Principal Components of Superhigh-Dimensional Statistical Features and Support Vector Machine for Improving Identification Accuracies of Different Gear Crack Levels under Different Working Conditions

Wang, Dong; Tsui, Kwok-Leung; Tse, Peter W.; Zuo, Ming J.

Research Article

Principal Components of Superhigh-Dimensional Statistical Features and Support Vector Machine for Improving Identification Accuracies of Different Gear Crack Levels under Different Working Conditions

Dong Wang,[1] Kwok-Leung Tsui,[1] Peter W. Tse,[1] and Ming J. Zuo[2,3]

[1]*Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*
[2]*School of Mechatronics Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, Sichuan 611731, China*
[3]*Department of Mechanical Engineering, University of Alberta, Edmonton, AB, Canada T6G 1H9*

Correspondence should be addressed to Kwok-Leung Tsui; kltsui@cityu.edu.hk

Gears are widely used in gearbox to transmit power from one shaft to another. Gear crack is one of the most frequent gear fault modes found in industry. Identification of different gear crack levels is beneficial in preventing any unexpected machine breakdown and reducing economic loss because gear crack leads to gear tooth breakage. In this paper, an intelligent fault diagnosis method for identification of different gear crack levels under different working conditions is proposed. First, superhigh-dimensional statistical features are extracted from continuous wavelet transform at different scales. The number of the statistical features extracted by using the proposed method is 920 so that the extracted statistical features are superhigh dimensional. To reduce the dimensionality of the extracted statistical features and generate new significant low-dimensional statistical features, a simple and effective method called principal component analysis is used. To further improve identification accuracies of different gear crack levels under different working conditions, support vector machine is employed. Three experiments are investigated to show the superiority of the proposed method. Comparisons with other existing gear crack level identification methods are conducted. The results show that the proposed method has the highest identification accuracies among all existing methods.

1. Introduction

Gears are commonly used in mechanical transmission systems to transmit power from one shaft to another. Because gear is a mechanical component, its performance degrades over time when it is used [1–4]. Gear crack evaluation is a special case only when the gear crack is taken as the first agent in gear performance degradation process [5]. Identification of different gear crack levels is beneficial in preventing any unexpected machine breakdown and reducing economic loss because gear crack leads to gear tooth breakage [6]. Vibration analysis is a major tool to diagnose gear faults

because vibration signals are easily collected from the casing of gearboxes [7, 8]. In the past years, vibration signal based wavelet analysis has been widely used to diagnose gear faults [9–11]. Additionally, in recent years, to diagnose the multistage gearbox used in a bucket wheel excavator system, Bartelmus and Zimroz [12–16] proposed a simple and effective health indicator for distinguishing the good and bad health conditions of the multistage gearbox. Data collected from such complex machines are nonstationary gearbox vibration signals under time-varying working conditions. Because the gearbox used in a bad health condition is more susceptible to external varying loads, the proposed health

indicator was designed to be a function of the instantaneous input speed. The results show that the proposed health indicator has a linear relationship with the instantaneous input speed when the gearbox is in the good or bad health condition. Moreover, as the gearbox degrades, the inclination of the linear relationship increases.

Wavelet analysis is one of the most popular methods for diagnosing gear faults. For the use of wavelet analysis, the selection of a proper wavelet basis is crucial because wavelet analysis aims to calculate the inner product between a signal and the wavelet basis. Recently, Rafiee et al. [17, 18] made an exhaustive study for investigating the performance of 324 mother wavelet candidates on gear fault feature extraction. Their results showed that Daubechies 44 has the most similar shape with gear fault features. Even though gear fault diagnosis becomes a hot topic, according to our literature review, only some methods, such as mean frequency of Scalogram [19], instantaneous energy density [20], regularization dimension [21], cyclic spectral analysis [22], and self-similarity [23], are highly related to identification of different gear crack levels. However, because those methods are based on signal processing, the explanation of the results obtained by using these methods requires expertise.

To automatically identify different gear crack levels, intelligent methods [24–26] are required to be developed. To fill out this gap, first, a series of experiments [5] were conducted to collect gear vibration signals under different working conditions including four different motor speeds and three different loads. Five different artificial gear crack levels, including crack level 0%, crack level 25%, crack level 50%, crack level 75%, and crack level 100%, were produced to simulate gear crack deterioration levels. The definition of these different gear crack levels will be illustrated in Section 3.1. Then, a weighted K-nearest neighbor algorithm [24] was proposed to identify three different gear crack levels including crack level 0%, crack level 25%, and crack level 50% under four different motor speeds and three different loads because identification of early gear crack levels is more useful to conduct preventive maintenance. Their results showed that the weighted K-nearest neighbor algorithm achieves high prediction accuracies to identify three different gear crack levels under the four different motor speeds and the three different loads.

In this paper, to improve the work reported in [24], identification of three gear crack levels under the different loads and speeds is extended to identification of five gear crack levels under the different loads and speeds. It means that two more crack deterioration levels including crack level 75% and crack level 100% are considered. The gear crack levels ranged from 0% to 100% with an increase of 25% are used to describe the whole gear crack level deterioration. Because the number of the gear crack levels increase, the gear crack level identification problem under different loads and speeds becomes complicated and difficult. Therefore, it is necessary to propose an advanced gear crack identification method.

The rest of this paper is organized as follows. The proposed method for identification of different gear crack levels under different loads and speeds is introduced in Section 2. Three experiments are investigated in Section 3 to illustrate how the proposed method works and comparisons with other existing gear crack level identification methods are conducted. Conclusions are drawn in Section 4.

2. The Proposed Method for Identification of Different Gear Crack Levels under Different Motor Speeds and Loads

The proposed method for identification of five different gear crack levels under different loads and speeds is summarized in Figure 1, where the mathematical formulas of the proposed method are introduced in the following subsections.

First, to represent different gear crack levels, statistical features must be extracted from a gear vibration signal. Traditionally, statistical features are directly extracted from the temporal gear vibration signal and its corresponding frequency spectrum. The frequency spectrum is obtained by conducting Fourier transform on the temporal gear vibration signal. These statistical features directly extracted from the gear vibration signal and its corresponding frequency spectrum can be regarded as the "global" statistical features. These "global" statistical features are useful if the signal-to-noise ratio of the gear vibration signal is high. In other words, the fault features caused by a gear crack can be clearly found in the gear vibration signal and its corresponding frequency spectrum. However, besides the fault features caused by a gear crack, there are many noises and unknown vibration components existing in the gear vibration signal. It is necessary to enhance the signal-to-noise ratio of the gear vibration signal before the statistical features are extracted. Unlike Fourier transform, which aims to decompose the gear vibration signal to the sum of globally complex exponentials, a continuous wavelet transform uses inner product operation to measure the local similarity between a gear vibration signal and a wavelet mother function. It should be noted that the wavelet mother function is a locally oscillated analyzing function and can be shifted and scaled. The smaller the scale, the more compressed the wavelet mother function. The larger the scale, the more stretched the wavelet mother function. The continuous wavelet transform at different scales facilitates detecting the different local characteristics of the gear vibration signal, such as the features generated by a gear crack. Therefore, in this paper, the continuous wavelet transform at different scales, namely, Scalogram [27, 28], is conducted on the gear vibration signal to highlight the local gear fault features. If statistical features are extracted from the continuous wavelet coefficients, these statistical features can be regarded as "local" statistical features and can be used to better represent different gear crack levels. Additionally, it is not difficult to find that the resulting wavelet coefficients are so redundant that the statistical features extracted from the resulting wavelet coefficients are redundant because the wavelet functions at some of these different scales have very similar shapes.

