

## Research Article

# Bearing Fault Diagnosis Based on Collaborative Representation Using Projection Dictionary Pair

Dan Ma , Yixiang Lu , Yushun Zhang, Hua Bao, and Xueming Peng

Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education,  
School of Electrical Engineering and Automation, Anhui University, Hefei, China

Correspondence should be addressed to Yixiang Lu; [lyxahu@ahu.edu.cn](mailto:lyxahu@ahu.edu.cn)

Received 21 July 2019; Revised 9 October 2019; Accepted 30 October 2019; Published 22 November 2019

Academic Editor: Silvio L. T. de Souza

Copyright © 2019 Dan Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In state analysis of rolling bearings using collaborative representation theory, how to construct an excellent redundant dictionary to collaboratively represent the acquired normal or abnormal data has been being a significant issue. Thus, a new method for fault detection and classification of rolling bearings is proposed in this paper. The proposed algorithm mainly consists of three components. First, a wavelet transform is employed to extract features, which takes advantage of the observation that vibration signals under different conditions have similar frequency spectra. This similarity ensures that we can collaboratively represent any test sample by using training samples. Second, under the similarity assumption, a dictionary pair learning strategy is employed to build an overcomplete dictionary pair, which is used to realize an optimal representation of the vibration signal. Meanwhile, the sparse constraint is also taken into account during dictionary training to enhance the robustness of the classification. Finally, the learned dictionary combined with collaborative representation is used to intelligently perform pattern classification of rolling bearings. The effectiveness and superiority of the method are verified by applying the proposed algorithm on the simulated and real vibration signals. The results show that, for different fault categories generated from different fault size and motor loads, our method can rapidly and accurately identify the fault category to which the input sample belongs.

## 1. Introduction

Among the majority of rotating machines, fairly important and frequently encountered components are rolling bearings, and the operating efficiency of an entire machine or an entire system is directly affected by their operating state. Therefore, the state detection and classification of rolling bearings have always been and will still be a research hotspot because of their low-cost maintenance and the reduction of unpredictable influences in some cases [1]. In the past few years, the state diagnosis of bearings has been intensively studied with respect to vibration and acoustic measurement analysis and the research has achieved many satisfactory results [2–4]. However, due to environmental noises and the shaft rotational speed, the acquired vibration signals of rolling bearings are always more complicated. Thus, how to determine the state of bearings via obtained vibration signals is still an important issue worth studying.

In the early stage of fault diagnosis, some statistics (e.g., the root mean square and kurtosis) were extensively used to extract the fault characteristics of the original signals in time domain. Although these statistical parameters can be computed using fairly simple methods [5, 6], they can not completely represent the fault information features due to the existence of unknown periods or frequencies when analyzing vibration signals with complex faults, which may lead to low classification precision. To obtain more detailed fault information, many new methods were developed in recent years. As most popular methods in fault detection, feature-based algorithms assume that each fault of a rolling bearing produces a specific vibration that can be measured and has a unique characteristic. Currently, with respect to state detection, a variety of feature-based detection techniques (e.g., frequency-domain methods and time-frequency methods) [7–10] have been studied and successfully applied in practice. In general, feature-based fault diagnosis

methods mainly involve feature extraction and pattern classification. First, a distinct feature can improve accuracy rate and reduce the complexity of the classifier. In [11], Wang et al. extracted a distinct feature (i.e., the singular values of a feature matrix) from the considered data using singular value decomposition (SVD) and empirical mode decomposition and employed a very simple classifier called the Mahalanobis distance to accomplish fault clustering. In [12], a characteristic information vector is extracted using the wavelet package transform, and then the modified features (after dimensionality reduction) are fed into twin support vector machine (SVM) for fault detection. Second, a good classifier can provide excellent predictions for both training samples and testing (or query) samples and have good compatibility with the extracted features. For example, an SVM can produce high accuracy classification and diagnosis due to its excellent generalization performance [13–15]. In addition, Wang [16] constructed redundant statistical features using a binary wavelet packet transform with a wavelet called the Daubechies 44 and then employed the simple Kankar nearest neighbors to identify the features, which achieved high prediction accuracies. However, in spite of the theoretical correctness of these methods, the selection of features needs a wealth of prior knowledge and the features vary greatly in practice, which complicates the adaptivity. Meanwhile, for some classifiers, the parameter selection and optimization of larger samples are also challenging tasks, which may increase the complexity of the model.

To improve the diagnostic accuracy rate, sparse representation (SR) theory has been proposed to resolve the problem of fault diagnosis. The SR uses a dictionary that is predefined or learned from the underprocessed data to represent the signal, which has greater flexibility and adaptivity than the traditional orthogonal transform. More importantly, the identifying information in the SR vector is convenient for pattern classification. Since the advent of SR, it has been intensively studied and widely used in signal processing (e.g., image processing [17, 18], compression sensing [19], and fault diagnosis [20, 21]). Meanwhile, the sparse characteristics of dictionary-based SR are also applied in various classification tasks. In [22], a classification model named the SR-based classification was devised and achieved an exciting result for face classification. However, the investigation shows that the sparsity of SR classification may experience the leakage of weak fault features, which leads to bad classification performance in the case of complicated data with multiple features.

To capture the weak features, collaborative representation (CR) was devised by Zhang et al. to conduct classification in a more effective way than the  $\ell_1$  norm sparsity (i.e., the SR classification), and it is less sparse than SR classification [23]. Since the birth of CR, it has been widely applied, and a great number of methods based on CR have been devised and successfully employed in visual classification [24–28]. Nevertheless, CR-based algorithms that are used for fault classification are few and far between. The success of CR classification lies in the fact that the acquired

images with natural similarity could be directly used as samples to produce a redundant dictionary [29], and the samples belonging to other classes can also be employed to express any test example. The dictionary is elaborately designed to make the atoms collaborate together to obtain optimal coefficients that can yield the minimum residuals for the classification. In the framework of CR classification, dictionary learning is considered to be more important than similar feature extraction; therefore, it can also be viewed as a learning method based on similar features. However, there are no such similarities in the raw vibration signals of rolling bearings, and thus a specific class of signal may not be accurately represented by all the samples that are contained in the dictionary. Fortunately, some artificial similarity can be generated by using some available feature extraction methods such as fast Fourier transform and wavelet transform. In this work, wavelet transform was employed to build similar features, and a feature matrix whose columns are formed using the wavelet coefficients was also constructed. Since the features are obtained directly from the signals, the learned dictionary has great flexibility. However, in the case where the training matrix contains plenty of similar samples, the traditional learning method will lead to a redundant dictionary with high correlation, which will result in poor signal reconstruction performance, and the ability of CR is largely affected by the capacity of the trained dictionary. Therefore, to achieve perfect signal reconstruction, a new strategy called dictionary pair learning (DPL) that was proposed in [30] is used to train a pair of dictionaries to analyze and synthesize the signals.

Inspired by the DPL and the traditional CR classification, we proposed an improved CR classification method to diagnose bearings in this paper. The main novelty of our work consists of two aspects: (1) the features are directly extracted from the input raw signals using the wavelet transform, and the feature matrix is trained by the DPL; and (2) in the fault diagnosis phase, the sparse and projection coefficients are combined to enhance the robustness of the CR classification and accuracy. Since the CR represents a query sample using the dictionary that comes from all of the raw samples without undergoing training, the CR-based classification method lacks robustness. After combining DPL and CR, the performance of the CR classification is improved and the collaboration between different categories is further enhanced. By using the proposed method, the problem of classification with small samples can also be properly resolved. Under the trained DPL, the case in which a query sample is represented using an incomplete subdictionary can be avoided, which will have a positive effect on the results of the fault diagnosis. The complete flow chart of the proposed diagnostic model is shown in Figure 1. Since the process of dictionary learning is time-consuming when the samples are highly dimensional, the PCA technique is employed to reduce the dimensionality before dictionary training, which can reduce the computational costs while maintaining the main features of the data. In fact, the function of PCA is mainly to remove the redundant components of the feature vectors.

