

Received October 16, 2018, accepted November 18, 2018, date of publication November 22, 2018,
 date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2882801

Gearbox Fault Diagnosis Based on a Novel Hybrid Feature Reduction Method

YU WANG^{ID}¹, SHUAI YANG^{ID}¹, AND RENÉ VINICIO SÁNCHEZ²

¹National Research Base of Intelligent Manufacturing, Chongqing Technology and Business University, Chongqing 400060, China

²GIDTEC, Universidad Politécnica Salesiana, Cuenca, Ecuador

Corresponding author: Shuai Yang (jerryyang@ctbu.edu.cn)

This work was supported in part by the Program of Chongqing Municipal Education Commission under Grant KJZH17123, in part by the National Key Research & Development Program of China under Grant 2016YFE0132200, in part by the Open Funding of Chongqing Technology and Business University under Grant KFJJ2017076, in part by the Research Start-Up Funds of Chongqing Technology and Business University under Grant 1856018, and in part by the Graduate Scientific Research Innovation Project of Chongqing Technology and Business University under Grant yjscxx2018-060-15.

ABSTRACT The dimensionality reduction of the high-dimensional feature space is a critical part for data preprocessing, which directly affects the accuracy of fault diagnosis. In this paper, a novel hybrid algorithm named principal component locally linear embedding (PCLLE) is introduced to compress the original high-dimensional feature. This approach combines the optimization objectives of the principal component analysis (PCA) and locally linear embedding (LLE), which attempts to find a mapping that meets the optimization goals of PCA and LLE at the same time. It is applied on the gearbox fault diagnosis. In the experiment, the extracted fault-sensitive feature is compressed by PCLLE method. Then, the compressed feature is embedded with five classifiers for fault detection. To evaluate the performance of the proposed new method, the traditional PCA and LLE methods are introduced for comparison. Experimental results show that the PCLLE algorithm has good performance during the classification process compared with the traditional PCA and LLE method.

INDEX TERMS Feature reduction, gearbox, fault diagnosis, principal component analysis, locally linear embedding, hybrid algorithm.

I. INTRODUCTION

The gearbox is a complex power system that contains mechanical components such as drive shafts, bearings, gears, and cabinet structures, which is used for changing the speed and power of the machinery [1]. Due to the harsh working environment, key components of the gearbox are prone to wear, fatigue and cracks etc [2]–[5]. Once a component in the gearbox fails, a chain reaction can occur, causing the transmission system to fail. According to statistics, the gear is the main fault component of the gearbox, accounting for 60% of the total number of faults in gearbox [6]. Therefore, fault monitoring and diagnosis of gear is crucial for the reliability maintenance of the entire gearbox.

Fault diagnosis for gearboxes is generally divided into the following three processes: (1) Acquisition of the operating signals under different fault conditions from sensors installed on gearbox. (2) Preprocessing of the original data and feature extraction. (3) Establishment of the fault identification model. The key process of fault diagnosis is to extract sensitive features of the original signal [7]–[9]. In order to obtain

sufficient fault information from the raw signal, the feature index is usually selected from the time domain, the frequency domain, and the time-frequency domain of the raw signal [10]–[11]. However, data redundancy or irrelevance may exist between features, resulting in problems such as slower calculation speed, overfitting, which will reduce the fault recognition accuracy of the classifier [12]. Therefore, it is necessary to reduce the dimensionality of the raw feature matrix to eliminate redundant information and enhance the classification effect.

A wealth of methods has been employed on the dimensionality reduction process of rotating machinery fault diagnosis. Principal Component Analysis (PCA) is currently the most widely used dimensionality reduction method. Based on the principle of variance maximization, PCA attempts to find a low-dimensional representation of original data through a linear mapping. It has shown great performance on fault diagnosis field [13]–[14]. However, PCA is a method for linear data, which is inapplicable for non-linear data. Thus, non-linear dimensionality reduction techniques were proposed.

Kernel Principal Component Analysis (KPCA) is an improved algorithm of PCA [15], [16]. It maps the input vector to a high-dimensional feature space through a kernel function, and uses PCA to calculate the principal component in a high-dimensional space. Isometric Feature Mapping (Isomap) is proposed by Tenenbaum as a nonlinear improvement of the Multidimensional scaling (MDS) in 2000. Through using the geodesic distance, the Euclidean distance in the MDS algorithm is replaced to maintain the inherent geometric properties of the data points [17]–[19]. Roweis proposed Locally Linear Embedding (LLE), tried to use the local linearity of the data to approximate the global linearity [20]. By assuming that adjacent sample points in the high-dimensional representation are also adjacent in the low-dimensional representation, Belkin and Niyogi proposed to use the concept of Graph Laplace to calculate the low-dimensional representation of the high-dimensional representation [21]–[22], which is named as Laplacian Eigenmaps (LE) in 2003. These algorithms mentioned above are unsupervised. In order to improve the classification accuracy, category label is introduced in some algorithms to achieve supervised dimensionality reduction, such as Linear Discriminant Analysis (LDA) and Kernel Linear Discriminant Analysis (KLDA) [23]–[24]. These algorithms essentially find the corresponding embedding map by low-dimensional manifolds in high-dimensional space [25].

All of those methods show unique merits in dimensionality reduction field. In order to combine the advantage belongs to different algorithms, hybrid algorithm is a great consideration. Some researchers combined different algorithms together and achieved excellent results in the experiment. Liu *et al.* [26] proposed an Incremental Supervised Local Linear Embedding (ISLLE) to integrate the LDA algorithm into the LLE algorithm in 2016. The proposed hybrid algorithm improved the performance of LLE through using label information. In this paper, another hybrid approach is proposed for dimensionality reduction. The main original contributions of this research are presented as follows:

- 1) A novel feature reduction approach named Principal Component Locally Linear Embedding (PCLLE) is introduced. Motivated by PCA and LLE, this novel hybrid method attempts to find a new mapping, which can transform the original high-dimensional feature matrix into a low-dimensional space while meets the optimization targets of PCA and LLE at the same time. This method combines the advantage of global and local dimensionality reduction techniques.
- 2) Fisher discriminant value is proposed to measure the theoretical separability of the compressed feature matrix in this paper. It serves as an effective technique to select the number of neighbors for each sample point, which is an important parameter in this proposed approach. The proposed method is utilized in gearbox fault diagnosis, which is combined with five frequently used classifiers to test the performance of the proposed method.

The rest of this paper is organized as follows. In Section 2, the hybrid dimensionality reduction method proposed in this paper are detailed. In Section 3, a set of data is introduced to validate the performance of the proposed method. In Section 4, the experiment on fault diagnosis of gearbox with five classifiers is carried out. And conclusions are drawn in Section 5.

II. METHODOLOGY

A. PCA

Principal Component Analysis (PCA) is the most widely used linear dimension reduction method [13]–[14]. The original matrix can be compressed by a set of orthogonal bases. Then, through the projection transformation, the original data matrix can be transformed into the set of bases, so that the covariance between sample points maintains zero, while the variance of the transformed result is as large as possible.