10 popular statistical features, including mean, standard deviation, root mean square, peak, skewness, kurtosis, crest factor, clearance factor, shape factor, and impulse factor, are applied to the Scalogram at the scales ranged from 1 to 45
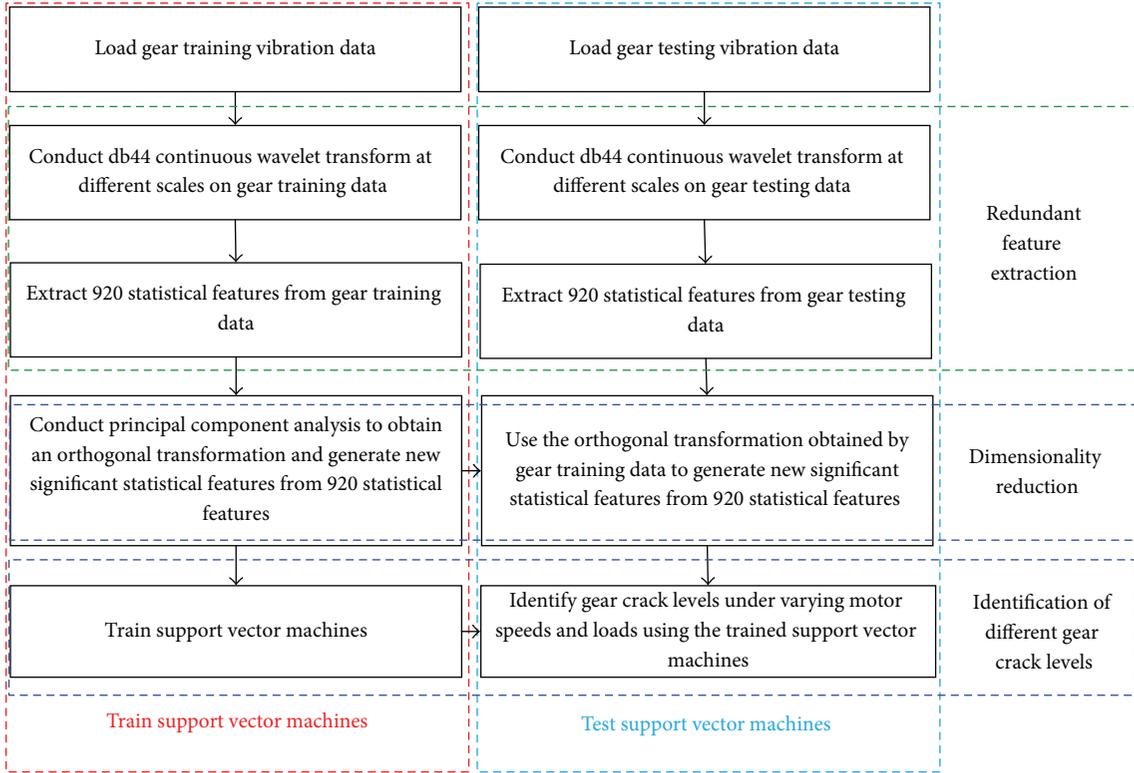
FIGURE 1: The flowchart of the proposed method for identification of different gear crack levels under different motor speeds and loads.

to generate 920 dimensional redundant statistical features for the representation of the gear crack level. The dimensionality of the redundant statistical features is calculated as follows. Based on the above ten statistical features and the Scalogram at the scales ranged from 1 to 45, $45 \times 10 = 450$ redundant statistical features are extracted. Besides the statistical features extracted from the Scalogram, it is necessary to extract the statistical features from the frequency spectra of the Scalogram, which exhibit different gear fault signatures from the Scalogram [29]. Consequently, there are $450 \times 2 = 900$ redundant statistical features extracted from the Scalogram and its frequency spectra. Considering 20 more "global" statistical features directly extracted from the original gear vibration signal and its corresponding frequency spectrum, there are $900 + 20 = 920$ redundant statistical features in all. Such redundant statistical features characterize the "global" and "local" features of the gear vibration signal. Moreover, according to our literature review, the use of the high-dimensional redundant features for identification of different gear faults is rarely reported.

Second, compared with the statistical features used in other intelligent gear crack level identification methods [24–26], in which only 10 to 30 statistical features were extracted, these 920 statistical features extracted by the proposed method are highly redundant and they are superhigh dimensional so that it is impossible to directly use all of the redundant statistical features to train and test statistical models. Before any statistical model is used, it is necessary to reduce

the dimensionality of the 920 redundant statistical features. There are many dimensionality reduction methods including linear and nonlinear methods [30]. For the nonlinear dimensionality methods, they have some disadvantages listed as follows. First, the calculation efficiencies of the nonlinear dimensionality methods are low. Second, they need much computer memory. Otherwise, the nonlinear dimensionality methods fail to generate low-dimensional features due to lack of computer memories. Third, even though these nonlinear dimensionality methods could be used to process training data, mapping of testing data to a low-dimensional space, namely, out-of-sample extension, is still questionable and often results in distinct estimation errors. Besides, through some numerical and real case studies, [30] concluded that nonlinear dimensionality reduction methods are not outperforming the traditional linear dimensionality reduction methods, such as principal component analysis. Therefore, a simple and efficient linear dimensionality reduction method called principal component analysis [31] is employed in this paper to generate new significant low-dimensional statistical features, namely, principal components, to distinguish different gear crack levels under the different working conditions.

At last, to ensure high training and testing accuracies, support vector machine [32, 33] is utilized to identify different gear crack levels under the different working conditions because it is able to use the kernel trick to map these new significant statistical features to a high-dimensional feature space, where linear classification is possible.
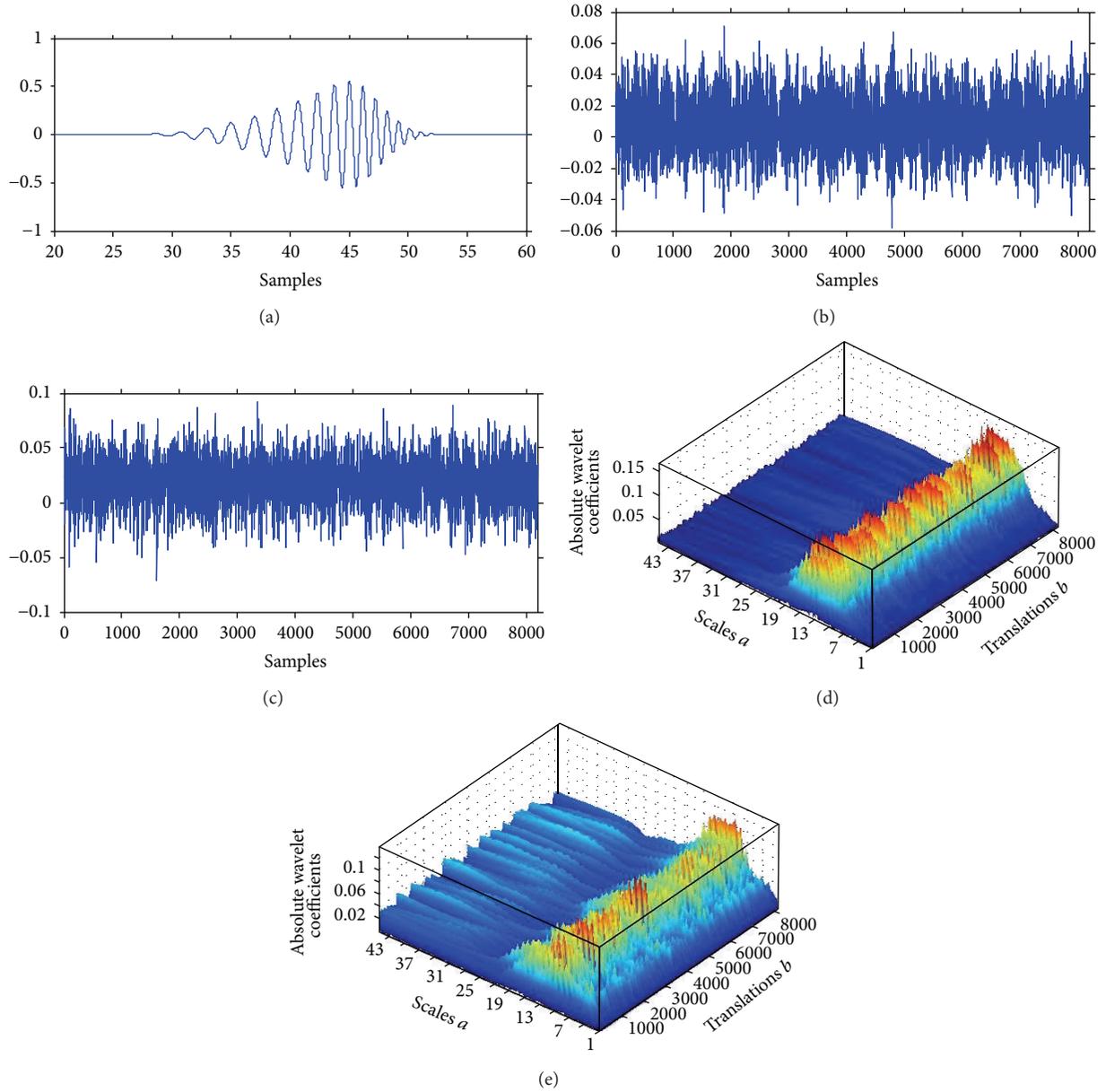
(a)

(b)

(c)

(d)

(e)

FIGURE 2: The analyses of the gear crack data: (a) the waveform of Daubechies 44; (b) the gear data with crack level 0%; (c) the gear data with crack level 100%; (d) the absolute wavelet coefficients of the gear data with crack level 0%; (e) the absolute wavelet coefficients of the gear data with crack level 100%.