The rest of the contents of this paper will be structured as follows. In Section 2, the principles of the SR and CR classification are reviewed, and the details of the improved augmented CR classification algorithm based on DPL are also described. In Section 3, quantitative evaluations and comparisons are presented on the basis of experimental results from the simulated and real vibration signals. Finally, we summarize the paper in Section 4.

## 2. CR Classification Based on DPL

**2.1. Brief Introduction of CR Classification.** SR is an effective signal representation method which represents a sample using a linear combination of training samples that belong to the same type. Suppose that the all samples in training set  $T$  are arranged to form an initial dictionary  $C'$ , that is,

$$C' = [C'_1, \dots, C'_j, \dots, C'_m] = [A_{1,1}, \dots, A_{j,1}, \dots, A_{m,n_j}], \quad (1)$$

where  $C'_j = [A_{j,1}, \dots, A_{j,i}, \dots, A_{j,n_j}] \in \mathbb{R}^{r \times n_j}$  comes from the  $j$ -th class that contains  $n$  samples and has  $r$  dimensions;  $m$  is the number of classes, and the total number of samples is  $\sum_{j=1}^m n_j$ . Thus, for any one test example  $q$  belonging to the  $j$ -th class in test set  $Q$ , it can be faithfully expressed by using those samples coming from the  $j$ -th class:

$$q = a_{j,1}A_{j,1} + a_{j,2}A_{j,2} + \dots + a_{j,n_j}A_{j,n_j}, \quad (2)$$

where  $a = [a_{j,1}, \dots, a_{j,n_j}]$  is the coefficient vector. However, in most cases, the number of training samples in any dataset  $T$  is limited and the representation of  $q$  only using  $C'_j$  is not optimal. Thus, it is a wise strategy to represent  $q$  using  $C'_j$  combined with other classes.

The innovation of CR is that it applies all samples in a dictionary to achieve an optimal representation for  $q$ , which can subsequently obtain a minimum reconstruction residual to determine the given query sample to which a category belongs. Therefore, to achieve the minimum reconstruction residual, the CR uses the  $\ell_2$  norm as the regularization term instead of the  $\ell_1$  norm. Thus, the CR problem can be formulated as

$$\hat{\alpha} = \arg \min_{\alpha} \|q - C' \alpha\|_2^2 + \rho \|\alpha\|_2^2, \quad (3)$$

where  $\rho$  is used for balancing the residual and the sparsity of the solution, the coefficient vector  $\hat{\alpha} = [\hat{\alpha}_1; \dots; \hat{\alpha}_j; \dots; \hat{\alpha}_m]$  corresponds to the training samples that are located in all categories, and  $\|\cdot\|_2$  denotes the  $\ell_2$  norm that introduces some sparsity and achieves a steady solution. Since the objective function in (3) is a quadratic and derivable function, it has a closed-form solution  $\hat{\alpha} = (C'^T C' + \rho \cdot I)^{-1} C'^T q$  and the computational complexity is low. Then, the residual  $\|q - C'_j \hat{\alpha}_j\|_2^2$  that is generated from the  $j$ -th class can be directly calculated using  $\hat{\alpha}$ , and the class that produces the minimum residual is considered as the category to which the  $q$  belongs. Therefore, the CR classification algorithm can be presented as Algorithm 1.

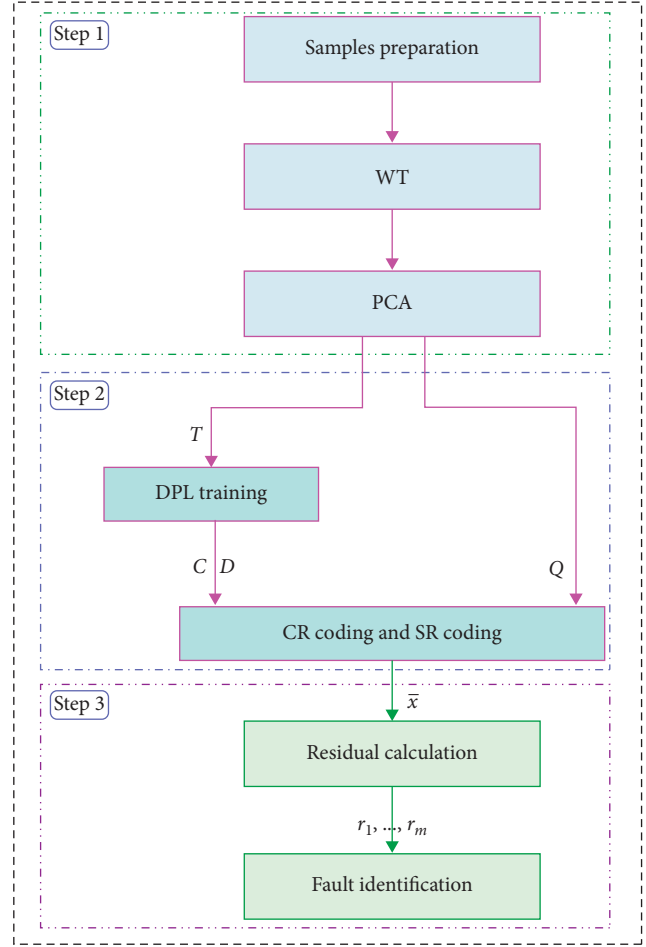


FIGURE 1: Flowchart of the proposed diagnostic model.

## 2.2. Dictionary Learning for DPL

**2.2.1. Dictionary Learning Model.** The collected vibration signals are always contaminated by noise, and the noise is superimposed on every point of the underlying vibration signal. Since the frequencies of noise mainly focus on the high-frequency band, they can be separated from the main content of the underlying vibration signal after wavelet decomposition, and the low-frequency coefficients of each sample will be more similar than their original forms. In this paper, we construct the feature matrix  $C_j \in \mathbb{R}^{r \times n_j}$  (subdictionary) using the coefficients of the wavelet transform computed directly from the raw bearing signals. Even though the dictionary  $C$  is better than that of the one formed by original raw samples, it is only a predefined nonadaptive dictionary. Thus, to obtain better results, we expect to construct a synthesis dictionary  $C = [C_1, \dots, C_j, \dots, C_m] \in \mathbb{R}^{r \times k}$  ( $k > r$ ) and an analysis dictionary  $D = [D_1; \dots; D_j; \dots; D_m] \in \mathbb{R}^{k \times r}$  by employing the DPL learning strategy, where  $C_j \in \mathbb{R}^{r \times n_j}$  is the learned subsdictionary that is used to reconstruct the raw signal, and  $D_j \in \mathbb{R}^{n_j \times r}$  is the learned projection subsdictionary that is used to analyze the raw signals from the  $j$ -th class.

Under each synthesis subsdictionary  $D_j$ , the coefficient matrix of corresponding dataset is  $D_j A_j$ . Then, the DPL

Input: dictionary  $C'$  and threshold  $\rho$

Main procedure:

- (1) Normalize the columns of  $C'$  to make them regularize ( $\ell_2$  norm)
  - (2) Let  $q$  be encoded on  $C$  through  $\hat{\alpha} = (C'^T C' + \rho \cdot I)^{-1} C'^T q$
  - (3) Calculate residuals  $r_j(q) = \|q - C'_j \hat{\alpha}_j\|_2 / \|\hat{\alpha}_j\|_2$ ,  $j = 1, \dots, m$
  - (4) Identify the class of  $q$  by using  $\text{TD}(q) = \text{argmin}_j r_j(q)$ , where  $\text{TD}(q)$  is the label of identified class to which the sample  $q$  belongs.
- Output: identified class to which  $q$  belongs.

