Intelligent diagnosis of rolling bearing compound faults based on device state dictionary set sparse decomposition feature extraction–hidden Markov model

HongChao Wang¹,² and WenLiao Du¹,²

Abstract
Identification of rolling bearing fault patterns, especially for the compound faults, has attracted notable attention and is still a challenge in fault diagnosis. Intelligent diagnosis method is an effective method for compound faults of rolling element bearing, and effective fault feature extraction is the key step to decide the intelligent diagnosis result to some extent. The sparse decomposition method could capture the complex impulsive characteristic components of rolling bearing more effectively than the other time–frequency analysis method when compound fault arises in rolling bearing. Based on the self-learning dictionary under different operating states of the device corresponding to the special features modes, an intelligent diagnosis method of rolling bearing compound faults based on device state dictionary set sparse decomposition feature extraction–hidden Markov model is proposed in the article. First, characteristic dictionaries of rolling bearing under different operating conditions are extracted by sparse decomposition self-learning method, and state dictionary set of rolling bearing is constructed. Then, the compound fault signals of bearing are transformed into sparse domain using the constructed dictionary set to extract sparse features. At last, the extracted sparse features are used as training and testing vectors of hidden Markov model, and satisfactory intelligent diagnosis results are obtained. The validity of the proposed method is verified by compound faults of rolling element bearing. In addition, the advantages of the proposed method are also verified by comparing with the other feature extraction and intelligent diagnosis methods, and the proposed method provides a feasible and efficient solution for fault diagnosis of rolling bearing compound faults.

Keywords
Intelligent diagnosis, rolling bearing, compound fault, dictionary set, sparse decomposition, hidden Markov model

Date received: 21 January 2020; accepted: 29 April 2020

Handling Editor: James Baldwin

Introduction
The rolling element bearing is one of the most commonly used rotating components, and it is meaningful to study effective fault diagnosis methods to avoid catastrophic accident. The fast Fourier transform (FFT) and envelope demodulation are the two classical signal-processing methods for this purpose. In recent years, kinds of related new signal-processing methods for fault diagnosis of rolling element bearing have been arising such as wavelet transform (WT),¹ tunable
quality-factor wavelet transform (TQWT),\textsuperscript{2} ensemble empirical mode decomposition (EEMD),\textsuperscript{3} Spectral kurtosis (SK),\textsuperscript{4} and modified fast kurtogram.\textsuperscript{5} However, most of the above techniques are only effective for single fault of rolling bearing, which means that if faults occur in different parts of rolling element bearing at the same time, most of the above methods would not work effectively: (1) coupling effect phenomenon will exist in the compound faults usually. (2) The gear-meshing signal with its harmonics will make a big interference if compound faults arise in the gearbox. Fault diagnosis of rolling bearing compound fault is not only challenging but also is a hot area. Amounts of literature arise in recently years, and most of them are mainly focusing on classification of compound faults using intelligent algorithms. The support vector machine (SVM) was improved in paper\textsuperscript{6} and the one-against-all multiclass support vector machine (MSVM) method was proposed, and which was combined with heterogeneous feature models and used in multiple combined fault diagnosis of bearings. The experimental results showed that the classification performance could be improved effectively by the proposed method. A step-by-step compound faults diagnosis method for equipment based on majorization–minimization (MM) and constraint sparse component analysis (SCA) was proposed to separate the compound faults.\textsuperscript{7} A novel method called deep decoupling convolutional neural network was proposed for intelligent compound faults diagnosis successfully.\textsuperscript{8} A novel method called multi-scale feature extraction (MFE) and MSVM with particle parameter adaptive (PPA) was proposed for intelligent multiple-fault diagnosis.\textsuperscript{9} To diagnose compound faults of locomotive roller bearings accurately, a novel hybrid intelligent diagnosis method was proposed,\textsuperscript{10} and the diagnosis results of the compound faults of the locomotive roller bearings verified that the proposed hybrid intelligent method may accurately recognize compound faults. A new method based on use of combined mode functions for selecting the intrinsic mode functions instead of the maximum cross correlation coefficient–based EEMD technique, sandwiched with, convolution neural networks, which were deep neural nets, used as fault classifiers.\textsuperscript{11} A multi-fault diagnosis method of rolling bearing elements by combing wavelet analysis with hidden Markov model (HMM) was proposed.\textsuperscript{12} An effective method for multi-fault diagnosis was presented with optimizing signal decomposition levels using wavelet analysis and SVM.\textsuperscript{13} A novel method of rolling bearing fault diagnosis based on a combination of EEMD, weighted permutation entropy, and an improved SVM ensemble classifier was proposed.\textsuperscript{14} The stationary WT and singular value decomposition were combined and the stationary wavelet singular entropy was proposed, which was used to extract fault feature of compound faults of rolling bearing. The extracted features were passed on to a kernel extreme learning machine classifier and satisfactory results could be obtained through experiment verification.\textsuperscript{15} The multi-scale wavelet entropy of rolling bearing compound faults was computed and used as training and test input of the kernel extreme learning machine classifier, and the results verified that the proposed method had slight much better diagnosis accuracy than other related methods.\textsuperscript{16}