2.1. Redundant Feature Extraction.

Continuous wavelet transform [27] aims to use an artificial wavelet mother function to calculate the inner product between a signal $x(t)$ and a wavelet mother function $\psi(t)$ at different scales $a > 0$ and translations b:

$$X(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^*\left(\frac{t-b}{a}\right) dt, \qquad (1)$$

where $*$ is the complex conjugate operator and $X(a,b)$ are the wavelet coefficients. From (1), it is seen that the continuous wavelet transform converts a one-dimensional signal to a two-dimensional signal (a time-scale representation), which generates redundant wavelet coefficients at different scales.

Besides, it is found that the wavelet mother function has a significant impact on the wavelet coefficients. Different wavelet mother functions result in different wavelet coefficients. Therefore, for the use of continuous wavelet transform, proper selection of the wavelet mother function becomes an open question. As mentioned in Introduction, Daubechies 44 is used in this paper and its temporal waveform is plotted in Figure 2(a). To show the redundant wavelet coefficients obtained by the continuous wavelet transform, the data with gear crack levels 0% and 100% are plotted in Figures 2(b) and 2(c), respectively, and their corresponding wavelet coefficients are plotted in Figures 2(d) and 2(e), respectively, in which the absolute wavelet coefficients are

used to enhance the three-dimensional visualization of the wavelet coefficients. From the results shown in Figures 2(d) and 2(e), it is seen that the each of the one-dimensional gear signals shown in Figures 2(b) and 2(c) is, respectively, transformed to a two-dimensional time-scale diagram, which are the redundant wavelet coefficients. As a result, compared with the original signals shown in Figures 2(b) and 2(c), the redundant wavelet coefficients provide more fault signatures. Following the use of the continuous wavelet transform, 10 statistical features shown in the following can be applied to quantify the wavelet coefficients at the scales ranged from 1 to 45 and their corresponding frequency spectra. Considering 20 more statistical features extracted from the original signal without being processed by the continuous wavelet transform and its frequency spectrum, there are $45 \times 10 + 45 \times 10 + 2 \times 10 = 920$ redundant statistical features in all. The statistical features used for quantifying wavelet coefficients at different scales (note that the same statistical features are used to quantify the frequency spectra of the wavelet coefficients at different scales and N is the length of the signal) are as follows.

Mean Value

$$F_1 = \frac{1}{N} \sum_{b=1}^{N} X(a,b). \tag{2}$$

Root Mean Square

$$F_2 = \sqrt{\frac{1}{N} \sum_{b=1}^{N} X(a,b)^2}. \tag{3}$$

Standard Deviation

$$F_3 = \sqrt{\frac{1}{N-1} \sum_{b=1}^{N} \left(X(a,b) - \frac{1}{N} \sum_{b=1}^{N} X(a,b) \right)^2}. \tag{4}$$

Skewness

$$F_4 = \frac{(1/N) \sum_{b=1}^{N} \left(X(a,b) - (1/N) \sum_{b=1}^{N} X(a,b) \right)^3}{\left(\sqrt{(1/N) \sum_{i=1}^{N} \left(x_i - (1/N) \sum_{i=1}^{N} x_i \right)^2} \right)^3}. \tag{5}$$

Kurtosis

F_5

$$= \frac{(1/N) \sum_{b=1}^{N} \left(X(a,b) - (1/N) \sum_{b=1}^{N} X(a,b) \right)^4}{\left(\sqrt{(1/N) \sum_{b=1}^{N} \left(X(a,b) - (1/N) \sum_{b=1}^{N} X(a,b) \right)^2} \right)^4}. \tag{6}$$

Crest Factor

$$F_6 = \frac{\max(|X(a,b)|)}{\sqrt{(1/N) \sum_{b=1}^{N} X(a,b)^2}}. \tag{7}$$

Clearance Factor

$$F_7 = \frac{\max(|X(a,b)|)}{\left((1/N) \sum_{b=1}^{N} \sqrt{|X(a,b)|} \right)^2}. \tag{8}$$

Shape Factor

$$F_8 = \frac{\sqrt{(1/N) \sum_{b=1}^{N} X(a,b)^2}}{(1/N) \sum_{b=1}^{N} |X(a,b)|}. \tag{9}$$

Impulse Factor

$$F_9 = \frac{\max(|X(a,b)|)}{\sqrt{(1/N) \sum_{b=1}^{N} |X(a,b)|}}. \tag{10}$$

Maximum

$$F_{10} = \max_b (X(a,b)). \tag{11}$$

2.2. Dimensionality Reduction. In the previous section, the superhigh-dimensional statistical features are extracted based on the use of the continuous wavelet transform. The dimensionality of the statistical features is 920. The direct use of all 920 statistical features for identification of different gear crack levels under the different working conditions will lead to the curse of dimensionality [34], which means that the number of the data used for supporting any result grows with the dimensionality exponentially. To relieve this problem, the dimensionality reduction is required to be conducted prior to the use of support vector machines. As discussed at the beginning of Section 2, in this paper, principal component analysis is chosen and its fundamental is introduced in the following [31]. Suppose that L training samples for different gear crack levels are obtained. Based on the L training samples and the 920 statistical features extracted from each sample, $L \times 920$ feature matrix is constructed as follows:

$$\mathbf{F} = \begin{bmatrix} F_{1,1}, F_{1,2}, \ldots, F_{1,920} \\ F_{2,1}, F_{2,2}, \ldots, F_{2,920} \\ \vdots \\ F_{k,1}, F_{k,2}, \ldots, F_{k,920} \\ \vdots \\ F_{L,1}, F_{L,2}, \ldots, F_{L,920} \end{bmatrix}. \tag{12}$$

From (12), it is seen that the statistical features used in (12) are very redundant. Principal component analysis aims to generate the significant new statistical features from the high-dimensional space and form a low-dimensional orthogonal space to express different gear crack levels. Each of the new generated features is called a principal component. Additionally, the first principal component has the greatest variance. The second principal component has the second greatest variance, and so on. Suppose that each column or feature of (12) has a zero mean and a unit variance. To achieve the above statement, the following optimization problem is constructed:

$$\mathbf{w} = \text{argmax} \left(\|\mathbf{Fw}\|^2 \right) = \text{argmax} \left(\mathbf{w}^T \mathbf{F}^T \mathbf{Fw} \right)$$
$$\text{subject to } \|\mathbf{w}\| = 1, \tag{13}$$

where T is the transpose operator. The Lagrange function of (13) is built as follows:

$$R(\mathbf{w}, \lambda) = \mathbf{w}^T \mathbf{F}^T \mathbf{F} \mathbf{w} - \lambda \mathbf{w}^T \mathbf{w}, \tag{14}$$

where λ is the Lagrange multiplier. The first partial derivative of R with respect to $\mathbf{w}$ is obtained as follows:

$$\frac{\delta R(\mathbf{w}, \lambda)}{\delta \mathbf{w}} = 2\mathbf{F}^T \mathbf{F} \mathbf{w} - 2\lambda \mathbf{w}. \tag{15}$$

If (15) is set to zero, the relationship between $\mathbf{w}$ and the Lagrange multiplier is the eigenfunction of the symmetric matrix $\mathbf{F}^T \mathbf{F}$ and it is written as follows:

$$\mathbf{F}^T \mathbf{F} \mathbf{w} = \lambda \mathbf{w}. \tag{16}$$

Suppose that the first column of $\mathbf{w}$ is the eigenvector corresponding to the largest eigenvalue, the second column of $\mathbf{w}$ is the eigenvector corresponding to the second largest eigenvalue, and so on. Consequently, the feature matrix shown in (12) can be mapped to a new space consisting of the principal components:

$$\mathbf{t} = \mathbf{F} \mathbf{w}, \tag{17}$$

where the first column of $\mathbf{t}$ is the first principal component t_1, the second column of $\mathbf{t}$ is the second principal component t_2, and so on. For fair comparison with the other existing gear crack level identification methods reported in [24], where seven statistical features were selected from 25 statistical features, in this paper, the first seven principal components obtained by the proposed method are used to train and test support vector machines. Besides, the testing data can be directly mapped to the new principal component space by using the established linear transformation matrix $\mathbf{w}$.