ALGORITHM 1: Classification algorithm.

model can be obtained by solving the following optimization problem:

$$\langle \hat{C}, \hat{D} \rangle = \arg \min_{C, D} \sum_{j=1}^m \|A_j - C_j D_j A_j\|_2^2 + \mu \|D_j \bar{A}_j\|_2^2, \quad (4)$$

where  $\bar{A}_j$  is the complementary matrix of  $A_j$ . This dictionary learning model (4) is different from most of the current dictionary learning methods, where the term  $\sum_{j=1}^m \|A_j - C_j D_j A_j\|_2^2$  is the reconstruction error and the other term is a regularizer that is used to penalize the projection coefficients that are not related to  $A_j$ .

**2.2.2. Optimization.** On the basis of obtaining the features of the raw signals, a dictionary pair  $C$  and  $D$  can be found by using an appropriate optimization method. Since the expression of the objective function in (4) is nonconvex, we introduce a variable matrix  $V$  to obtain a closed-form solution. Thus, the optimization problem in (4) can be reformulated as

$$\langle \hat{C}, \hat{V}, \hat{D} \rangle = \arg \min_{C, V, D} \sum_{j=1}^m \|A_j - C_j V_j\|_2^2 + \lambda \|D_j A_j - V_j\|_2^2 + \mu \|D_j \bar{A}_j\|_2^2, \quad (5)$$

Obviously, the converted optimization problem in (5) has a closed-form solution. As with general multivariable optimization problems, it can be resolved iteratively through two alternative optimization steps. At the beginning of iteration, the corresponding parameters that are involved in the objective function need to be initialized. In this paper, the columns of dictionary pair  $C$  and  $D$  are randomly initialized using the  $\ell_2$  norm. In addition, to reduce the unnecessary time consumption, a stopping criterion is set; that is, when the residual is less than a very small positive value (e.g.,  $10e^{-5}$ ), the iterative operation terminates. When  $C$  and  $D$  are fixed, the optimization problem in (5) can be simplified to the following form:

$$\hat{V} = \arg \min_V \sum_{j=1}^m \|A_j - C_j V_j\|_2^2 + \lambda \|D_j A_j - V_j\|_2^2, \quad (6)$$

Since the objective function in (6) is separable, the variable  $V$  can be obtained by solving  $m$  suboptimization problems and each  $V_j$  has the following expression:

$$\hat{V}_j = (C_j^T D_j + \lambda \cdot I)^{-1} (\lambda D_j A_j^T + C_j^T A_j). \quad (7)$$

Similarly, when  $V$  is fixed,  $C$  and  $D$  can be obtained as follows:

$$\begin{cases} \hat{C} = \arg \min_C \sum_{j=1}^m \|A_j - C_j V_j\|_2^2, \\ \hat{D} = \arg \min_D \sum_{j=1}^m \lambda \|D_j A_j - V_j\|_2^2 + \mu \|D_j \bar{A}_j\|_2^2. \end{cases} \quad (8)$$

In addition, the optimal value of  $C$  and  $D$  can be solved class by class as follows:

$$\begin{cases} \hat{C}_j = A_j V_j^T (V_j V_j^T)^{-1}, & j = 1, 2, \dots, m, \\ \hat{D}_j = \lambda V_j A_j^T (\lambda A_j A_j^T + \mu \bar{A}_j \bar{A}_j^T + \gamma I)^{-1}, & j = 1, 2, \dots, m, \end{cases} \quad (9)$$

where  $\gamma$  is a very small positive value. According to the above analysis, we know that both  $C$  and  $D$  have closed-form expressions that can be efficiently computed in each iteration. Once  $C$  and  $D$  are obtained, we can use them to collaboratively represent the test samples. To assess the convergence of the DPL method, the relation between the value of objection function and iteration is illustrated in Figure 2.

**2.3. Proposed Classification Algorithm.** A classification algorithm based on collaborative representation using DPL is proposed in this subsection. The analysis dictionary  $D$  can produce a coding coefficient vector  $\hat{x} = Dq$  for the given sample  $q$ . In addition, the samples from the  $j$ -th category can be reconstructed via the synthesis subdictionary  $C_j$  using the coding coefficient vector  $\hat{x}_j$  with a fairly small residual  $\|q - C_j \hat{x}_j\|_2$ . Now, according to the CR classification, we can discriminate the fault classes by using the minimum residual. Since there are some correlations among the atoms in  $C$ , more than one linear combination can be used to represent the corresponding  $x$  with the same residual. Thus, to find the appropriate (or correct) linear combination that coincides with the actual situation (i.e., the sample belonging to the  $j$ -th class should be represented by using the atoms from the  $j$ -th class), we introduce the sparsity of the coding coefficient vector into the CR representation to augment the dense representation to improve the performance of CR-based classification [31].

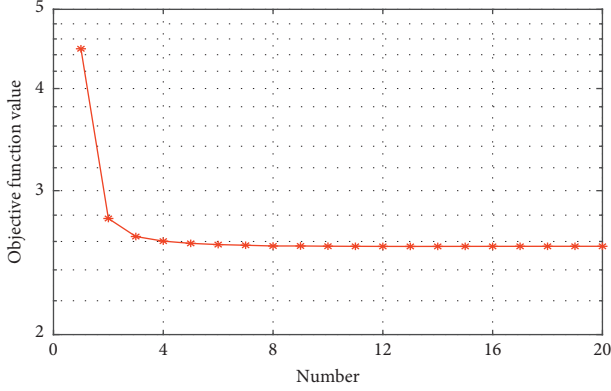


FIGURE 2: Convergence process of dictionary training.

To find a sparse representation, the orthogonal matching pursuit (OMP) [32] is used to perform the following optimization:

$$\begin{aligned} \tilde{x} &= \min_x \|q - Cx\|_2, \\ \text{s.t. } &\|x\|_0 \leq k_0. \end{aligned} \quad (10)$$

In general,  $\tilde{x}$  contains large positive coefficients whose indexes are related to the right category, and it has either small positive or negative coefficients for other indices. Thus, if  $\hat{x}$  and  $\tilde{x}$  are properly fused, the probability of finding the correct linear combination will be greatly enhanced. For simplicity, the improved CR coefficient is written as follows:

$$\bar{x} = \frac{\tilde{x} + \hat{x}}{\|\tilde{x} + \hat{x}\|_2}. \quad (11)$$

Obviously,  $\bar{x}$  amplifies the coefficients of the correct classes and reduces the relative energy of the unamplified coefficients; thus,  $C\bar{x}$  can effectively approximate the test sample. Meanwhile, this strategy also enhances the stability of CR classification. Subsequently, we can calculate the residuals corresponding to all the classes as follows:

$$r_j(q) = \|q - C_j \bar{x}_j\|_2^2, \quad j = 1, 2, \dots, m. \quad (12)$$

The test example is classified using  $\text{TD}(q) = \arg \min_j r_j(q)$ , where  $\text{TD}(q)$  is the label of the identified class to which the sample  $q$  belongs. To study the adaptiveness of the dictionaries learned from DPL to the test samples, the relation between training accuracy and each iteration of DPL training is shown in Figure 3. It can be seen that the accuracy rate gradually rises with the increase in iterations.

**2.4. Computational Complexity.** The complexity of the proposed algorithm is investigated in this subsection. Since the final classification step is a simple computation of the residuals, the vast majority of the computational complexity of the proposed method comes from the training process of a pair of dictionaries. Recall that  $C$ ,  $D$ , and  $V$  have closed-form solutions, and the computational costs of updating  $C_j$ ,  $D_j$ , and  $V_j$  are  $O(d^3 + rmd + d^2r)$ ,  $O(r^3 + rmd + dr^2)$ , and

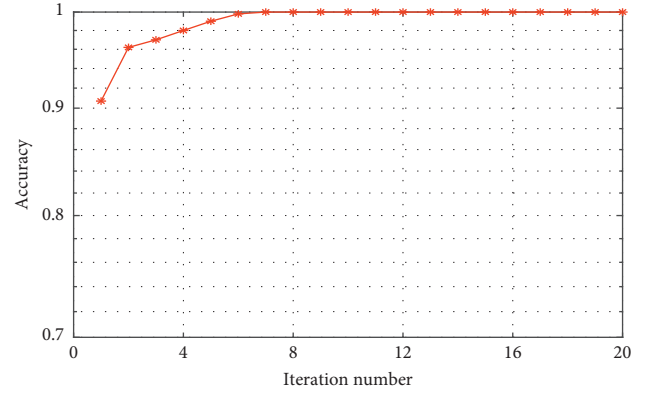


FIGURE 3: Accuracy rate at each iteration of the dictionary training.