Based on the above literature, it could be concluded that there are two key steps in intelligent diagnosis of rolling bearing compound faults: feature extraction and efficient intelligent classification algorithm. As a relative signal-processing method, sparse decomposition could capture the implicit characteristics of vibration signal when fault arises in rotating machinery, and it has been used widely in fault diagnosis of rotating machinery and other areas.\textsuperscript{2,17–20} However, most of papers relating to fault diagnosis of rolling bearing basing on sparse decomposition mainly focus on single fault of rolling bearing, and it has great application potential in feature extraction of rolling bearing compound fault. In recent years, kinds of intelligent classification algorithms such as deep belief network (DBN),\textsuperscript{21,22} generalized linear regression model,\textsuperscript{23} hidden Markov random field,\textsuperscript{24} hybrid MLPNN-ICA,\textsuperscript{25} hybrid CNN-MLP,\textsuperscript{26} and so on have been arising. However, most of these intelligent methods are used mainly in image classification. Besides, HMM is an effective and mature intelligent algorithm, which has been used widely in fault diagnosis of rotating machinery\textsuperscript{27–29} and the authors of this article also has done some work on HMM.\textsuperscript{30} So, the article proposes an intelligent diagnosis method of rolling bearing compound faults based on device state dictionary set sparse decomposition feature extraction-HMM. First, characteristic dictionaries of rolling bearing under different operating conditions are extracted by sparse decomposition self-learning method, and state dictionary set of rolling bearing is constructed. Then, the compound fault signals of bearing are transformed into sparse domain using the constructed dictionary set to extract sparse features. At last, the extracted sparse features are used as training and testing vectors of HMM, and satisfactory intelligent diagnosis results are obtained.

The organization of the article is as following: section “The sparse feature extraction method based on device state dictionary set” is dedicated to the sparse feature extraction method based on device state dictionary set. Section “HMM” and section “The flow chart of the proposed method” are dedicated to the basic theory of HMM and the flow chart of the proposed method, respectively. Experiment is carried out in
The sparse feature extraction method based on device state dictionary set

**The construction method of sparse feature**

Suppose there exists $L$ known target classes and the total training sample sets are represented by $Y = [Y_1, Y_2, \ldots, Y_L]$, and $Y_i$ is the $i$th training sample set. Let $D_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,n}]$ represent the sub-dictionary of the $i$th target class, and $D_i$ could be obtained using the following function

$$
\{D_i, X_i\} = \arg \min_{D_i, X_i} \| Y_i - D_i X_i \|_2^2 + \lambda \| X_i \|_0
$$

(1)

where $\lambda > 0$. All the sub-dictionaries are gathered into a larger redundant dictionary set $D = [D_1, D_2, \ldots, D_L]$, so that the dictionary contains all the sub-dictionaries to express the target class. $D$ is named as dictionary set in order to distinguish it from $D_i$. Assuming $y$ is a test sample signal and its sparse coefficients $X = [X_1, X_2, \ldots, X_L]$ under $D$ are calculated, and $X_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,m}]$ are the sparse coefficients corresponding to $D_i$. Furthermore, the potential components of each sub-dictionary could be obtained using the sub-dictionary $D_i$ and its corresponding coefficients $X_i$

$$
y_i = D_i X_i = \sum_{m} x_{i,m} d_{i,m} \quad (1 \leq i \leq L)
$$

(2)