2.3. Identification of Different Gear Crack Levels under Different Working Conditions. To automatically identify different gear crack levels under different motor speeds and loads, support vector machine [32] is used in this paper and it is a kind of supervised learning method which has been widely investigated in the past years for solving various classification and regression problems. Given the training data of two different gear crack levels $T = \{(\mathbf{y}_i, z_i) \mid \mathbf{y}_i \in \mathbb{R}^p, z_i \in \{-1, 1\}\}_{i=1}^n$, where $\mathbf{y}_i$ is the statistical feature vector with a dimensionality of p and z is the binary classification label, if the training data are linearly separable, a linear decision function can be determined by solving the following optimization problem [32]:

$$\arg\min \quad \frac{1}{2} \|\boldsymbol{\omega}\|^2 \tag{18}$$
$$\text{subject to} \quad z_i (\boldsymbol{\omega} \cdot \mathbf{y}_i - c) \geq 1,$$

where $\boldsymbol{\omega}$ is the normal vector to the decision function, $\cdot$ is the dot product, and c is the offset of the decision function. The objective function of (18) aims to maximize the distance between two hyperplanes, where there are no training data between them. It means that the linear decision function

creates the maximum distance between the linear decision function and the nearest training data.

Considering the noise with the slack variables ξ_i and the error penalty constant C, (18) is revised as [32]

$$\arg\min_{\boldsymbol{\omega}, \xi} \quad \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^{n} \xi_i \tag{19}$$
$$\text{subject to} \quad z_i (\boldsymbol{\omega} \cdot \mathbf{y}_i - c) \geq 1 - \xi_i, \quad \xi \geq 0.$$

Solving (19) is equivalent to solving the following Lagrangian problem with Lagrange multipliers α_i and β_i:

$$\arg\min_{\boldsymbol{\omega}, \xi, \alpha, \beta, b} \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^{n} \xi_i$$
$$- \sum_{i=1}^{n} \alpha_i \left[z_i (\boldsymbol{\omega} \cdot \mathbf{y}_i - c) - 1 + \xi_i \right] - \sum_{i=1}^{n} \beta_i \xi_i, \tag{20}$$
$$\alpha_i, \beta_i \geq 0.$$

Taking the derivatives of (20) with respect to $\boldsymbol{\omega}$ and c, respectively, it is derived that

$$\boldsymbol{\omega} = \sum_{i=1}^{n} \alpha_i z_i \mathbf{y}_i, \quad \sum_{i=1}^{n} \alpha_i z_i = 0. \tag{21}$$

Substituting (21) into (20), (20) becomes a dual quadratic optimization problem [32]:

$$\arg\max_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j z_i z_j \mathbf{y}_i \cdot \mathbf{y}_i \tag{22}$$
$$\text{subject to} \quad C \geq \alpha_i \geq 0, \quad i = 1, \ldots, n, \quad \sum_{i=1,}^{n} \alpha_i z_i = 0.$$

After solving (22), the linear decision function is obtained as follows:

$$\text{sign} \left(\sum_{i=1,}^{n} \alpha_i z_i (\mathbf{y}_i \cdot \mathbf{y}) + c \right). \tag{23}$$

To extend the linear classification problem to the nonlinear classification problem, kernel trick can be used to map the training data to a high-dimensional space, where the linear classification problem is possibly solved, prior to the establishment of (23). Considering the kernel function, (23) can be revised as [32]

$$\text{sign} \left(\sum_{i=1}^{n} \alpha_i z_i K(\mathbf{y}_i, \mathbf{y}) + c \right), \tag{24}$$

where $K(\mathbf{y}_i, \mathbf{y})$ is the linear or nonlinear kernel function which should satisfy Mercer's theorem. There are three popular kernel functions including linear, polynomial, and Gaussian radial basis functions [32]. Generally, the Gaussian radial basis function is the preferable choice for the use of the support vector machine because, unlike the linear kernel,
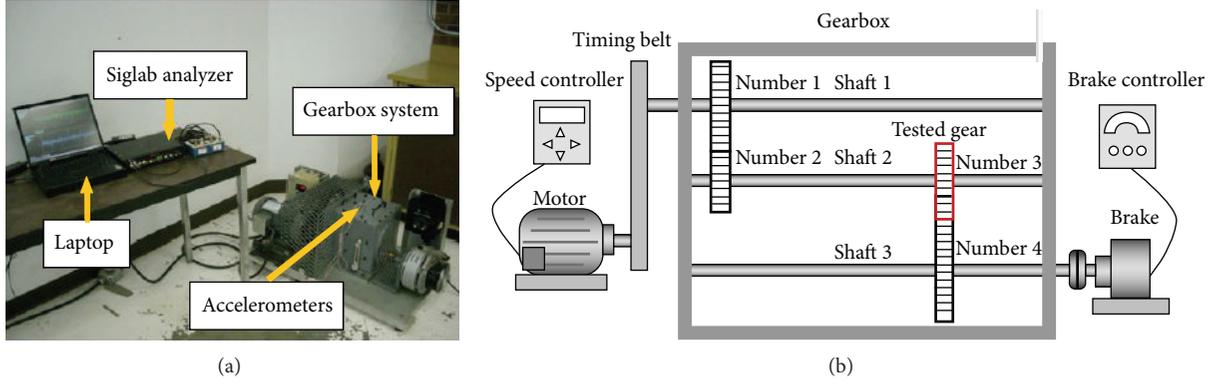
(a)

(b)

FIGURE 3: Experimental setup [24]: (a) a real picture; (b) a sketch map.

TABLE 1: The geometries of different gear crack levels.

Gear crack level	Depth (mm)	Width (mm)	Thickness (mm)	Angle (degree)
0%	0	0	0	0
25%	$0.25d$	$0.25c$	0.4	
50%	$0.5d$	$0.5c$	0.4	45
75%	$0.75d$	$0.75c$	0.4	
100%	d	c	0.4	

it is able to handle the nonlinear classification problem. Furthermore, the number of hyperparameters used in the Gaussian radial basis function is less than that used in the polynomial function. Therefore, in this paper, the Gaussian radial basis function is chosen and (24) is revised as

$$\text{sign}\left(\sum_{i=1}^{n}\alpha_i z_i \exp\left(-\gamma\left\|\mathbf{y}_i - \mathbf{y}\right\|^2\right) + c\right), \quad (25)$$

where γ is the kernel parameter and $\|\cdot\|$ is the modulus of the feature vector. To classify five different gear crack levels, the popular one-against-all strategy and one-against-one strategy can be used [32, 33].

3. Instance Studies and Comparisons with Other Existing Advanced Gear Crack Level Identification Methods

3.1. Experimental Platform. In this paper, one of the coauthors [5] designed the experiments to collect the different gear crack level data under different motor speeds and loads from the experimental setup shown in Figures 3(a) and 3(b). The experimental setup included a gearbox, a 3-hp ac motor, which was used to drive the input shaft of the gearbox, and a magnetic brake, which was used to provide different loads. Four different rotation motor speeds including 1200 rpm, 1400 rpm, 1600 rpm, and 1800 rpm and three different loads including no load, half load, and full load were used. Gears 1, 2, 3, and 4 had 48, 16, 24, and 40 teeth, respectively. Gear 3 was the tested gear used in the experimental setup. Some artificial gear crack levels denoted as crack levels 0%, 25%, 50%, 75%, and 100% were produced to simulate all gear crack deterioration levels and their geometries are tabulated in