$O(d^3 + rmd + d^2m)$  (where  $r$  denotes the dimension of the features,  $m$  is the number of samples in each class, and  $d$  is the number of atoms in each subdictionary), respectively, in each iteration. Among  $C_j$ ,  $D_j$ , and  $V_j$ , the largest computational cost is the update of  $D_j$  because the calculation of an  $r \times r$  matrix is involved in the training of  $D_j$  and the number of subdictionary atoms and training samples in each class are smaller than  $r$ . Fortunately, the factor  $(\lambda A_j A_j^T + \mu \bar{A}_j \bar{A}_j^T + \gamma I)^{-1}$  in  $D_j$  can be precomputed before the training process since it only relates to  $A_j$  and the other predetermined parameters, which can greatly reduce the training time. In summary, the complexity of the proposed algorithm is  $O(r^3)$ .

### 3. Experiments

In this section, one set of simulated vibration signals, three sets of laboratory bench signals, and one set of real-operating signals were constructed to evaluate the performance of the proposed algorithm. To exhibit the advantages of the proposed algorithm, four other SVM comparison classification methods which include the SVM [13],  $k$  nearest neighbor (KNN) [33], decision tree (DT) [34], and back-propagation neural network (BPNN) [35] are also employed to conduct the experiments. The wavelet used in our method is the Haar wavelet, and the reason is that (1) the Haar wavelet has the shortest filter length which can avoid excessive smoothing of the low-frequency component of signal and effectively ensure that the main features are not lost; and (2) Haar wavelet transform has the character of easy calculation, which ensures that the construction of feature matrix using wavelet coefficients is very efficient. Since the parameters have important impacts on the performance of a given method, the parameters involved in the proposed algorithm were optimized by hand to obtain the best results in the experiments; that is, the corresponding values were set as  $\lambda = 0.05$ ,  $\gamma = 0.003$ ,  $\mu = 0.001$ ,  $\rho = 10e^{-3}$ , and  $k_0 = 50$ , respectively. Moreover, 70 atoms were chosen in each subdictionary, which are randomly selected from all the samples in each class. For the four other machine learning algorithms that were used for the comparison, the optimal values of the free parameters were taken from the reference papers.

As to the evaluation of performance of the proposed method, an indicator called the accuracy (Acc) is adopted in this paper, which is defined as

$$\text{Acc} = \frac{\sum_{i=1}^N I(\text{ID}_i = \text{TL}_i)}{N} \times 100\%, \quad (13)$$

where  $N$  denotes the number of test samples,  $I(\cdot)$  represents the indicator function, and  $\text{ID}_i$  and  $\text{TL}_i$  represent the predicted label and the true label of the  $i$ -th test sample, respectively. In general, the Acc is calculated by using a confusion matrix [36].

**3.1. Experiment Using Simulated Data.** In this subsection, a set of simulated signals are generated to test the proposed method. In general, when a rotating bearing has a fault in some key components, the vibration of the bearing will be intensified and impact the performance, which results in the instantaneous high-frequency resonance of the bearing system. Therefore, the vibration signals of some failures are usually characterized using impulsive series. According to this, we generate the simulated signals using the following model [37]:

$$\begin{aligned} y(t) &= r(t) + e(t), \\ r(t) &= r_i(t)r_{f_a}(t) + r_a(t), \end{aligned} \quad (14)$$

where  $r_i(t)$  is an impulsive component,  $r_{f_a}(t)$  is the resonance component,  $r_a(t)$  is the vibration component that is produced by the mechanical equipment, and  $e(t)$  denotes the additive Gaussian noise. In this experiment, six different frequencies (corresponding to one normal state and five fault states) were used to generate six fault vibration signals which are shown in Figure 4. The vibration signals are formulated as follows:

$$\begin{aligned} \text{normal: } y(t) &= 0.02 \cos(20\pi \times t) \sin(2\pi f_a \times t) \\ &\quad + 0.02 \cos(20\pi \times t), \\ \text{faults: } y_j(t) &= e^{-nt_m} \sin(2\pi f_a \times t) \\ &\quad + 0.02 \cos(20\pi \times t), \quad j = 1, 2, 3, 4, 5, \\ t_m &= \text{mod}\left(t_d, \frac{1}{f_m}\right), \quad m = 1, 2, 3, 4, 5, \end{aligned} \quad (15)$$

where  $n$  denotes the damping factor contributing to the signal that affects the pulse duration,  $f_a$  and  $f_m$  denote carrier centers and characteristic frequencies, respectively, and  $t_d$  is the duration of the simulated signal. The sampling frequency was set to 12 kHz, and  $10^5$  points were collected for the simulated signal under the condition that the value of the damping factor is 500. Considering the stability of the classification and the simplicity of the training, 200 samples (each sample contains 2000 points) were generated from the simulated signal for each class, in which 100 samples were randomly selected to train the subdictionary and the other 100 were used as test samples. After the PCA, the dimension

of the features was reduced to 300. In addition, under the trained dictionary pair, the reconstruction precisions of each test sample are assessed using all subdictionaries  $C_j$  and  $D_j$  in the experiment, and the corresponding errors are shown in Figure 5. It can be clearly seen that when a test sample belongs to a certain class, it produces the smallest error for that corresponding class because almost all of the energy of the given test example is projected in its corresponding analysis subdictionary  $D_j$ , which ensures that the reconstructed sample using the synthesis subdictionary  $C_j$  closely approximates the original sample.

Moreover, the DPL greatly contributes to the accuracy of the CR-based classification methods. A comparative experiment was conducted to verify this point, and the results are reported in Figure 6. From the figure, one can clearly see that the accuracy that is obtained by the proposed algorithm is 100% in all cases, while the result using the traditional CR classification without the DPL strategy is only approximately 91% and is also unstable. Meanwhile, it should be noted that the sparsity constraint that is involved in DPL also makes a certain contribution to the classification accuracy.

By applying the four comparison methods on the simulated samples, we further investigate the performance of the proposed method. The results that are obtained by our algorithm and the comparison methods are listed in Table 1. From the table, it can be seen that the classification accuracy obtained by our algorithm is still 100% in all cases, while the results produced by other four methods are all lower than ours. This result indicates that the bearing states can be precisely classified using our algorithm. The reason is that the features are directly extracted from the raw data without an artificial operation, and the method is adaptive to the input data. By carefully analyzing Table 1, we know that the performance of the DT algorithm is the worst among the used methods. This is because the features in DT are extracted by finding the smallest error after an artificial operation. Obviously, this minimum value is not accurate enough and it will lower the accuracy of DT, especially when there is more than one class. In addition, the classification accuracies of the other three methods are between the DT method and the proposed algorithm. The SVM achieves the second best results for all the fault states. The main reason is that only 100 samples are tested for each class, which is a small-sample problem at which the SVM is particularly skilled. However, the problem that is encountered in this paper is not just a small-sample problem, but is also a multiclass problem; thus, the SVM only outperforms the KNN, BPNN, and DT. Meanwhile, skewed datasets that arise in the multiclass problem also degrade the performance of the SVM. Moreover, the confusion matrix that is used to compute the Acc for the proposed method is listed in Figure 7. The elements on the main diagonal represent the correct classification accuracy for each state, while the elements on the other places denote the wrong classification accuracy. Figure 7 shows that the values on the main diagonal are all equal to 1, which means that all the test samples are correctly classified by our method.