The matrix X is the original data, let $\text{Cov} = (1/n)XX^T$ denote the covariance matrix. According to equation (1), the elements on the diagonal represent the variance of the sample, and the remaining elements stand for the covariance between the sample points. Therefore, the variance and the covariance of the original data point are unified into the matrix Cov.

$$\text{Cov} = \frac{1}{n}XX^T = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n X_1^2 & \dots & \frac{1}{n}\sum_{i=1}^n X_1X_D \\ \dots & \dots & \dots \\ \frac{1}{n}\sum_{i=1}^n X_DX_1 & \dots & \frac{1}{n}\sum_{i=1}^n X_D^2 \end{bmatrix} \quad (1)$$

According to the optimization goal of PCA, the covariance between sample points needs to be maintained as zero. When the covariance matrix Cov is diagonalized, the goal is achieved. Since the covariance matrix is symmetrical, the eigenvectors corresponding to eigenvalues are orthogonal based on its nature. Let $E = (e_1, e_2, \dots, e_n)$ denotes the eigenvector of the covariance matrix Cov, the covariance matrix can be diagonalized as follows:

$$E^TCE = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix} \quad (2)$$

where C in the equation (2) stands for covariance matrix Cov, which is diagonalized through the matrix transformation with its eigenvector E . λ is the corresponding eigenvalues. Based on the above analysis, PCA attempts to find a linear mapping W that maximizes the following function:

$$C_{pca}(W) = \max_w \sum_{i=1}^n W^T(x_i - \bar{x})(x_i - \bar{x})^T W \quad (3)$$

Where, x_i represents the i -th sample of a set of data, and \bar{x} represents the mean of the sample. $X = x_i - \bar{x}$ represents the sample after averaging. The optimization goal of PCA is to find an optimal W to make the covariance matrix

$C = (x_i - \bar{x})(x_i - \bar{x})^T$ diagonal. Therefore, the feature vector $\mathbf{E} = (e_1, e_2, \dots, e_d)$ corresponding to the largest d eigenvalues is obtained as the projection matrix \mathbf{W} . The objective function can be represented as:

$$C_{PCA}(W) = \max_w \text{tr} (\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}) \quad (4)$$

The compressed sample space is presented as $\mathbf{Y} = \mathbf{W}^T \mathbf{x}$. The linear mapping \mathbf{W} is formed by d principal eigenvectors of the covariance matrix. PCA has been successfully applied for fault diagnosis [13], [14]. However, PCA is a linear technique that retains the global properties of the data, the local relationship between sample points can be neglected.

B. LLE

In contrast to PCA, locally linear embedding (LLE) is a local technique for dimensionality reduction, which attempts to preserve local properties of the data manifold. Based on the hypothesis, the data manifold is locally linear, LLE constructs a neighbor graph representation of the data points. In LLE, a high-dimensional sample points are constructed by the linear combination of their nearest neighbors [27]. In the low-dimensional representation of the data, LLE attempts to retain the linear construction weights as well as possible.

The main steps of LLE algorithm are as follows:

- 1) Find the neighbor points for each sample point. For given sample matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n] \in R^{D \times n}$, the pairwise similarity is measured using Euclidean distance, correlation, or geodetic distance equidistant calculation method etc. And the nearest K points are chosen as the neighbors. Assume that $\{\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{ik}\} \in R^{k \times n}$, ($i = 1, 2, \dots, n$) are the K neighbors of each point, X_{ik} denotes the K neighbor of X_i .
- 2) The local reconstruction weight matrix of the sample should be calculated, which represents the linear relationship between each sample point and its neighbors. According to the analysis, the reconstruction weight can be obtained by minimizing the cost function:

$$C_{lle}(W) = \min_v \sum_{i=1}^n \left\| \mathbf{X}_i - \sum_{j=1}^K W_{ij} \mathbf{Z}_{ij} \right\|^2 \quad (5)$$

$$\text{s.t. } \sum_{j=1}^K W_{ij} = 1, i = 1, 2, \dots, n \quad (6)$$

where \mathbf{Z}_{ij} represents the j -th neighbor of sample point i . W_{ij} represents the j -th linear reconstruction coefficient of sample point i . Set the non-adjacent weight coefficient of \mathbf{W} as zero, the high-dimension linear reconstruction coefficient is obtained.

- 1) [3)] let $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n] \in R^{d \times n}$ ($d < D$) denotes the low-dimensional projection of sample $\mathbf{X} \in R^{D \times n}$. Since the goal of LLE is to maintain the linear relationship of these high-dimensional data in low dimensional representation, the reconstruction weight

\mathbf{W} that reconstruct data point X_n from its neighbors can also reconstruct its low-dimensional representation data point Y_n . As a consequence, the low-dimensional data representation \mathbf{Y} amounts to minimizing the following cost function:

$$\begin{aligned} C_{LLE}(W) &= \min_Y \sum_{i=1}^n \left\| \mathbf{Y}_i - \sum_{j=1}^K W_{ij} \mathbf{Y}_j \right\|^2 = \min_Y \left\| \mathbf{Y} - \mathbf{Y} \mathbf{W}^T \right\|^2 \\ &= \min_Y \text{tr} \left\{ (\mathbf{Y} - \mathbf{Y} \mathbf{W}^T)(\mathbf{Y} - \mathbf{Y} \mathbf{W}^T)^T \right\} \\ &= \min_Y \text{tr} \left\{ \mathbf{Y}(\mathbf{I} - \mathbf{W}^T)(\mathbf{I} - \mathbf{W})\mathbf{Y}^T \right\} \end{aligned} \quad (7)$$

In this equation, ‘tr’ denotes trace function, \mathbf{I} is the identity matrix. Let $\mathbf{M} = (\mathbf{I} - \mathbf{W}^T)(\mathbf{I} - \mathbf{W})$. It is shown that the low-dimensional representations \mathbf{Y} that minimize this cost function can be found by computing the eigenvectors corresponding to the smallest d nonzero eigenvalues of matrix \mathbf{M} .

C. PCLLE

PCA is a global linear technique, it constructs the low-dimensional representation of the data that describes as much of the variance in the data as possible by finding a linear basis of low-dimensional data [25]. PCA enhances the global properties of the data, but neglects the local relationship between samples. Therefore, some nonlinear information may be lost through the process. In order to make up for the weakness of PCA, LLE is introduced to build a novel hybrid algorithm. LLE is a typical local nonlinear technique. In LLE, the high-dimensional data and its low-dimensional representation sharing the same reconstruction weight based on the assumption that the data manifold is locally linear. However, the global distribution properties of the sample point may be out of consideration in LLE. To combine the advantages of the two methods, a hybrid algorithm is proposed as an improved algorithm of PCA and LLE, which is called principal component locally linear embedding (PCLLE). Similar to PCA, PCLLE constructs a low-dimensional representation of the data by a new mapping that can meet the optimization goal of PCA and LLE at the same time. The details are as follows.