Table 1. The crack thickness was 0.4 mm because the available thinnest knife in the coauthor's lab was 0.4 mm. For the four gear crack levels, the crack depths were 0.25, 0.5, 0.75, and 1 times the half of the chordal tooth thickness, respectively, because the tooth will break rapidly when the crack depth is more than half of the chordal tooth thickness. Here, the chordal tooth thickness $2d$ is the tooth thickness at the pitch line. The crack widths were 0.25, 0.5, 0.75, and 1 times the face width c equal to 25 mm, respectively. The crack angle was 45 degree. The diagrammatic sketches of the chordal tooth thickness, crack width, and crack angle are plotted in Figures 4(a) and 4(b), respectively. The different gear crack levels are shown in Figures 5(a) to 5(d), respectively. The vibration signals were measured by two acceleration sensors, which were produced by PCB Electronics with model number 352C67. These two sensors were mounted on the casing of the gearbox in the vertical and horizontal directions, respectively. In [5], it was reported that the vertical direction is more sensitive to identification of the different crack levels. Therefore, the vibration signals collected from the vertical direction were used in this paper. The sampling frequency was set to 5120 and, for each sample, the sampling points were 8192. For each working condition, two samples were collected. Consequently, there were 2 samples ×3 loads ×4 speeds ×5 gear health conditions = 120 samples in all. By considering the different combination of different motor speeds and loads, the similar three experiments designed in [24] were used in this paper. Compared with the experiments designed in [24], two more gear crack levels including 75% and 100% were considered in this paper, which makes the gear crack level identification difficult. The designed experiments used in this paper are tabulated in Table 2.
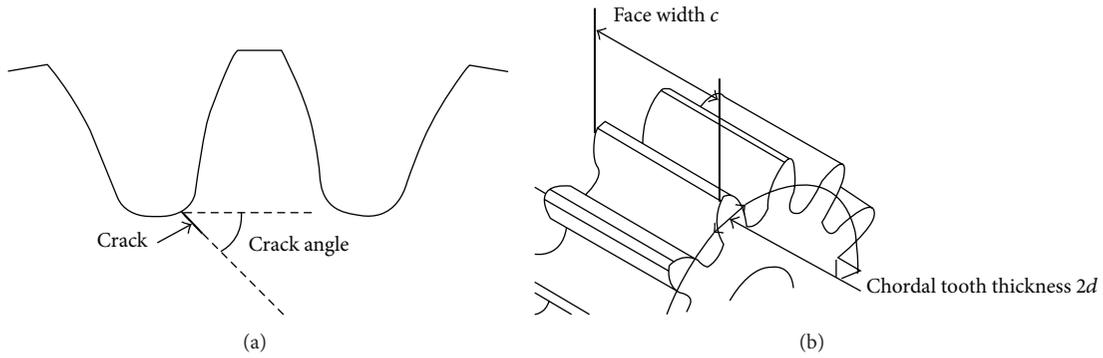
(a)

(b)

FIGURE 4: The diagrammatic view [24] of (a) the crack angle and (b) the face width and the chordal tooth thickness.
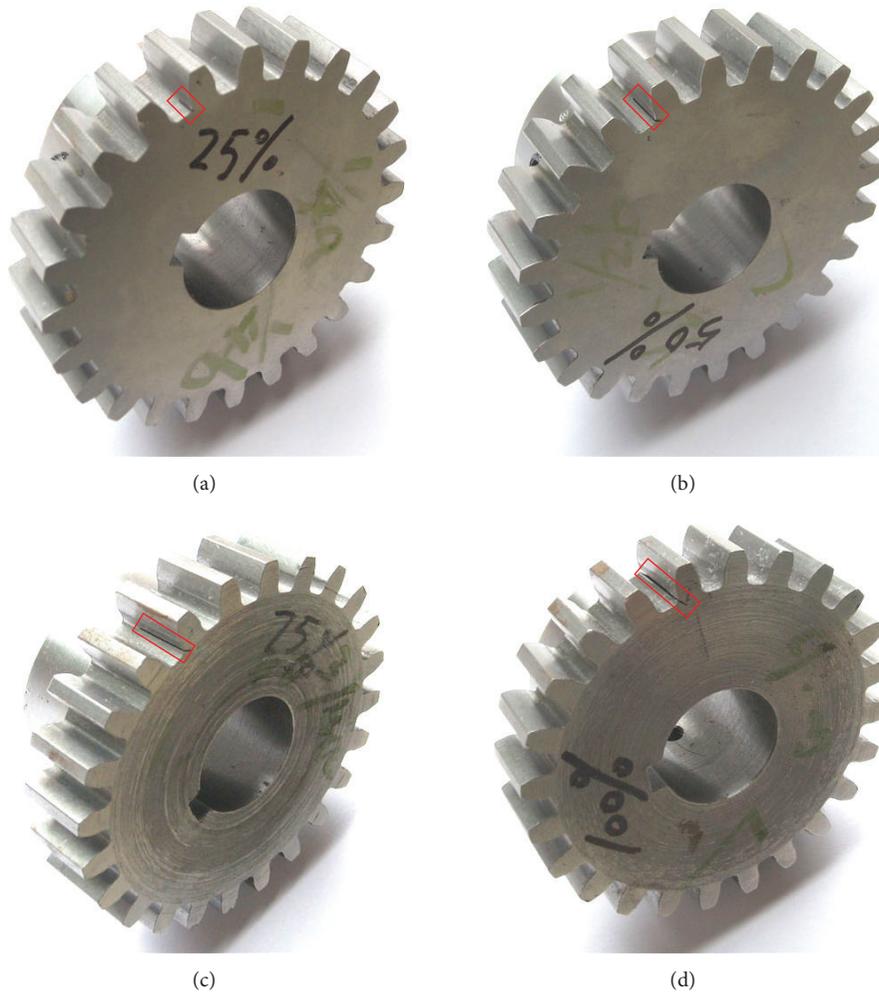

(a)

(b)

(c)

(d)

FIGURE 5: The different gear crack levels [24]: (a) crack level 25%; (b) crack level 50%; (c) crack level 75%; (d) crack level 100%.

In the first experiment, for each gear crack level, 24 samples were collected from the machine under 4 different motor speeds and 3 different loads. Therefore, $24 \times 4 \times 3 \times 5$ samples were collected in all. Then, half of the samples ($12 \times 4 \times 3 \times 5$) were used in the training phase. The other samples ($12 \times 4 \times 3 \times 5$) were used in the testing phase.

In the training phase of the second experiment, for each gear crack level, 12 samples were collected from the machine under 2 different motor speeds and 3 different loads. Therefore, $12 \times 2 \times 3 \times 5$ samples were collected. Then, in the testing phase, for each gear crack level, other 12 samples were collected from the machine under other 2 different motor

TABLE 2: The design of the three experiments (N, H, and F denotes no load, half load, and full load, resp.).

Experiment	Number of training/testing samples	Gear crack levels of training/testing	Motor speeds of training/testing samples (rpm)	Loads of training/testing samples	Classification labels
	12/12	0%/0%	1200–1800/1200–1800	N, H, F/N, H, F	1
	12/12	25%/25%	1200–1800/1200–1800	N, H, F/N, H, F	2
1	12/12	50%/50%	1200–1800/1200–1800	N, H, F/N, H, F	3
	12/12	75%/75%	1200–1800/1200–1800	N, H, F/N, H, F	4
	12/12	100%/100%	1200–1800/1200–1800	N, H, F/N, H, F	5
	12/12	0%/0%	1200, 1600/1400, 1800	N, H, F/N, H, F	1
	12/12	25%/25%	1200, 1600/1400, 1800	N, H, F/N, H, F	2
2	12/12	50%/50%	1200, 1600/1400, 1800	N, H, F/N, H, F	3
	12/12	75%/75%	1200, 1600/1400, 1800	N, H, F/N, H, F	4
	12/12	100%/100%	1200, 1600/1400, 1800	N, H, F/N, H, F	5
	8/16	0%/0%	1200–1800/1200–1800	N/H, F	1
	8/16	25%/25%	1200–1800/1200–1800	N/H, F	2
3	8/16	50%/50%	1200–1800/1200–1800	N/H, F	3
	8/16	75%/75%	1200–1800/1200–1800	N/H, F	4
	8/16	100%/100%	1200–1800/1200–1800	N/H, F	5

speeds and 3 different loads. Therefore, $12 \times 2 \times 3 \times 5$ samples were collected. The design of experiment 2 aims to investigate the influence of different motor speeds on the prediction accuracies of the proposed method.

In the training phase of the third experiment, for each gear crack level, 8 samples were collected from the machine under 4 different motor speeds and 1 load. Therefore, $8 \times 4 \times 1 \times 5$ samples were collected. Then, in the testing phase, for each gear crack level, 16 samples were collected from the machine under another 4 different motor speeds and 2 different loads. Therefore, $16 \times 4 \times 2 \times 5$ samples were collected. The design of experiment 3 aims to investigate the influence of different loads on the prediction accuracies of the proposed method.