Finally, to compare the computational burden of our method with other four methods, the running times are

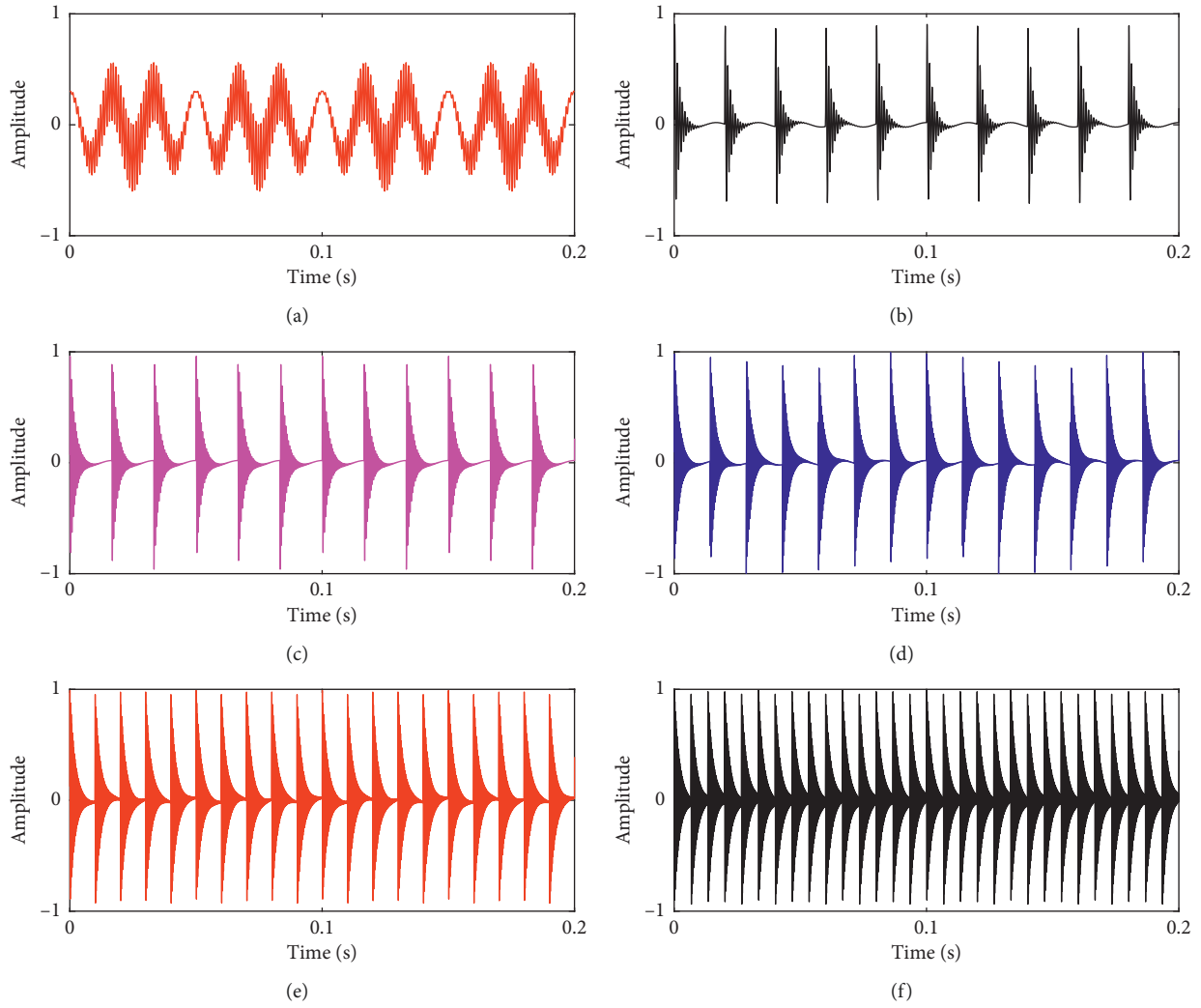


FIGURE 4: Simulated vibration signals with different parameters: (a) normal signal using  $f_a = 1000$  Hz; (b) fault signal 1 using  $f_a = 1000$  Hz and  $f_m = 50$  Hz ( $m = 1$ ); (c) fault signal 2 using  $f_a = 2000$  Hz and  $f_m = 60$  Hz ( $m = 2$ ); (d) fault signal 3 using  $f_a = 3000$  Hz and  $f_m = 70$  Hz ( $m = 3$ ); (e) fault signal 4 using  $f_a = 4000$  Hz and  $f_m = 100$  Hz ( $m = 4$ ); (f) fault signal 5 using  $f_a = 4500$  Hz and  $f_m = 150$  Hz ( $m = 5$ ).

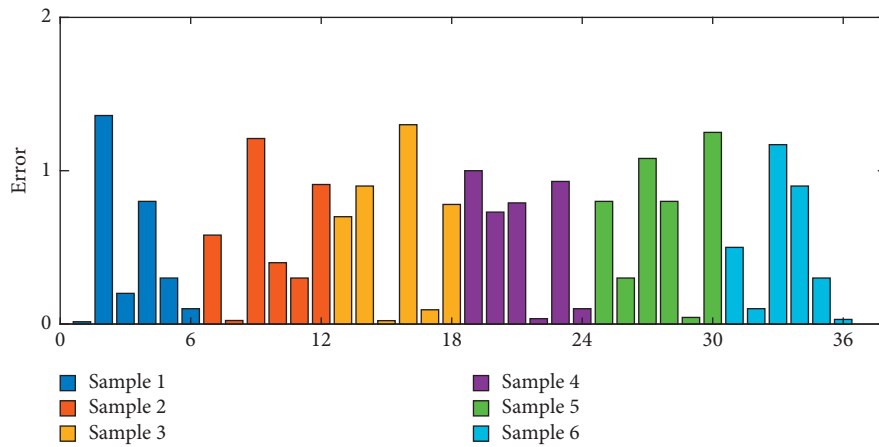


FIGURE 5: Reconstruction errors of six test samples using six subdictionaries (corresponding to 6 classes).

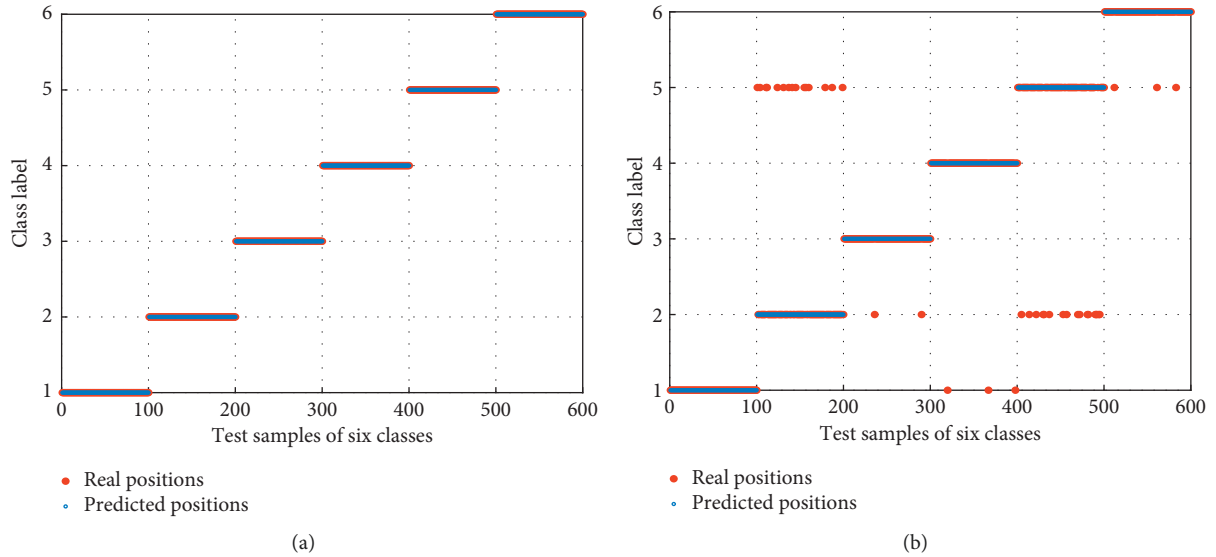


FIGURE 6: Classification results of the simulated signals: (a) results without dictionary learning; (b) results using dictionary learning.