Let \mathbf{Y} denote the dimensionality reduction result of the original matrix \mathbf{X} . Assuming that there is a projection matrix \mathbf{P} that performs a projection transformation on \mathbf{X} , while maintaining the variance information and linear structure of \mathbf{X} at the same time. Thus, exists $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$. Therefore, the optimization goals of PCA and LLE can be represented as:

$$\begin{aligned} C_{PCA}(W) &= \max_w \text{tr} (\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}) \\ &= \max_P \text{tr} (\mathbf{P}^T \mathbf{X} \mathbf{X}^T \mathbf{P}) \end{aligned} \quad (8)$$

$$\begin{aligned} C_{LLE}(W) &= \min_Y \text{tr} (\mathbf{Y} \mathbf{M} \mathbf{Y}^T) \\ &= \min_P \text{tr} \left\{ (\mathbf{P}^T \mathbf{X}) \mathbf{M} (\mathbf{P}^T \mathbf{X})^T \right\} \\ &= \min_P \text{tr} \left\{ \mathbf{P}^T (\mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{P} \right\} \end{aligned} \quad (9)$$

And the optimization goals of PCA and LLE are combined to construct a new objective function in this paper. As a consequence, finding the low-dimensional representation \mathbf{Y} amounts to maximizing the following function:

$$C(W) = \max_P \operatorname{tr} \frac{\mathbf{P}^T \mathbf{X} \mathbf{X}^T \mathbf{P}}{\mathbf{P}^T (\mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{P}} \quad (10)$$

In the above function, a shared weight \mathbf{P} should be found to meet the optimization goals of PCA and LLE at the same time. In PCLLE, it is the direction of projection matrix \mathbf{P} rather than length that affects the result, in order to find a unique solution, the constraint $\mathbf{P}^T (\mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{P} = 1$ is added. Let $\mathbf{S} = \mathbf{X} \mathbf{X}^T$, $\mathbf{D} = \mathbf{X} \mathbf{M} \mathbf{X}^T$. Then the objective function can be rewritten as follows:

$$C(W) = \max_P \operatorname{tr} \frac{\mathbf{P}^T \mathbf{S} \mathbf{P}}{\mathbf{P}^T \mathbf{D} \mathbf{P}} \quad (11)$$

$$\text{s.t. } \mathbf{P}^T \mathbf{D} \mathbf{P} = 1 \quad (12)$$

Thus, the above solution problem can be transformed into: the maximum value of the solution function $C_S(\mathbf{P}) = \mathbf{P}^T \mathbf{S} \mathbf{P}$ under the condition $C_D(\mathbf{P}) = \mathbf{P}^T \mathbf{D} \mathbf{P} - 1 = 0$. A Lagrangian function is constructed to solve the problem.

$$L(\mathbf{P}) = \mathbf{P}^T \mathbf{S} \mathbf{P} - \lambda(\mathbf{P}^T \mathbf{D} \mathbf{P} - 1) \quad (13)$$

$$\frac{dL(\mathbf{P})}{d\mathbf{P}} = 2\mathbf{S}\mathbf{P}^* - 2\lambda\mathbf{D}\mathbf{P}^* = 0 \quad (14)$$

Solve the above formula, the result can be got as:

$$\mathbf{D}^{-1} \mathbf{S} \mathbf{P}^* = \lambda \mathbf{P}^* \quad (15)$$

In the above equation, \mathbf{P}^* is the extreme solution of $L(\mathbf{P})$. When \mathbf{D}^{-1} is non-singular, the solution of the objective function is equivalent to calculate the eigenvalues and eigenvectors of $\mathbf{D}^{-1} \mathbf{S}$. The eigenvector corresponding to the largest d eigenvalues is the projected projection matrix, which is determined by following equation.

$$\sum_{i=1}^d \eta_i \left/ \sum_{i=1}^D \eta_i \right. \geq \delta \quad (16)$$

Where η stands for the contribution rate of the eigenvalue. When the cumulative contribution rate of the current d eigenvalues reaches the specified value δ , the previous d eigenvalues are taken to form a projection matrix.

III. VALIDATION OF THE PROPOSED METHOD

In this paper, a hybrid feature reduction algorithm is proposed for fault diagnosis. Based on the above preparation, the general steps of PCLLE algorithm can be summarized as follows:

Step 1: For the given matrix $\mathbf{X} = [X_1, X_2, \dots, X_n] \in R^{D \times n}$, the zero-mean matrix $\bar{\mathbf{X}}$ of the original data should be calculated.

Step 2: To describe the variance information in the data, the covariance matrix of $\bar{\mathbf{X}}$ is obtained in consideration of PCA, denoted as matrix \mathbf{S} . The calculation process is as shown in Equation (1).

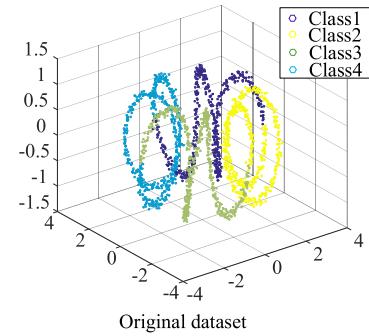


FIGURE 1. Visualization of the helix data generated for validation.

Step 3: Find the neighbor points for each sample point, the pairwise similarity is measured by Euclidean distance in experiment. And the most similar K points are selected as the neighbors.

Step 4: Based on equation (5-7), the local reconstruction weight matrix of the high-dimensional data is found in consideration of LLE, which is denoted as \mathbf{W} . As a consequence, the matrix $\mathbf{M} = (I - \mathbf{W}^T)(I - \mathbf{W})$ can be obtained. Based on equation (10), $\mathbf{D} = \mathbf{X} \mathbf{M} \mathbf{X}^T$.

Step 5: Since the matrix \mathbf{S} and \mathbf{D} are obtained, the objective function of PCLLE can be written as equation (11). The projection matrix \mathbf{P} can be found according to equation (12-15).

Step 6: By using the first several eigenvalues, sorted in descending order, of the eigenvalues, the number of components in original data can be reduced.

In order to verify the effectiveness of the proposed method, a set of helix data is introduced to evaluate the dimensionality reduction performance of PCA, LLE and PCLLE. The helix data is generated using MATLAB, which contains a total of 2000 sample points with 3 dimensions. Each of the 3 dimension data can be expressed as:

$$\mathbf{X}_{\text{helix}} = [X_1 \ X_2 \ X_3] + \vartheta_{\text{noise}} \quad (17)$$

$$X_1 = (2 + \cos 8t) * \cos t \quad (18)$$

$$X_2 = (2 + \cos 8t) * \sin t \quad (19)$$

$$X_3 = \sin 8t \quad (20)$$

$$t = 2\pi \sum_{k=1}^n k/n \quad (21)$$

where X_1, X_2, X_3 are the three column vectors that denote the 3 dimensions of the helix data. The notation ϑ_{noise} denotes noise term, n is the number of total sample point.. For this paper, a 2000-by-3 matrix of normally distributed random numbers are used as noise. The detail of the data point is partly shown in TABLE 1.