3.2. Comparisons of the Proposed Method with the Four Existing Gear Crack Level Identification Methods Reported in [24]. For experiment 1, support vector machines are trained and tested by the five different gear crack levels under four different motor speeds and the three different loads. It means that all different working conditions are considered in the training of support vector machines, which makes the different crack levels relatively easy compared with experiments 2 and 3, in which only part of the working conditions are used to train support vector machines. In vibration analysis, different motor speeds and loads have great influence on the amplitudes and waveforms of vibration signals, the changes of which result in the changes of the statistical feature values. Therefore, the design of experiments 2 and 3 makes the different gear crack level identification complicated. For the use of SVM, Gaussian radial basis function was used and its kernel width was optimized to 0.14. The prediction accuracies of the proposed method and the four existing advanced gear crack level identification methods reported in [24] are tabulated in Table 3, where method 1 is the direct use of K-nearest

TABLE 3: Comparisons of the prediction accuracies obtained by using the proposed method and the four methods reported in [16] (unit: %). (Note: methods 1 to 4 are used to distinguish three gear health conditions, while the proposed method is used to classify five gear health conditions.)

Prediction accuracy	Experiment 1	Experiment 2	Experiment 3
K-nearest neighbor without statistical feature selection [16]	89.58	87.42	87.67
K-nearest neighbor with random statistical feature selection [16]	80.17	77.29	78.08
K-nearest neighbor with Euclidean distance evaluation [16]	96.53	99.15	92.10
The weighted K-nearest neighbor method [16]	96.68	99.61	92.62
The proposed method	100	100	100

neighbor without statistical feature selection, method 2 is K-nearest neighbor with random statistical feature selection, method 3 is K-nearest neighbor with Euclidean distance evaluation used for statistical feature selection and without weighting technique, and method 4 is the weighted K-nearest neighbor method proposed in [24]. From the result shown in Table 3, it is found that the proposed method has the highest prediction accuracies among all methods. Besides, the prediction accuracies achieve 100%.

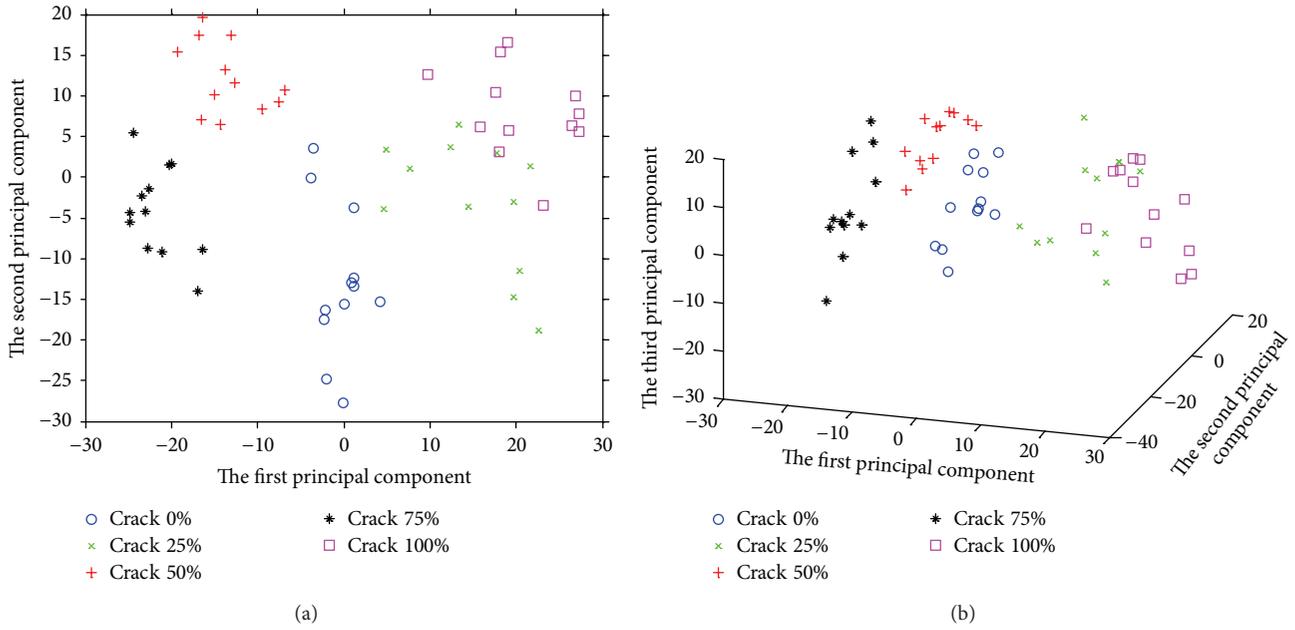The reasons why the proposed method has such high prediction accuracies are explained as follows. First, the

FIGURE 6: The principal components of the training data for experiment 1: (a) the first two principal components; (b) the first three principal components.
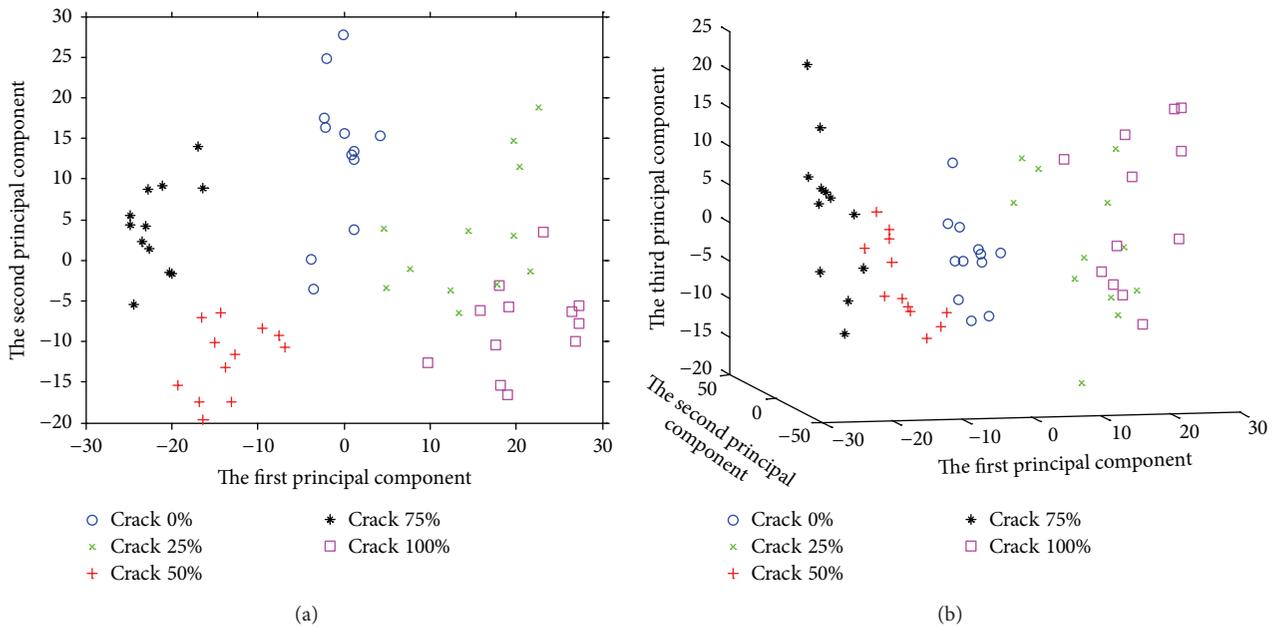

FIGURE 7: The principal components of the testing data for experiment 1: (a) the first two principal components; (b) the first three principal components.