TABLE 1: Recognition rates obtained by different classification methods on simulated signals.

Method	SVM	BPNN	KNN	DT	Ours
Feature dimension	300	300	300	300	300
Acc (%)	98.2	92.1	94.5	83.7	100.0

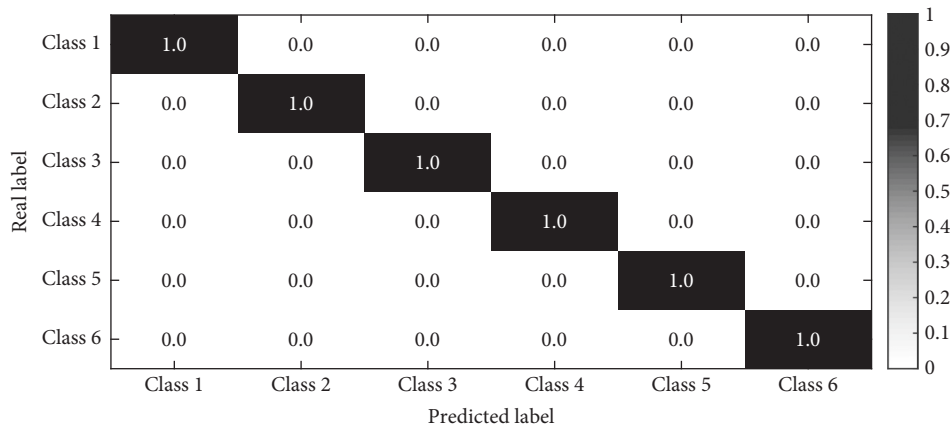


FIGURE 7: Confusion matrix obtained by the proposed method for the simulated signals.

assessed using Matlab (R2018a, 64 bit) on a personal computer with an i5-7500 CPU (3.40 GHz) and 8.00 GB of memory. The running times that are consumed by the SVM, KNN, DT, BPNN, and ours method are 74.73 s, 126.03 s, 4.07 s, 80.86 s, and 3.85 s, respectively. Obviously, the proposed method shows significant superiority over the comparison algorithms. This is because (1) the feature extraction step is fairly simple, which dramatically decreases the computational time and (2) the proposed method does not need to estimate too many parameters, which also saves a large portion of the execution time.

**3.2. Experiment on Real Data.** In this subsection, two types of real data are employed to further evaluate the performance of

the proposed method. One is the vibration signals that are generated from bearings with predefined faults, and the other is generated from bearing without predefined faults. The parameters that are involved in the tested methods are the same as those that are used in the above simulation experiment.

**3.2.1. Real Data from Bearings with Predetermined Faults.** The used experimental data of rolling bearings are provided by Case Western Reserve University [38]. First, we provide a brief introduction of the equipment (Figure 8) that is used to generate the real vibration signals. The objective bearings (6203 wind turbine end bearing (FE) and 6205 drive end bearing (DE)) with single faults were tested under different



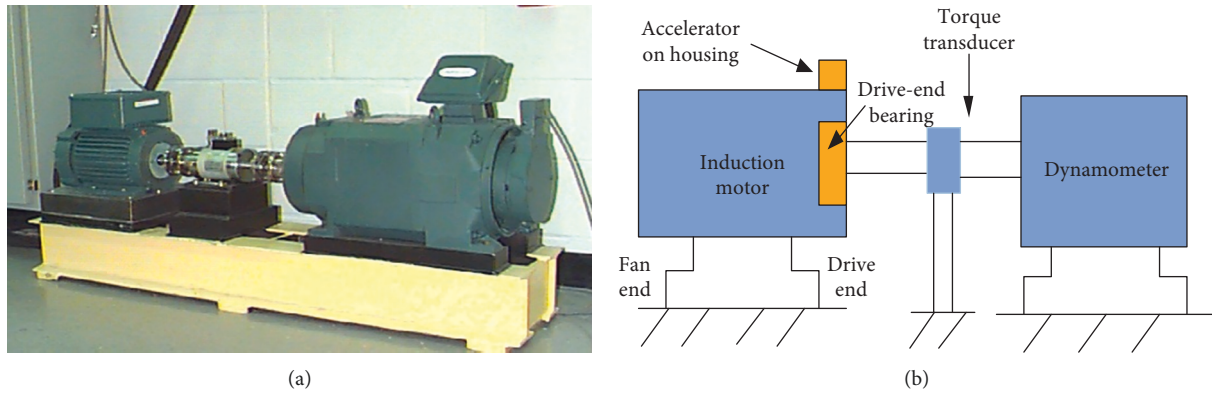


FIGURE 8: Classification results of the simulated signals: (a) results without dictionary learning; (b) results using dictionary learning.