FIGURE 1 is the visualization of the original data. The original data is separated into 4 classes, which are presented by different colors. In the next step, the original data is compressed in two dimensions by three feature reduction methods, which are PCA, LLE, and PCLLE respectively. The low-dimensional embedded results are shown in FIGURE 2.

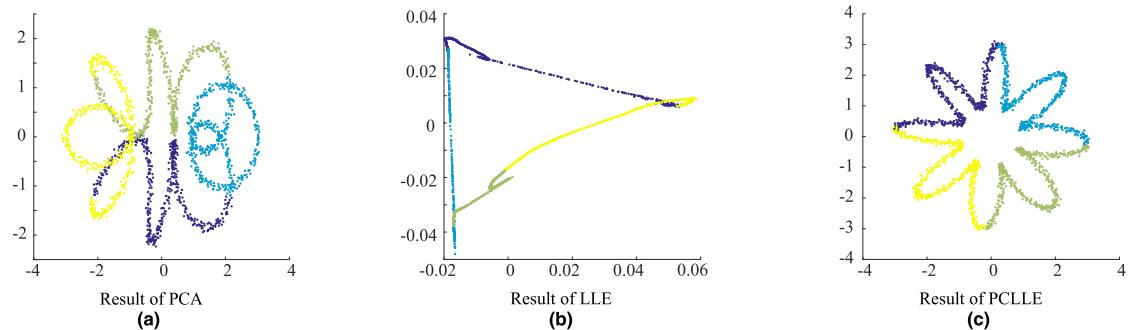


FIGURE 2. Data compressed by (a) PCA, (b) LLE, and (c) PCLLE.

TABLE 1. The original data used for validation.

Category Label	d_1	d_2	d_3
Class 1	2.9310	-0.0105	-0.0026
	2.9298	-0.0213	0.0748
	2.9731	0.0345	0.1280
	2.9972	0.0410	0.0584

Class 2	-0.0202	3.0313	0.0487
	-0.0565	2.9763	0.0172
	-0.0047	2.9512	0.1550
	-0.1730	3.0328	0.1872

Class 3	-2.9896	0.0086	0.0668
	-3.0810	-0.0900	0.1629
	-3.0235	-0.0600	0.0269
	-3.0353	-0.0633	0.0269

Class 4	-0.0468	-2.9929	-0.0137
	0.0356	-2.9172	0.0639
	-0.0843	-2.9005	0.1365
	0.0973	-2.9956	0.1028

Figure 2 (a)-(c) show the results of feature space compressed by PCA, LLE, and PCLLE respectively. As it can be seen from FIGURE 2(a), class 2 of the data space compressed by PCA intersects with class 1 and class 3. In FIGURE 2(b), the data space compressed by LLE overlaps in class 1 and class 2. FIGURE 2(c) presents the data compressed by PCLLE, the results show that features compressed by PCLLE can be perfectly separated in the two-dimension space. The validation results indicate the excellent performance of PCLLE method, which can be applied on the fault diagnosis of gearbox.

IV. FAULT DIAGNOSIS WITH PCLLE

In this paper, PCLLE model is proposed to compress the original feature matrix of the vibration signal, and several classifiers are introduced to validate the performance of

the compressed feature spaces. According to FIGURE 3, the diagnosis process is summarized as below.

- 1) Collect vibration signals under different fault conditions from sensors installed on gearbox;
- 2) Calculate the original features of the vibration signal from the time domain and the time-frequency domain;
- 3) The low-dimensional representation of the high-dimensional feature matrix is calculated by the proposed method. According to the above analysis, an optimal projection matrix P should be found based on equation (11-15), which is used to perform the projective transformation on the original feature matrix. Its low-dimensional representation can be obtained;
- 4) Divide the compressed feature into training part and testing part. Calculate the fault diagnosis accuracy using testing data.

A. DATA COLLECTION

For the experiments in this study, the vibration signal was collected in a spur gearbox. The experimental system is shown in FIGURE 4, which is driven by a motor. As shown in Figure 4(a), a coupling (4 in FIGURE 4(a)) connects the input shaft of the motor (6 in FIGURE 4(a)) and the gearbox (2 in FIGURE 4(a)), and the speed is input to the gearbox via the coupling. During operation of the gearbox drive system, faults and defects in the gearbox will cause changes in the physical effects of the structure, which are reflected in the vibration signals. The output shaft of the gearbox is connected with an electromagnetic torque break (5 in FIGURE 4(a)) through a belt transmission (1 in FIGURE 4(a)). Figure 4(b) shows the configuration of the gearbox, two spur gears are mounted on the input and output shafts of the gearbox respectively (number of teeth: G1 = 53, and G2 = 80).

In order to evaluate the proposed method, three faulty conditions and a healthy condition are applied to the experiment. As shown in TABLE 2, four fault conditions are formed. The load of the torque break can be manually adjusted by a controller (9 in FIGURE 4(a)). In this experiment, 3 different load conditions are chosen for each pattern (0V, 10V and 30V respectively). Signals are collected by an accelerometer (3 in FIGURE 4(a)) mounted on top of the gearbox. In the data acquisition process, the sampling frequency is 50k HZ and

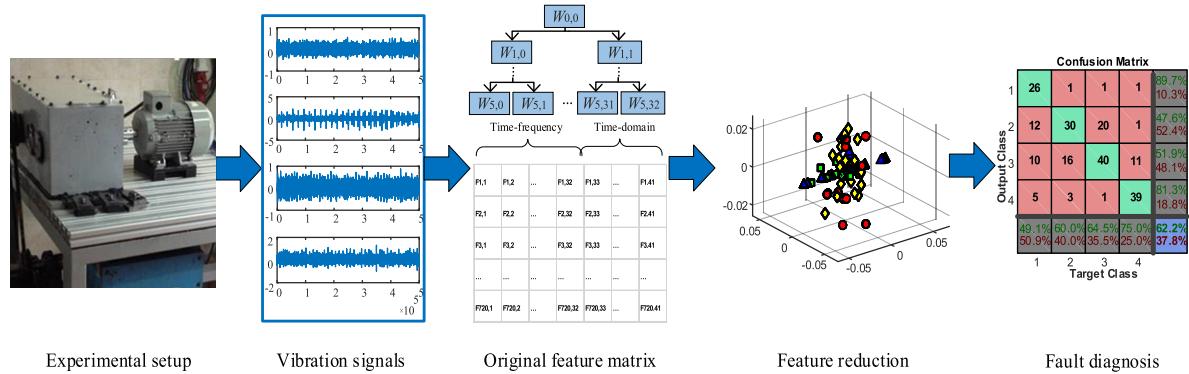


FIGURE 3. General procedures of the fault diagnosis experiment.