statistical features used in the proposed method are very redundant and their number is high to 920. The redundant statistical features provide more gear crack fault signatures. Second, the principle components are the new significant statistical features generated from the redundant statistical features to represent different gear crack levels. For experiment 1, the first two principal components and the first

three principal components of the training data are plotted in Figures 6(a) and 6(b), respectively. From the result shown in Figure 6(a), it is seen that only the gear crack levels 25% and 100% are overlapped with each other. From the result shown in Figure 6(b), the five different gear crack levels are separable in a three-dimensional principal component space. The first two principal components and the first
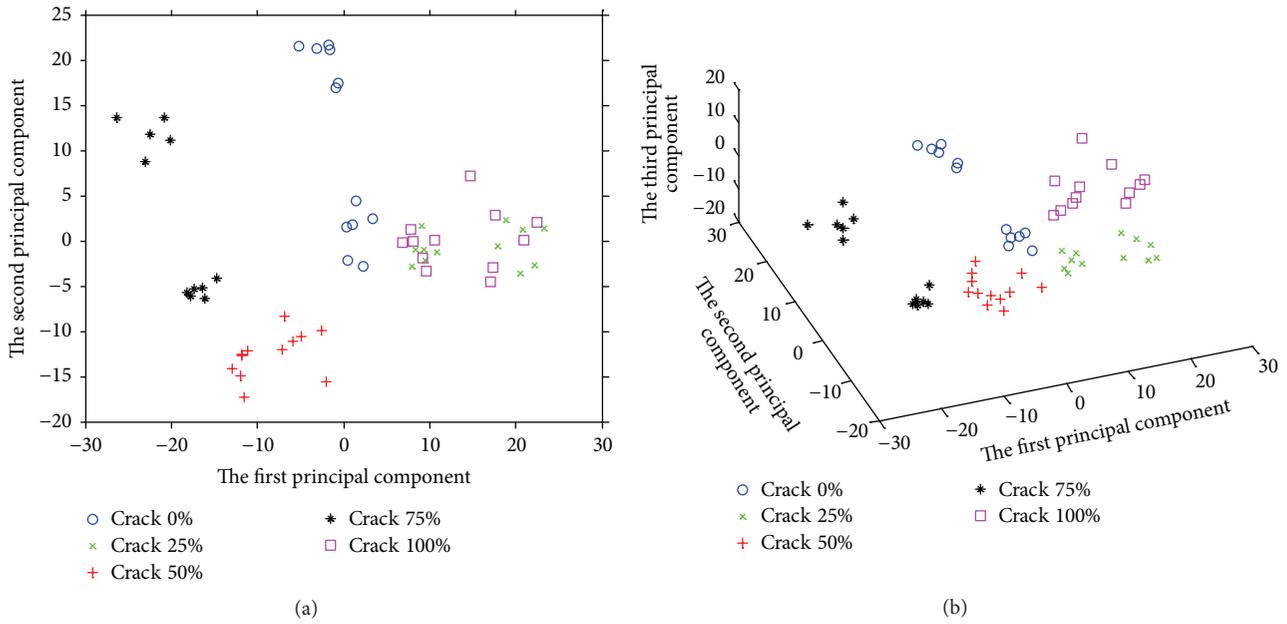
FIGURE 8: The principal components of the training data for experiment 2: (a) the first two principal components; (b) the first three principal components.
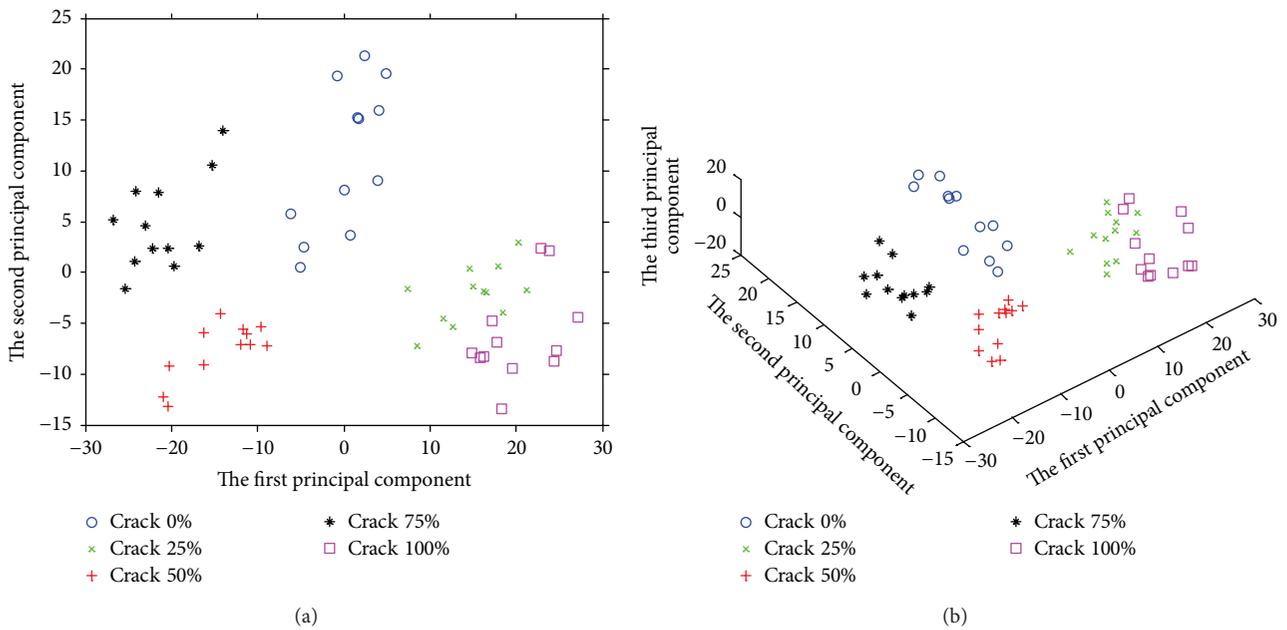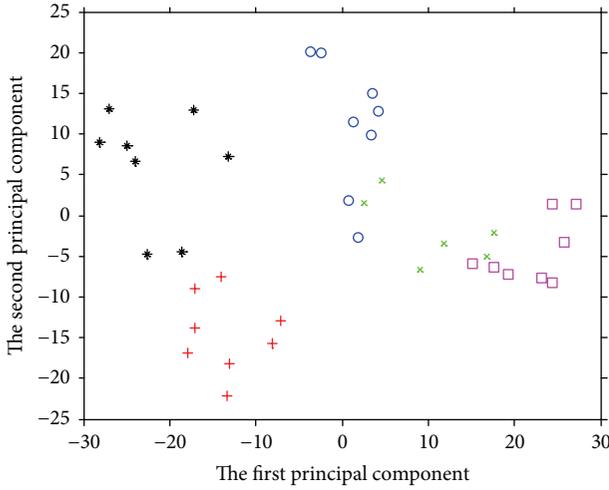

FIGURE 9: The principal components of the testing data for experiment 2: (a) the first two principal components; (b) the first three principal components.
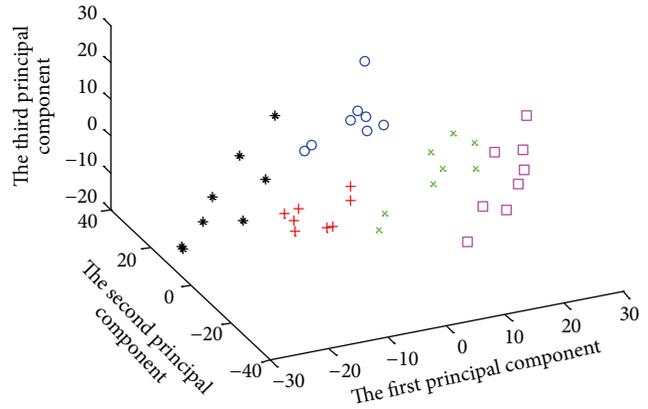
three principal components of the testing data are shown in Figures 7(a) and 7(b), respectively, where it is found that the principal components are the new significant statistical features to distinguish the five different gear crack levels. It can be inferred that, with the number of the principal components increasing, the five different gear crack levels are well separable. For experiments 2 and 3, the first two