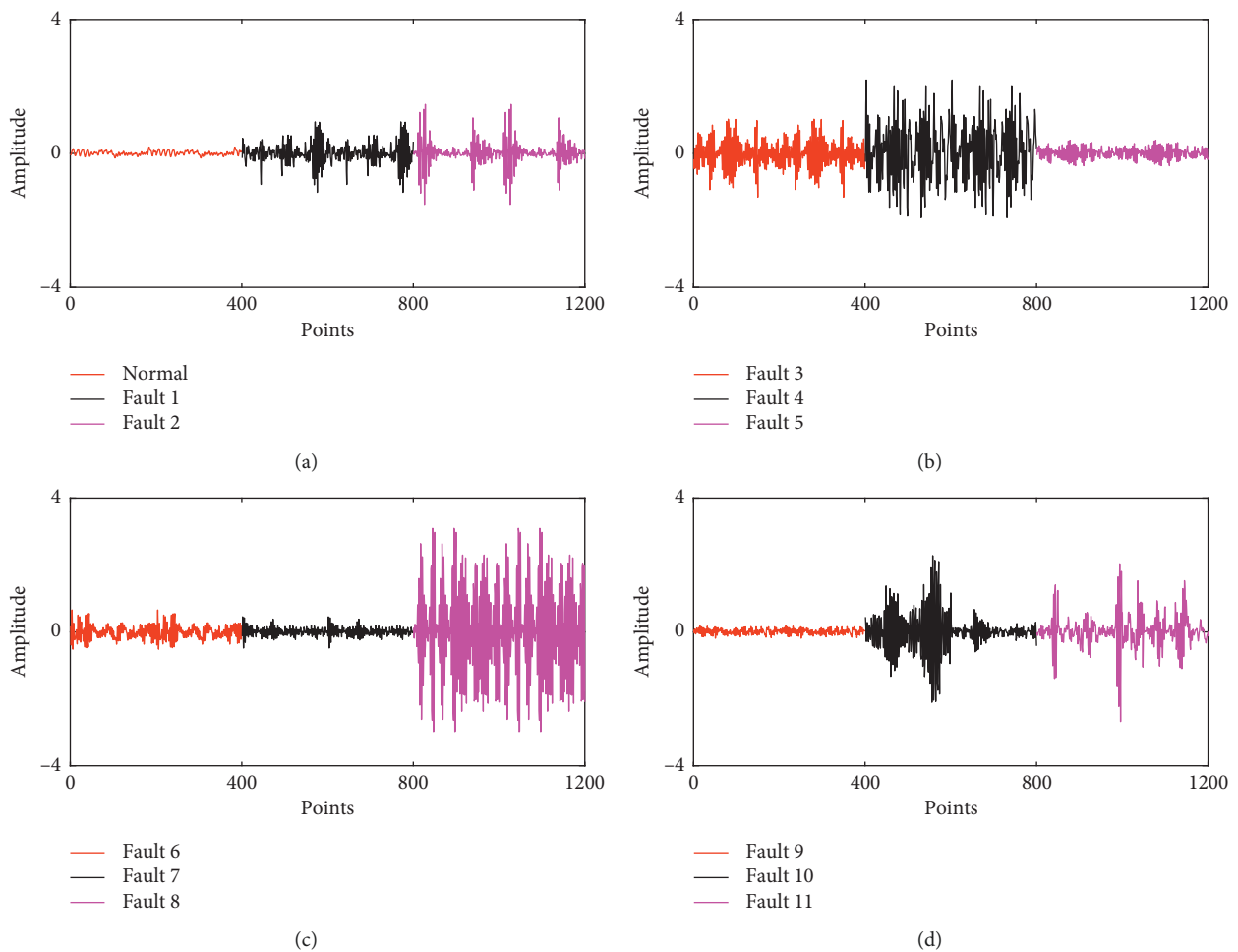


FIGURE 9: Vibration signals under normal and different fault states: (a) normal and faults 1-2; (b) faults 3-5; (c) faults 6-8; (d) faults 9-11.

loads (0–3 horsepower). The single faults with 4 different defect diameters (0.028, 0.021, 0.014, and 0.007 inches) are introduced to the rolling bearings using electrodischarge machining. Two sampling frequencies (12 kHz and 48 kHz) were used to collect the vibration signals. In this experiment, we selected twelve running states of rolling bearings with different types of faults and fault severities. To visually observe the real vibration signal, twelve raw signals (DE with

different faults) with a sampling frequency 12 kHz are illustrated in Figure 9. Based on the raw fault signals, we generated 200 samples (each sample contains 2000 points) for each class, and the detailed information for the generation of raw samples is reported in Table 2.

Once the samples were acquired, we randomly and equally divided them into two sets. One set (100 samples) is used to train the dictionaries, and the other (100 samples) is

TABLE 2: Detail information of the drive end for the generation of raw samples.

States	Diameter (inch)	Number of samples				Class
		1797 rpm (0 HP)	1772 rpm (1 HP)	1750 rpm (2 HP)	1730 rpm (3 HP)	
Normal	0.000	50	50	50	50	1
Ball fault	0.007	50	50	50	50	2
	0.014	50	50	50	50	3
	0.021	50	50	50	50	4
	0.028	50	50	50	50	5
Inner-race fault	0.007	50	50	50	50	6
	0.014	50	50	50	50	7
	0.021	50	50	50	50	8
	0.028	50	50	50	50	9
Outer-race fault	0.007	50	50	50	50	10
	0.014	50	50	50	50	11
	0.021	50	50	50	50	12

TABLE 3: Recognition rates obtained by different classification methods on real signals.

Method	Feature dimension	Acc (%)			Average Acc (%)
		DE (12 kHz)	DE (48 kHz)	FE (12 kHz)	
SVM	300	97.6	96.7	95.4	96.6
BPNN	300	90.8	91.0	93.1	91.6
KNN	300	95.4	94.7	95.6	95.2
DT	300	80.5	82.7	82.1	81.7
Ours	300	100.0	100.0	100.0	100.0

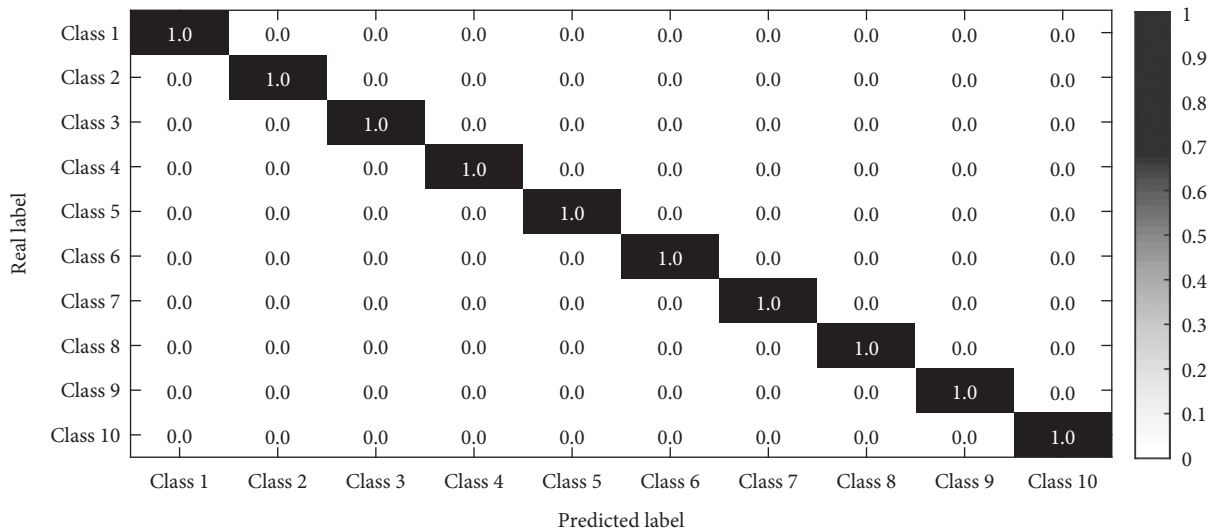


FIGURE 10: Confusion matrix obtained by the proposed method for the vibration signals in Figure 9.

used to test the proposed model. To obtain reliable results, this experiment was repeated 30 times to produce averaged results as the final output. The Accs that are obtained by our method and the other comparison methods are listed in Table 3, and the corresponding confusion matrix for our method is shown in Figure 10. It can be clearly seen that the classification accuracy that is obtained by our work is still 100%, which indicates that the performance of the proposed method is very effective and stable. By observing Table 3, one can find that there is a sharp distinction between the proposed work and the DT; namely, our work gets the best result, and the DT obtains the lowest accuracy in all fault

cases. Overall, the results that are obtained in this experiment accord with those of the simulated experiment; that is, our method is also effective for real vibration signals and outperforms other classification methods.

### 3.2.2. Real Data from Bearings without Predetermined Faults.

To prove the performance of our method for bearings in real operating situation, the vibration signals from the accelerated degradation of bearings (without predetermined faults) were assessed in this subsection. The data are provided by the Changxing Sumyoung Technology and the

TABLE 4: Parameters of bearing used for accelerated degradation.

Parameter	Value	Parameter	Value
Outer-race diameter	39.80 mm	Inner-race diameter	29.30 mm
Bearing mean diameter	34.55 mm	Ball diameter	7.92 mm
Number of balls	8	Contact angle	0°
Load rating	6.65/12.82 kN	RPM	2100/2250/2400

TABLE 5: Recognition rates obtained by different classification methods on real vibration signals from accelerated degradation of bearing.

Method	KNN	SVM	DT	BPNN	Ours
Feature dimension	300	300	300	300	300
Acc (%)	95.4	97.6	89.5	92.6	100.0

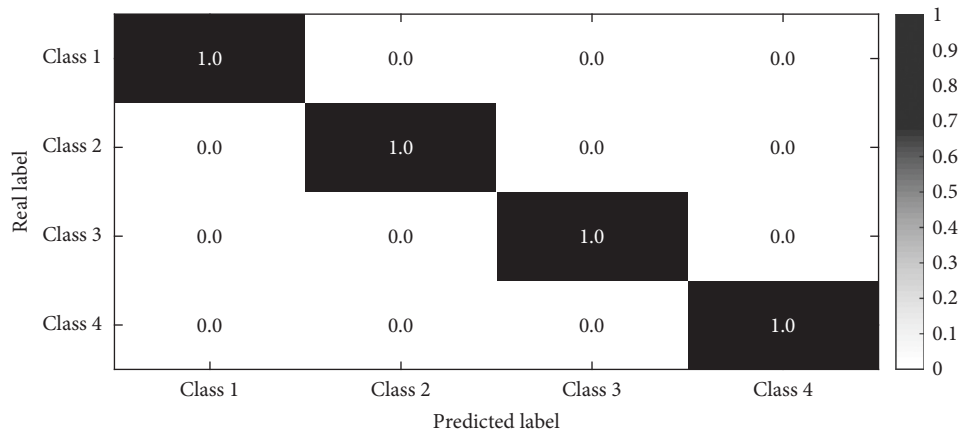


FIGURE 11: Confusion matrix obtained by the proposed method for vibration signals from accelerated degradation of bearing.

