, fault diagnosis of rotating machinery.

Acknowledgements

Supported by National Natural Science Foundation of China (Grant No. 51475053).

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 2 January 2017 Accepted: 14 January 2018

Published online: 28 February 2018

References

1. Y C Guo, R G Parker. Analytical determination of mesh phase relations in general compound planetary gears. *Mechanism and Machine Theory*, 2011, 46(12): 1869–1887.
2. W Smith, L Deshpande, R Randall, et al. Gear diagnostics in a planetary gearbox: a study using internal and external vibration signals. *International Journal of Condition Monitoring*, 2013, 3(2): 36–41.
3. S Sheng, H Link, W LaCava, et al. Wind turbine drivetrain condition monitoring during GRC phase 1 and phase 2 testing. *Contract*, 2011, 303: 275–300.
4. P Samuel, D J Pines. A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of Sound and Vibration*, 2005, 282(1): 475–508.
5. Z G Tian, M J Zuo, S Wu. Crack propagation assessment for spur gears using model-based analysis and simulation. *Journal of Intelligent Manufacturing*, 2012, 23(2): 239–253.
6. J Liu, Y M Shao. Dynamic modeling for rigid rotor bearing systems with a localized defect considering additional deformations at the sharp edges. *Journal of Sound and Vibration*, 2017, 398: 84–102.
7. W Wang, A K Wong. Autoregressive model-based gear fault diagnosis. *Transactions-American Society of Mechanical Engineers Journal of Vibration and Acoustics*, 2002, 124(2): 172–179.
8. J M Ha, B D Youn, H Oh, et al. Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines. *Mechanical Systems and Signal Processing*, 2016, 70: 161–175.
9. H Al-Bugharbee, I Trendafilova. A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling. *Journal of Sound and Vibration*, 2016, 369: 246–265.
10. Z P Feng, M Liang, Y Zhang, et al. 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.
11. V CMN Leite, J G B da Silva, G F C Veloso, et al. Detection of localized bearing faults in induction machines by spectral kurtosis and envelope analysis of stator current. *IEEE Transactions on Industrial Electronics*, 2015, 62(3): 1855–1865.
12. B Liang, S D Iwnicki, Y Zhao. Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis. *Mechanical Systems and Signal Processing*, 2013, 39(1): 342–360.
13. C Li, V Sanchez, G Zurita, et al. Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement. *ISA Transactions*, 2016, 60: 274–284.
14. B P Tang, W Y Liu, T Song. Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution. *Renewable Energy*, 2010, 35(12): 2862–2866.
15. J L Chen, Z P Li, J Pan, et al. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 2016, 70: 1–35.
16. Z W Gao, C Cecati, S X Ding. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Transactions on Industrial Electronics*, 2015, 62(6): 3757–3767.
17. V Muralidharan, V Sugumaran. Rough set based rule learning and fuzzy classification of wavelet features for fault diagnosis of monoblock centrifugal pump. *Measurement*, 2013, 46(9): 3057–3063.
18. Y G Lei, J Lin, M J Zuo, et al. Condition monitoring and fault diagnosis of planetary gearboxes: A review. *Measurement*, 2014, 48: 292–305.
19. Y G Lei, M J Zuo. Gear crack level identification based on weighted K nearest neighbor classification algorithm. *Mechanical Systems and Signal Processing*, 2009, 23(5): 1535–1547.
20. M Kang, M R Islam, J Kim, et al. A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Transactions on Industrial Electronics*, 2016, 63(5): 3299–3310.
21. S Theodoridis, K Koutroumbas. *Pattern Recognition, Fourth Edition*. Burlington, MA: Academic Press, 2008.
22. H C Peng, F H Long, C Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1226–1238.
23. J Yang, V Honavar. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems and Their Applications*, 1998, 13(2): 44–49.
24. B Li, P L Zhang, H Tian, et al. A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox. *Expert Systems with Applications*, 2011, 38(8): 10000–10009.
25. Z L Liu, J Qu, M J Zuo, et al. Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis. *The International Journal of Advanced Manufacturing Technology*, 2013, 67(5-8): 1217–1230.
26. Y G Lei, Z J He, Y Y Zi, et al. New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. *Mechanical Systems and Signal Processing*, 2008, 22(2): 419–435.
27. Z J Shen, X F Chen, X L Zhang, et al. A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM. *Measurement*, 2012, 45(1): 30–40.
28. B S Yang, K J Kim. Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals. *Mechanical Systems and Signal Processing*, 2006, 20(2): 403–420.
29. Q Hu, Z J He, Z S Zhang, et al. Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mechanical Systems and Signal Processing*, 2007, 21(2): 688–705.
30. A Widodo, B S Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 2007, 21(6): 2560–2574.
31. M Yuwono, Y Guo, J Wall, et al. Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in heating ventilation and air conditioning systems. *Applied Soft Computing*, 2015, 34: 402–425.
32. C T Yiakopoulos, K C Gryllias, I A Antoniadis. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, 2011, 38(3): 2888–2911.
33. G F Wang, C Liu, Y H Cui. Clustering diagnosis of rolling element bearing fault based on integrated autoregressive/autoregressive conditional heteroscedasticity model. *Journal of Sound and Vibration*, 2012, 331(19): 4379–4387.
34. M E Celebi, H A Kingravi, P A Vela. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, 2013, 40(1): 200–210.
35. J Barros, R I Diego. On the use of the Hanning window for harmonic analysis in the standard framework. *IEEE Transactions on Power Delivery*, 2006, 21(1): 538–539.
36. Z P Feng, M J Zuo. Vibration signal models for fault diagnosis of planetary gearboxes. *Journal of Sound and Vibration*, 2012, 331(22): 4919–4939.
37. J Harmouche, C Delpha, D Diallo. Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part II. *Signal Processing*, 2015, 109: 334–344.
38. P J Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational & Applied Mathematics*, 1987, 20(20): 53–65.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com