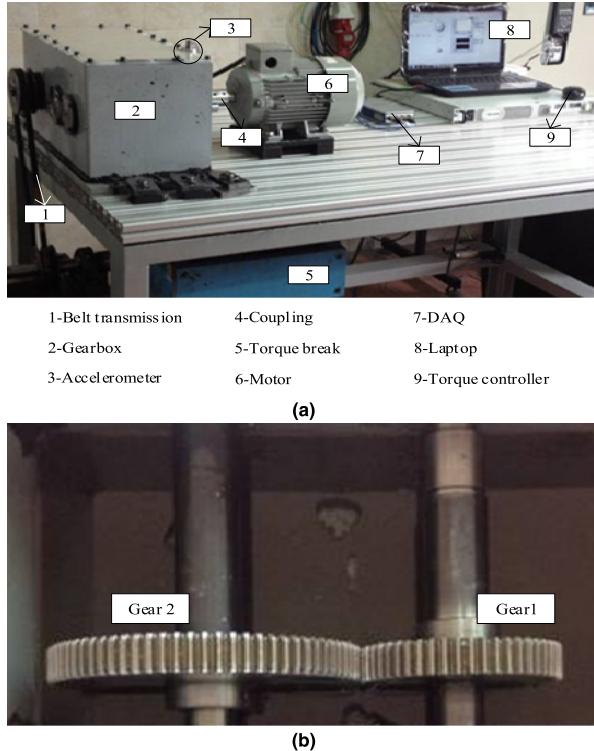


FIGURE 4. Experimental configurations. (a) The experimental platform. (b) The configuration of the gearbox.

the time length of each sample is 5s. The data acquisition box (7 in FIGURE 4(a)) transmits the collected vibration signals to the laptop (8 in FIGURE 4(a)).

The whole data set consists of 720 sample points of four different fault conditions, and each fault condition includes 180 sample points. In the experimental process, the order of the sample is disrupted, and the disordered sample is split into two sets: 504 samples (70% of the data set) for training and 216 samples for testing (30% of the data set).

B. FEATURE PROCESS

In this study, 41 features are selected to describe the vibration signals of gearbox. As listed in TABLE 3, there are 9 time-domain statistical features (peak-to-peak value,

TABLE 2. Condition patterns of the experiment.

Category Label	Gear1	Gear2	Samples Number
Class 1	Health tooth	Health tooth	180
Class 2	Chipped tooth 25%	Health tooth	180
Class 3	Health tooth	Chipped tooth 25%	180
Class 4	Chipped tooth 25%	Chipped tooth 25%	180

standard deviation, kurtosis, root mean square, waveform factor, peak factor, kurtosis factor, impulse factor, and margin factor) and the 32 time-frequency features [28-31] included in this work. The time-frequency features are extracted by Wavelet Packet Transform (WPT). In the experiment, WPT (shannon entropy and daubechies 10) is applied to perform 5-layer decomposition of the vibration signal, and results in 32 decomposed frequency bands [32-33]. Through calculation, the energy sum for each frequency band can be obtained, which represents the 32 features of vibration signal in time-frequency domain (w_1-w_{32}). Since there are 720 samples, a total of 41 characteristic signals in the time domain and time-frequency domain are extracted, then a 720×41 original feature matrix is obtained.

The PCLLE algorithm is applied to reduce the dimensionality of the original feature matrix. Initially, average the original data, and the covariance matrix of the original feature matrix is calculated according to equation (1). Then, find the K nearest neighbors of each sample point. The Euclidean distance is proposed to measure the similarity between sample points. Since the main function of LLE is to preserve the local linear relationship between samples. When the value of K is too large, it is hard to reflect the local relationship of the sample. Therefore, the nearest neighbor point K is valued between 1-10. In the next step, calculate the local reconstruction weight matrix of the data according to equation (5), it is used to present the linear relationship between each sample point and its neighbors. According to equation (11-15), an optimal projection matrix P can be obtained after the covariance matrix and the local reconstruction weight matrix are obtained.

TABLE 3. The 41 features generated from vibration signals.

Time-domain features		
1	Peak-to-peak value	$P_k = \max(x) - \min(x)$
2	Standard deviation	$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n x_i - \bar{x} ^2}$
3	Kurtosis	$Ku = E(x - \bar{x})^4 / \sigma^4$
4	Root mean square	$X_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
5	Waveform factor	$S = X_{rms} / \left(\sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} x_i } \right)$
6	Peak factor	$C = X_{peak} / X_{rms}$
7	Kurtosis factor	$K_r = \sum_{i=1}^n x_i^4 / n X_{rms}^2$
8	Impulse factor	$I = X_{peak} / \left(\sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} x_i } \right)$
9	Margin factor	$M = X_{peak} / \left(\sqrt{\frac{1}{n} \sum_{i=1}^n x_i } \right)^2$
Time-frequency features		
Wavelet packet energy: w1, w2, ..., w32		

After the projection transformation of the original feature matrix, a compressed feature matrix can be obtained. In this paper, the spatial dimensionality of the original feature is reduced to 6.

Fisher discriminant analysis is a common method of feature extraction, the goal is to integrate samples of the same category and separate samples of different categories as much as possible [24], [34]–[35]. In this paper, the fisher discriminant value is employed to measure the within-class scatter and between-class scatter of the compressed feature, and the feature with the smallest within-class scatter and largest between-class scatter is selected. Assuming that the sample contains c categories, the within-class scatter S_w and the between-class scatter S_b of the sample can be expressed as:

$$S_w = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{N_i} (y_j(i) - y(i)) (y_j(i) - y(i))^T \quad (22)$$

$$S_b = \frac{1}{N} \sum_{i=1}^c N_i (y(i) - \bar{y}) (y(i) - \bar{y})^T \quad (23)$$

Where N represents the total number of samples. $y(i)$ represents the mean of the i -th sample, $y_j(i)$ represents the j th sample in the i -th class, and \bar{y} stands for the average of all samples. The fisher discriminant value is defined as follows:

$$F = S_w / S_b \quad (24)$$

Based on the above formula, the fisher discriminant values of the features compressed by PCA, LLE, and PCLLE algorithm are calculated, the results are shown in TABLE 4. When

TABLE 4. Fisher discrimination value of compressed features.

Compressed features	K	S_w	S_b	F
PCA	—	3.9841e+09	4.0483e+08	0.1016
	1	0.2679	0.987	0.3684
	2	0.2538	0.1303	0.5133
	3	1.9330	1.6280	0.8422
	4	2.7003	0.2352	0.0871
	5	39.5655	4.2468	0.1073
	6	1.6471e+03	157.5870	0.0957
	7	6.6845e+04	6.4006e+03	0.0958
	8	5.2422e+05	5.3716e+04	0.1025
	9	1.0406e+05	1.0911e+05	0.1049
LLE	10	4.0234e+07	3.7737e+06	0.0938
	1	0.0020	1.0900e-04	0.0532
	2	0.0020	2.8566e-04	0.1439
	3	0.0020	1.5899e-04	0.0784
	4	0.0021	7.7260e-05	0.0375
	5	0.0019	3.3015e-04	0.1725
	6	0.0019	3.4992e-04	0.1813
	7	0.0020	2.9421e-04	0.1490
	8	0.0020	2.2550e-04	0.1119
	9	0.0020	1.9110e-04	0.0954
PCLLE	10	0.0020	4.1089e-04	0.2094

the K value is taken as 4, the fisher discriminant value of the feature compressed by LLE and PCLLE is the smallest.