Xi'an Jiaotong University ([https://www.mediafire.com/folder/m3sij67rizpb4/XJTU-SY\\_Bearing\\_Datasets](https://www.mediafire.com/folder/m3sij67rizpb4/XJTU-SY_Bearing_Datasets)). The vibration signals were obtained by running a bearing from a normal state until reaching the various faults under different conditions, which can accurately reflect the real operating situations of bearings. The 4 different fault conditions were adopted in this experiment, which are inner-race fault, cage fault, ball fault, and outer-race fault. For simplicity, we labeled them as 1, 2, 3, and 4, respectively, and the parameters of the bearings are listed in Table 4. In addition, 200 samples (each sample contains 2000 points) were generated for each fault class, in which 100 samples were randomly selected to train the subdictionary, and the other 100 were used as test samples.

The accuracies that were obtained by all the fault diagnosis methods are reported in Table 5, and the confusion matrix for the proposed method is illustrated in Figure 11. As presented in the previous experiments, our proposed method acquires a 100% recognition rate in all fault states, the DT has the worst performance, and the Accs that are obtained by the rest are between the proposed method and the DT algorithm. These results agree with the results of the simulated experiment. The results also indicate that our method can identify the fault states of the bearings operating in real situations.

## 4. Conclusion

In the paper, a classification method using CR based on DPL has been proposed to detect and classify the states of bearings. We construct a feature matrix that is used to train the dictionary pair, and the Haar wavelet was employed to decompose the raw samples. Two operations are adopted to improve the accuracy of classification: one is the introduction of the sparse constraint into the dictionary training of DPL, and the other is the fusion of sparse coefficients and projection coefficients. The results that are obtained by testing the simulated vibration signals and the real vibration signals indicate that the proposed work provides fairly good classification effects in all cases, and the recognition performance is very stable. However, since the OMP algorithm is exploited to find the sparse vector, the computation burden of our method is also increased. Moreover, the size of training samples used in the proposed method is relatively small. In fact, long samples are common in many practical situations, and a DPL-based fault diagnosis algorithm that considers samples with large size will be studied in the future.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was funded by the Key Science Program of Anhui Education Department (KJ2018A0012) and the Research Fund for Doctor of Anhui University (J01003266) and was also supported by the National Natural Science Foundation of China (NSFC) (61402004 and 61601003).

## References

- [1] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 469–480, 2014.
- [2] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol. 32, no. 8, pp. 469–480, 1999.
- [3] S. A. McInerney and Y. Dai, "Basic vibration signal processing for bearing fault detection," *IEEE Transactions on Education*, vol. 46, no. 1, pp. 149–156, 2003.
- [4] J. R. Stack, R. G. Harley, and T. G. Habetler, "An amplitude modulation detector for fault diagnosis in rolling element bearings," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 5, pp. 1097–1102, 2004.
- [5] H. R. Martin and F. Honarvar, "Application of statistical moments to bearing failure detection," *Applied Acoustics*, vol. 44, no. 1, pp. 67–77, 1995.
- [6] R. B. W. Heng and M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition," *Applied Acoustics*, vol. 53, no. 1–3, pp. 211–226, 1998.
- [7] W. Liu, J. Han, and X. Lu, "A new gear fault feature extraction method based on hybrid time-frequency analysis," *Neural Computing and Applications*, vol. 25, no. 2, pp. 387–392, 2014.
- [8] D. Yu, Y. Yang, and J. Cheng, "Application of time–frequency entropy method based on Hilbert–Huang transform to gear fault diagnosis," *Measurement*, vol. 40, no. 9–10, pp. 823–830, 2007.
- [9] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using continuous wavelet transform," *Applied Soft Computing*, vol. 11, no. 2, pp. 2300–2312, 2011.
- [10] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mechanical Systems and Signal Processing*, vol. 41, no. 1–2, pp. 127–140, 2013.
- [11] Z. Wang, C. Lu, Z. Wang, H. Liu, and H. Fan, "Fault diagnosis and health assessment for bearings using the Mahalanobis-Taguchi system based on EMD-SVD," *Transactions of the Institute of Measurement and Control*, vol. 35, no. 6, pp. 798–807, 2013.
- [12] Z. Liu, W. Guo, J. Hu, and W. Ma, "A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and twin SVM," *ISA Transactions*, vol. 66, pp. 249–261, 2017.
- [13] X. Zhang, B. Wang, and X. Chen, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," *Knowledge-Based Systems*, vol. 89, pp. 56–85, 2015.
- [14] Y. Li, M. Xu, Y. Wei, and W. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80–94, 2016.
- [15] 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.
- [16] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: revisited," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 201–208, 2016.
- [17] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.
- [18] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.
- [19] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [20] Z. Du, X. Chen, H. Zhang, and R. Yan, "Sparse feature identification based on union of redundant dictionary for wind turbine gearbox fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6594–6605, 2015.
- [21] Z. Feng, Y. Zhou, M. J. Zuo, F. Chu, and X. Chen, "Atomic decomposition and sparse representation for complex signal analysis in machinery fault diagnosis: a review with examples," *Measurement*, vol. 103, pp. 106–132, 2017.
- [22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [23] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2224–2231, Providence, RI, USA, June 2012.
- [24] J. Waqas, Z. Yi, and L. Zhang, "Collaborative neighbor representation based classification using  $l_2$ -minimization approach," *Pattern Recognition Letters*, vol. 34, no. 2, pp. 201–208, 2013.
- [25] P. Zhu, L. Zhang, Q. Hu, and S. C. K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Proceedings of the European Conference on Computer Vision*, pp. 822–835, Florence, Italy, October 2012.
- [26] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2200–2208, 2014.
- [27] R. Timofte and L. Van Gool, "Weighted collaborative representation and classification of images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, pp. 1606–1610, Tsukuba, Japan, November 2012.
- [28] C. Zhang, J. Cheng, and Q. Tian, "Multiview label sharing for visual representations and classifications," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 903–913, 2018.
- [29] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *Proceedings of the IEEE International Conference on*

- Computer Vision*, pp. 471–478, Barcelona, Spain, November 2012.
- [30] S. Gu, L. Zhang, W. Zuo et al., “Projective dictionary pair learning for pattern classification,” in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 793–801, Montreal, Canada, 2014.
  - [31] N. Akhtar, F. Shafait, and A. Mian, “Efficient classification with sparsity augmented collaborative representation,” *Pattern Recognition*, vol. 65, pp. 136–145, 2017.
  - [32] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
  - [33] 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 Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1793–1803, 2016.
  - [34] N. E. I. Karabadi, H. Seridi, I. Khelif, N. Azizi, and R. Boulkroune, “Improved decision tree construction based on attribute selection and data sampling for fault diagnosis in rotating machines,” *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 71–83, 2014.
  - [35] Z. Zhao, Q. Xu, and M. Jia, “Improved shuffled frog leaping algorithm-based BP neural network and its application in bearing early fault diagnosis,” *Neural Computing and Applications*, vol. 27, no. 2, pp. 375–385, 2016.
  - [36] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*, Springer, Berlin, Germany, 2010.
  - [37] H. Zhang, X. Chen, Z. Du, and B. Yang, “Sparsity-aware tight frame learning with adaptive subspace recognition for multiple fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 94, pp. 499–524, 2017.
  - [38] <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