and three principal components of training data and testing data are plotted in Figures 8–11, respectively, where it is found that the five different gear crack levels are separable in three-dimensional principal component space. At last, because support vector machine uses the kernel trick to map a low-dimensional feature space to a high-dimensional feature space, where it is possible for the features to be
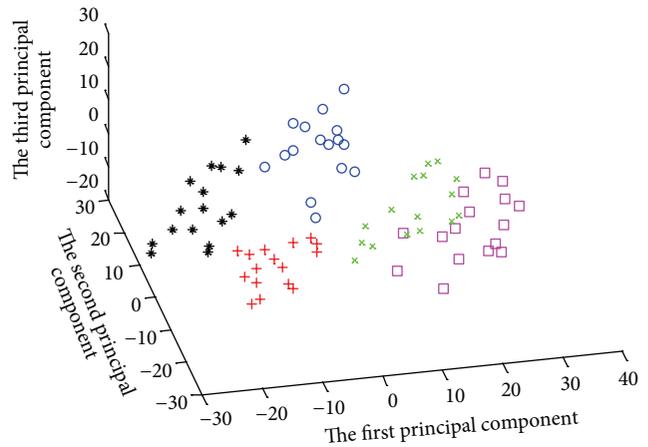
FIGURE 10: The principal components of the training data for experiment 3: (a) the first two principal components; (b) the first three principal components.
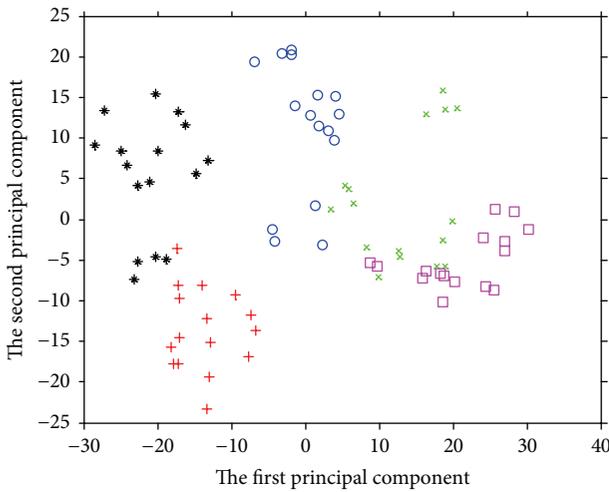
FIGURE 11: The principal components of the testing data for experiment 3: (a) the first two principal components; (b) the first three principal components.

linearly separated in a high-dimensional space, this technique enhances the prediction accuracy of the five different gear crack levels.

4. Conclusions

In this paper, an intelligent gear crack level identification method under different working conditions is proposed.

The major idea of the proposed method is to use superhigh-dimensional redundant statistical features to represent five different gear crack levels under different working conditions. The number of the redundant features is high to 920, which is obtained by using 10 statistical features extracted from the Scalogram and its frequency spectra. Then, to reduce the dimensionality of the redundant statistical features and relieve the curse of dimensionality, principal component

analysis is performed on the redundant statistical features to generate new significant statistical features. At last, support vector machines with a Gaussian radial basis function are used to identify the five different gear crack levels under the four different motor speeds and three different loads. The comparisons with the four existing gear crack level identification methods show that the proposed method has the highest prediction accuracies among all existing methods and the prediction accuracies obtained by the proposed method are high to 100% for the three different experiments. The reasons for such high prediction accuracies of the proposed method are summarized as follows.

First, extraction of high-dimensional redundant statistical features from the continuous wavelet transform at different scales is helpful to mine more gear fault signatures under different working conditions because these high-dimensional redundant statistical features can be used to reflect the global and local characteristics of the gear crack level data. Second, the new significant statistical features, namely, the principal components, generated from these high-dimensional redundant statistical features are useful to distinguish different crack levels under different working conditions. From the principal components shown in Figures 6 to 11, it is obvious to find that as the number of the principal components increases from 1 to 3, the five different crack levels are well separable. At last, support vector machine uses the kernel trick to map the principal components to a high-dimensional feature space, where the separation of the five different crack levels is more notable.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration*, vol. 324, no. 3–5, pp. 1141–1157, 2009.

[2] D. Wang, P. W. Tse, W. Guo, and Q. Miao, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis," *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.

[3] M. Yang and V. Makis, "ARX model-based gearbox fault detection and localization under varying load conditions," *Journal of Sound and Vibration*, vol. 329, no. 24, pp. 5209–5221, 2010.

[4] X. Wang, V. Makis, and M. Yang, "A wavelet approach to fault diagnosis of a gearbox under varying load conditions," *Journal of Sound and Vibration*, vol. 329, no. 9, pp. 1570–1585, 2010.

[5] X. Fan, Y. Wang, W. Li, S. Wu, and M. J. Zuo, "Experiment design of gear crack growth," Tech. Rep., Reliability Research Lab, Department of Mechanical Engieering, University of Alberta, Edmonton, Canada, 2005.

[6] S. Wu, M. J. Zuo, and A. Parey, "Simulation of spur gear dynamics and estimation of fault growth," *Journal of Sound and Vibration*, vol. 317, no. 3–5, pp. 608–624, 2008.

[7] E. Bechhoefer and M. Kingsley, "A review of time synchronous average algorithms," in *Proceedings of the Annual Conference of the Prognostics and Health Management Society (PHM '09)*, October 2009.

[8] W. J. Wang and P. D. McFadden, "Decomposition of gear motion signals and its application to gearbox diagnostics," *Journal of Vibration and Acoustics*, vol. 117, no. 3, pp. 363–369, 1995.

[9] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 199–221, 2004.

[10] R. X. Gao and R. Yan, *Wavelets: Theory and Applications for Manufacturing*, Springer, New York, NY, USA, 2010.

[11] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.

[12] W. Bartelmus and R. Zimroz, "A new feature for monitoring the condition of gearboxes in non-stationary operating conditions," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1528–1534, 2009.

[13] W. Bartelmus, "Root cause and vibration signal analysis for gearbox condition monitoring," *Insight—Non-Destructive Testing and Condition Monitoring*, vol. 50, no. 4, pp. 195–201, 2008.

[14] W. Bartelmus, "Gearbox damage process," *Journal of Physics: Conference Series*, vol. 305, no. 1, Article ID 012029, 2011.

[15] W. Bartelmus, "New focus on gearbox condition monitoring for failure prevention technology," *Key Engineering Materials*, vol. 588, pp. 184–191, 2014.

[16] W. Bartelmus, "Editorial statement," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 1–4, 2013.

[17] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.

[18] J. Rafiee and P. W. Tse, "Use of autocorrelation of wavelet coefficients for fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1554–1572, 2009.

[19] H. Öztürk, M. Sabuncu, and I. Yesilyurt, "Early detection of pitting damage in gears using mean frequency of scalogram," *Journal of Vibration and Control*, vol. 14, no. 4, pp. 469–484, 2008.

[20] S. J. Loutridis, "Instantaneous energy density as a feature for gear fault detection," *Mechanical Systems and Signal Processing*, vol. 20, no. 5, pp. 1239–1253, 2006.

[21] Z. Feng, M. J. Zuo, and F. Chu, "Application of regularization dimension to gear damage assessment," *Mechanical Systems and Signal Processing*, vol. 24, no. 4, pp. 1081–1098, 2010.

[22] Z. Feng, M. J. Zuo, R. Hao, F. Chu, and M. El Badaoui, "Gear damage assessment based on cyclic spectral analysis," *IEEE Transactions on Reliability*, vol. 60, no. 1, pp. 21–32, 2011.

[23] S. J. Loutridis, "Self-similarity in vibration time series: application to gear fault diagnostics," *Journal of Vibration and Acoustics*, vol. 130, no. 3, Article ID 031004, 2008.

[24] Y. Lei and M. J. Zuo, "Gear crack level identification based on weighted K nearest neighbor classification algorithm," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1535–1547, 2009.

[25] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1419–1430, 2010.

[26] Y. Hai, K.-L. Tsui, and M. J. Zuo, "Gear crack level classification based on multinomial logit model and cumulative link model," in *Proceedings of the 3rd Annual IEEE Prognostics and System Health Management Conference (PHM '12)*, pp. 1–6, IEEE, Beijing, China, May 2012.

[27] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Elsevier/Academic Press, Amsterdam, The Netherlands, 3rd edition, 2009.

[28] Z. Peng, F. Chu, and Y. He, "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram," *Journal of Sound and Vibration*, vol. 253, no. 5, pp. 1087–1100, 2003.

[29] R. Yan and R. X. Gao, "Multi-scale enveloping spectrogram for vibration analysis in bearing defect diagnosis," *Tribology International*, vol. 42, no. 2, pp. 293–302, 2009.

[30] L. J. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality reduction: a comparative review," *Journal of Machine Learning Research*, vol. 10, pp. 66–71, 2009.

[31] I. Jolliffe, *Principal Component Analysis*, Wiley Online Library, 2005.

[32] V. N. Vapnik, *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, New York, NY, USA, 1998.

[33] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, Berlin, Germany, 2009.