According to the calculation results, fisher value of the feature compressed by PCLLE is 0.0375. Comparing with 0.0871 of LLE and 0.1016 of PCA, the fisher value obtained by PCLLE is much smaller. It shows that features compressed by PCLLE is more separable than PCA and LLE in theory. On the other hand, the within-class scatter S_w of PCLLE is around 0.002, which indicates that the distance between the compressed feature matrix of the same category is small. It shows that PCLLE has netter stability in dimension reduction process.

FIGURE 5 shows the low-dimensional embedded results of the mentioned algorithms. FIGURE 5(a)-(b) show the six features compressed by PCA. FIGURE 5(c)-(d) show the six features compressed by LLE. FIGURE 5(e)-(f) shows the six features compressed by PCLLE.

C. MULTI-CLASSIFICATION WITH DIFFERENT METHODS

In order to evaluate the performance of the proposed method, five classifiers are introduced to compare the classification accuracy in this case. They are random forest (RF) [36]–[37], back propagation neural network (BPNN) [38]–[39], support vector machine (SVM) [40], K-nearest neighbor (KNN) [41], [42], and Softmax classifier,

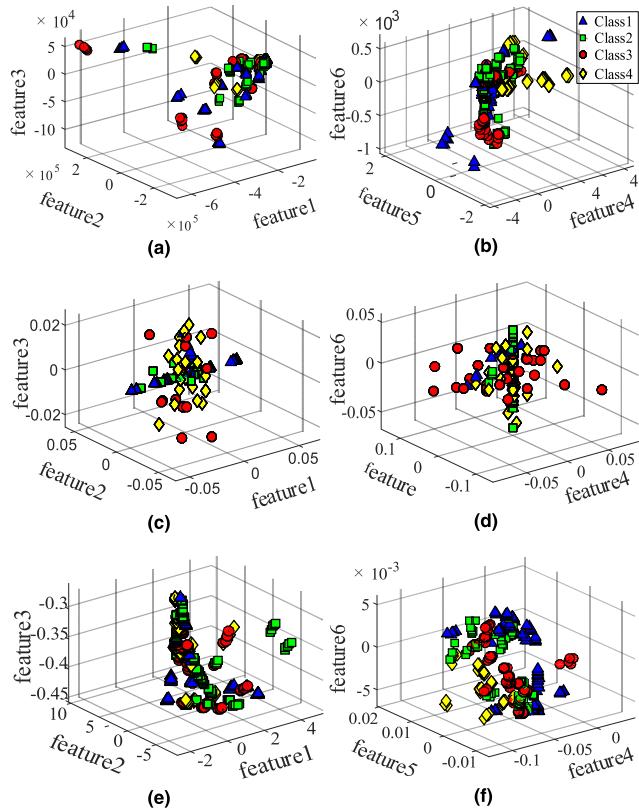


FIGURE 5. Visualization of the compressed feature space. (a-b) Feature space compressed by PCA. (c-d) Feature space compressed by LLE. (e-f) Feature space compressed by PCLLE.

respectively. The diagnosis process is implemented by using MATLAB. As mentioned before, the 70% of the sample points is chosen to train the classifiers, and the 30% of the sample points is selected for testing. For each classifier, the experiment is repeated 10 times, and the training and testing data are selected randomly from the compressed feature matrix in each trial. The average classification accuracy of the 10 trials for each classifier is the measuring term for this comparison. The comparison results are shown in FIGURE 6 and TABLE 5.

A RF classifier with 100 decision trees are embedded in this paper. The experiment is repeated 10 times with same parameters, and the average classification accuracy is shown in FIGURE 6 and TABLE 5. According to the results, the PCLLE method performs best among the three methods with the average classification accuracy of 99.31%. The result is 0.75% higher than PCA method, and 3.15% higher than LLE method.

The BPNN model is developed with one hidden layer, which contains 100 neurons, and the number of the maximum pre-training epochs of BPNN are 1000. The classification accuracy of PCA, LLE and PCLLE with BPNN are 91.06%, 87.88%, and 97.70%, respectively. The result of PCLLE is 6.64% higher than PCA, and 9.82% than LLE.

For SVM model, the selection of kernal function is critical. In the experiment, the Gaussian Radial Basis Function (RBF)

TABLE 5. Average accuracy of the various classification with the compressed three feature spaces for ten trials.

Feature space	Value (%)				
	RF	BPNN	SVM	KNN	Softmax
PCA	98.56	91.06	96.36	99.35	55.53
LLE	96.16	87.88	97.02	98.85	42.12
PCLLE	99.31	97.70	98.93	98.62	54.47

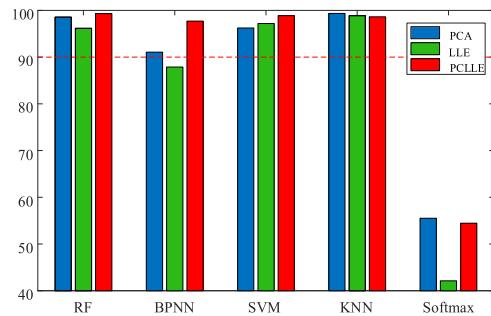


FIGURE 6. Comparison results for different compressed feature with five classifiers.

is introduced with kernal scale parameter of 0.1. Based on the classification result, the accuracy of PCA, LLE and PCLLE with SVM are 96.36%, 97.02%, and 98.93%, respectively. PCLLE presents the highest accuracy with SVM, which is 2.57% higher than PCA, and 1.91% than LLE.

In this paper, a 1-nearest neighbor (1NN) model is introduced for experiment. According to the results, the three feature reduction methods perform well on 1NN classifier, and the highest average accuracy of 99.35% is PCA, followed by 98.85% of LLE, and 98.62% of PCLLE. Based on the result, PCA is 0.5% higher than LLE and 0.73% higher than PCLLE.

For softmax classifier, the Mean Squared Error (MSE) is chosen for cost function, and the ‘trainscg’ function is applied for training function. The maximum epoch is selected as 1000. The three feature reduction methods show low classification accuracy on softmax classifier. PCA obtained the average accuracy of 55.53%, followed by 54.47% of PCLLE and 42.12% of LLE.

Based on the above analysis and FIGURE 6, the proposed PCLLE method presents relatively high classification accuracy on the five classifiers. When using RF, BPNN and SVM classifiers, the classification accuracy of PCLLE is significantly higher than that of PCA and LLE. At the same time, the average classification accuracy of PCLLE on RF, BPNN, SVM and KNN classifiers reached 98.64%, is about 2.31% higher than PCA’s 96.33%, and 3.69% higher than LLE’s 94.95%. Since the selection of classifiers is critical for fault detection, it is significant to select a suitable classifier. The classification results of the five classifiers indicate that the proposed method PCLLE is more adaptable to different

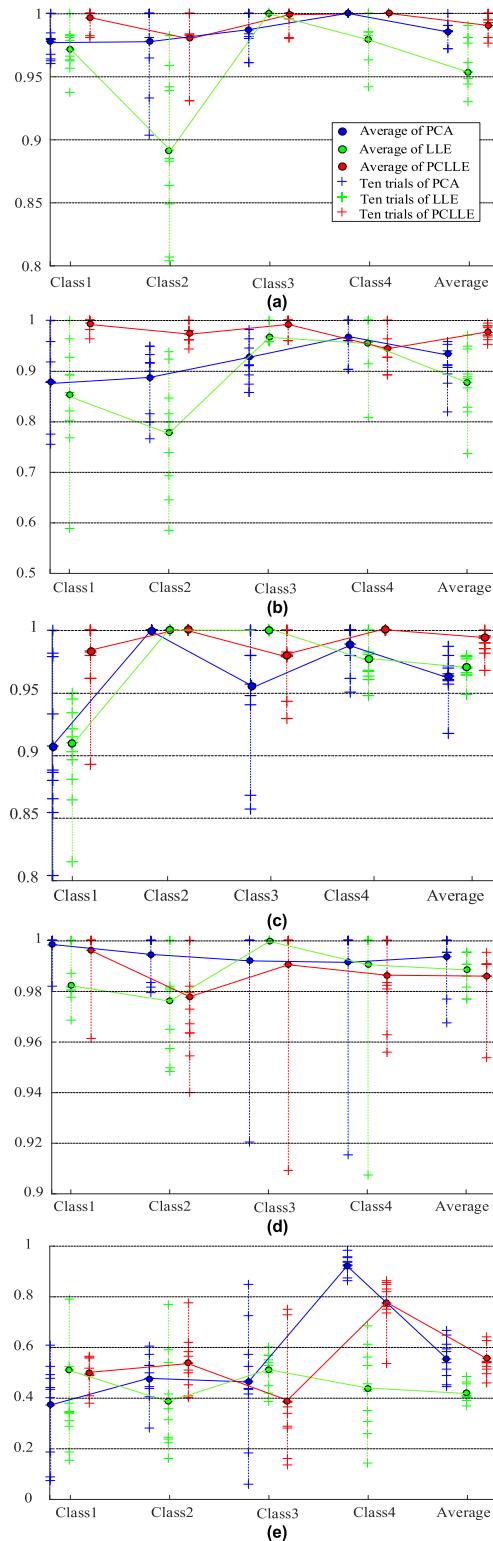


FIGURE 7. Details of the ten trials using different algorithms.
(a) Diagnosis results of RF. **(b)** Diagnosis results of BPNN. **(c)** Diagnosis results of SVM. **(d)** Diagnosis results of KNN. **(e)** Diagnosis results of softmax.

classifier, which means the PCLLE method could present high classification accuracy in different classifier, and reduce the accuracy bias that caused by different classifier.

To further evaluate the performance of the proposed three methods, FIGURE 7 presents the detail of the ten trials. The floating scatter points in the graph represent the results of the 10 trials, where blue represents the experimental results of PCA, green represents the experimental results of LLE, and red represents the experimental results of PCLLE. The vertical axis represents the classification accuracy, and the horizontal axis represents the classification result of the four faults conditions and the average accuracy. The FIGURE 7(a-e) shows the detailed diagnosis results generated by RF, BPNN, SVM, KNN, and Softmax model respectively.

In RF classifier, the proposed PCLLE method not only achieves the highest average classification accuracy, but also maintains the highest average classification accuracy for the four fault conditions. As can be seen from the figure, the fluctuation of the classification accuracy generated by PCA and LLE is very intense when classifying class2. Meanwhile, the fluctuation of PCLLE is small.

In BPNN model, the results of PCA and LLE methods fluctuate intensely when identifying class 1 and 2. On the contrary, the classification accuracy of the four fault conditions of PCLLE is very small, and the average accuracy of the 10 tests fluctuated within a very small range.

When using the SVM model, Although the accuracy of the three methods for class1 classification fluctuated greatly, the results of PCA and LLE are pretty scattered, while in the 10 trials of PCLLE, most of the results were concentrated in high accuracy except for several point. Although LLE performs well when classifying class2 and class3, its classification accuracy for class1 is low, indicating that LLE is unstable.

Based on the classification results of KNN classifier, we can see the three methods perform high accuracy with only weak differences.

In Softmax classifier, the classification results of the three methods are low. Based on the result, PCA can achieve higher average accuracy, followed by PCLLE. However, the diagnostic results of PCA are more volatile, that is, PCA is not very stable. Although the average accuracy of the LLE fluctuates within a small range, the accuracy of the LLE in the diagnosis of a single category is large and the accuracy that can be achieved is low.

Based on the above analysis, when combined with different classifiers, PCLLE maintain pretty high and steady classification accuracy, shows the outstanding adaptability of the proposed method. On the other hand, when PCLLE is combined with different classifiers, the classification accuracy fluctuation of PCLLE method in each trial are relatively small, that indicates the excellent robustness of the method.

V. CONCLUSION

The compression of high-dimensional feature space is a primary process for data preprocessing of fault detection. Based on the optimization goals of PCA and LLE method, an hybrid feature reduction method PCLLE is proposed, the method attempts to find a projection matrix that meets the optimiza-

tion goals of PCA and LLE at the same time. To evaluate the reliability of the proposed method, a set of helix data is generated firstly for validation, the result shows its excellent performance in dimensionality reduction. Then, the proposed method is used for gearbox fault diagnosis. In the experiments, to facilitate parameter selection, Fisher discriminant value is chosen to analyze the separability of the compressed feature. According to the analysis of Fisher value, the feature space compressed by PCLLE shows better separability and stability. Finally, five classifiers are introduced to this paper, the proposed dimensionality reduction method is experimentally verified and compared, and the conclusions are drawn as follows:

- 1) Based on the analysis of the average classification accuracy generated by five classifiers, the feature compressed by the proposed PCLLE method can achieve high classification accuracy in experiments. Meanwhile, it maintains high classification accuracy on different classifier compared with PCA and LLE method, which shows the excellent adaptability of the proposed method.
- 2) Based on the further analysis of the ten trials done in the experiment, the fluctuation of classification accuracy in each trial are relatively small on the introduced four classifiers (RF, BPNN, SVM, KNN), which indicates the good robustness of the method.

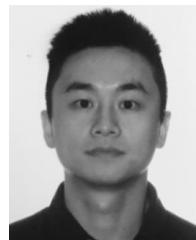
REFERENCES

- [1] C. Li, D. Cabrera, J. V. de Oliveira, R.-V. Sánchez, M. Cerrada, and G. Zurita, "Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram," *Mech. Syst. Signal Process.*, vol. 76–77, pp. 157–173, Aug. 2016.
- [2] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1419–1430, Mar. 2010.
- [3] C. Li and M. Liang, "Time–frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform," *Mech. Syst. Signal Process.*, vol. 26, no. 1, pp. 205–217, Jan. 2015.
- [4] B. Chen, Z. Zhang, C. Sun, B. Li, Y. Zi, and Z. He, "Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors," *Mech. Syst. Signal Process.*, vol. 33, pp. 275–298, Nov. 2012.
- [5] L. Song, H. Wang, and P. Chen, "Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory," *IEEE Trans. Fuzzy Syst.*, to be published, doi: 10.1109/TFUZZ.2018.2833820.
- [6] X. H. Li, M. P. Jia, and F. Y. Xu, "A new unbiased frequency response function estimator in experimental modal analysis," *J. Vib. Meas. Diagnostics*, vol. 23, no. 3, pp. 168–170, 2003.
- [7] Q. He, "Time–frequency manifold for nonlinear feature extraction in machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 200–218, Feb. 2013.
- [8] C. Sankavaram, A. Kodali, K. R. Pattipati, and S. Singh, "Incremental classifiers for data-driven fault diagnosis applied to automotive systems," *IEEE Access*, vol. 3, pp. 407–419, 2015.
- [9] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Mar. 2018.
- [10] C. Li et al., "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3565–3580, Aug. 2018.
- [11] Y. Li, Y. Wei, K. Feng, X. Wang, and Z. Liu, "Fault diagnosis of rolling bearing under speed fluctuation condition based on Vold–Kalman filter and RCMFE," *IEEE Access*, vol. 6, pp. 37349–37360, 2018.
- [12] C. Li, V. Sanchez, G. Zurita, M. C. Lozada, and D. Cabrera, "Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement," *ISA Trans.*, vol. 60, pp. 274–284, Jan. 2016.
- [13] W. Sun, J. Chen, and J. Li, "Decision tree and PCA-based fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 21, no. 3, pp. 1300–1317, Apr. 2007.
- [14] J. Fortuna and D. Capson, "Improved support vector classification using PCA and ICA feature space modification," *Pattern Recognit.*, vol. 37, no. 6, pp. 1117–1129, Jun. 2004.
- [15] Y. Zhang and C. Ma, "Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS," *Chem. Eng. Sci.*, vol. 66, no. 1, pp. 64–72, Jan. 2011.
- [16] J. Wu, J. Wang, and L. Liu, "Feature extraction via KPCA for classification of gait patterns," *Hum. Movement Sci.*, vol. 26, no. 3, pp. 393–411, May 2007.
- [17] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [18] Y. Zhang, B. Li, Z. Wang, W. Wang, and L. Wang, "Fault diagnosis of rotating machine by isometric feature mapping," *J. Mech. Sci. Technol.*, vol. 27, no. 11, pp. 3215–3221, Nov. 2013.
- [19] W. Cui, C. Wu, W. Meng, B. Li, Y. Zhang, and L. Xie, "Dynamic multi-dimensional scaling algorithm for 3-D mobile localization," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 12, pp. 2853–2865, Dec. 2016.
- [20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [21] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
- [22] J.-B. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 7440–7450, Jun. 2011.
- [23] A. Samant and H. Adeli, "Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis," *Comput. Aided Civil Infrastruct. Eng.*, vol. 15, no. 4, pp. 241–250, Jun. 2003.
- [24] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 230–244, Feb. 2005.
- [25] M. H. C. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 377–391, Mar. 2006.
- [26] Y. Liu, Y. Zhang, Z. Yu, and M. Zeng, "Incremental supervised locally linear embedding for machinery fault diagnosis," *Eng. Appl. Artif. Intell.*, vol. 50, pp. 60–70, Apr. 2016.
- [27] L. van der Maaten, E. Postma, and J. van den Herik, "Dimensionality reduction: A comparative review," *Rev. Lat. Am. Lit. Arts*, vol. 10, pp. 1–35, Feb. 2007.
- [28] W. Li, Z. Zhu, F. Jiang, G. Zhou, and G. Chen, "Fault diagnosis of rotating machinery with a novel statistical feature extraction and evaluation method," *Mech. Syst. Signal Process.*, vol. 50–51, pp. 414–426, Jan. 2015.
- [29] T. Karacay and N. Akturk, "Experimental diagnostics of ball bearings using statistical and spectral methods," *Tribol. Int.*, vol. 42, no. 6, pp. 836–843, Jun. 2009.
- [30] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, May 2013.
- [31] T. I. Patargas, C. T. Yiakopoulos, and I. A. Antoniadis, "Performance assessment of a morphological index in fault prediction and trending of defective rolling element bearings," *Nondestruct. Test. Eval.*, vol. 21, no. 1, pp. 39–60, Aug. 2006.
- [32] X. Fan and M. J. Zuo, "Gearbox fault detection using Hilbert and wavelet packet transform," *Mech. Syst. Signal Process.*, vol. 20, no. 4, pp. 966–982, 2006.
- [33] Z. Zhang, Y. Wang, and K. Wang, "Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network," *J. Intell. Manuf.*, vol. 24, no. 6, pp. 1213–1227, Dec. 2013.
- [34] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1389–1401, Jul. 2004.

- [35] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.
- [36] C. Li *et al.*, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mech. Syst. Signal Process.*, vols. 76–77, pp. 283–293, Aug. 2016.
- [37] M. Cerrada *et al.*, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 87–103, Mar. 2016.
- [38] P.-Y. Chou, J.-T. Tsai, and J.-H. Chou, "Modeling and optimizing tensile strength and yield point on a steel bar using an artificial neural network with taguchi particle swarm optimizer," *IEEE Access*, vol. 4, pp. 585–593, 2016.
- [39] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4034–4042, Sep. 2013.
- [40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 2015.
- [41] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.
- [42] R. V. Sánchez, P. Lucero, R. E. Vásquez, M. Cerrada, J.-C. Macancela, and Y. D. Cabrera, "Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3463–3473, Jan. 2018.



YU WANG was born in Sichuan, China, in 1994. She received the B.S. degree in management from Chongqing Technology and Business University, Chongqing, China, in 2017. Since 2017, she is a graduate with the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, majoring in management science and engineering. Her research interests are in data analysis and machinery health maintenance.



SHUAI YANG was born in Chongqing, China, in 1986. He received the B.S. and M.S. degrees in mechanical engineering from Chongqing University, China, in 2012, and the Ph.D. degree in mechanical engineering from the University of Ottawa, Ottawa, ON, Canada, in 2017.

From 2017 to 2018, he was a Research Assistant with the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing, China.

Since 2018, he has been an Associate Professor and also a Research Assistant with the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University. He has authored a book, over 15 articles, and over five inventions. His research interests include vibration control, fault diagnosis, and health management of manufacturing system.

Dr. Yang was a recipient of the Excellent Reward of Graduate Studies from the Canadian Society of Engineering in 2016.



RENÉ VINICIO SÁNCHEZ received the B.S. degree in mechanical engineering from the Universidad Politécnica Salesiana (UPS), Ecuador, in 2004, and the Ph.D. degree in industrial technologies research from UNED, Spain, in 2017. He is currently a Professor with the Department of Mechanical Engineering, UPS. His research interests are in machinery health maintenance, pneumatic and hydraulic systems, artificial intelligence, and engineering education.

• • •