Meanwhile, the test sample single $y$ could be expressed as

$$
y = \sum_{j=1}^{L} y_j
$$

(3)

The sub-dictionary $D_i$ of signal $y_i$ has strong adaptability to the test sample signal $y$. That is to say, the sub-dictionary $D_i$ could be more likely to be activated to approach or represent the test sample signal $y_i$ is assumed to represent the class label of the test sample $y_i$, and the sub-dictionary $D_i$ is more likely to be activated. That is to say, the non-zero term in its sparse coefficients of $D_i$ is most likely to appear in $X_i$. Sparse coefficients are often used directly as sparse features for classification in image or speech signal-processing area.

However, in mechanical signals, the dimension of sparse coefficients is often high. In this article, a sparse feature construction method based on energy distribution is proposed.

The test sample signal $y$ could be decomposed into the sum of a series of sub-components under the redundant dictionary set $D = [D_1, D_2, \ldots, D_L]$ using the following equation

$$
y = \sum_{j=1}^{L} l_j
$$

(4)

where $l_j = D_j X_j$. The energy of each sub-component could be defined as following

$$
E_i = \sum_{n} \left( l_i(n) \right)^2
$$

(5)

The normalized energy shown in equation (6) is used as sparse feature to prevent the occurrence of large amounts of energy

$$
\tilde{E}_i = \frac{E_i}{E} = \frac{E_i}{\sum_{m} E_{m}}
$$

(6)

The last obtained normalize sparse eigenvectors are shown in equation (7)

$$
F = [\tilde{E}_1, \tilde{E}_2, \ldots, \tilde{E}_L]
$$

(7)

In summary, the flow chart of sparse feature construction method based on dictionary learning is shown in Figure 1, and it could be divided into two main steps: the construction of redundant dictionary set $D = [D_1, D_2, \ldots, D_L]$ and sparse features $F = [\tilde{E}_1, \tilde{E}_2, \ldots, \tilde{E}_L]$. The shift invariant sparse coding (SISC) method is used to construct the redundant dictionary set $D = [D_1, D_2, \ldots, D_L]$, and the feature-sign search (FSS) method which will be discussed in section “Fast algorithm for sparse decomposition” is used to calculate the sparse coefficients of $D_i$.

**Fast algorithm for sparse decomposition**

Although there are many algorithms for solving sparse coefficients such as matching pursuit (MP), basis pursuit (BP), and so on, all of these algorithms have the problem of large amount of computation. The FSS method solves the analytic solution by guessing the sign of coefficients, which is more efficient than MP and BP. The objective optimization function of BSS
could be expressed as equation (8) when the dictionary set $D$ is known

$$\min_{S} \| y - DS \|_2^2 + \beta \sum_j |s_j|$$

(8)

This problem existing in equation (8) is a $L_1$ regular least squares problem. The $L_1$ norm problem of $|s_j|$ could be ignored if the sign of each element in coefficients $S$ is known: when $s_j < 0$, there is $|s_j| = -s_j$; when $s_j > 0$, there is $|s_j| = s_j$; when $s_j = 0$, there is $|s_j| = 0$. Then, equation (8) could be transformed into unconstrained quadratic optimization problem which could achieve effective solution. The following are main steps of FSS:

1. Initialize each element of the sparse coefficient with their corresponding sign, that is, $s_i = 0$, $\theta_i = 0$, $\theta_i \in \{-1, 0, 1\}$. Besides, the initialization set is initialized as $activeset = \{\}$.  

2. For all the coefficients with value of 0, select $i = \operatorname{argmax}_j |\partial\| y - DS \|_2^2 / \partial s_i|$.  
   a. If $|\partial\| y - DS \|_2^2 / \partial s_i| > \beta$, then $\theta_i = -1$, $activeset = \{i\} \cup activeset$.  
   b. If $|\partial\| y - DS \|_2^2 / \partial s_i| < \beta$, then $\theta_i = 1$, $activeset = \{i\} \cup activeset$.  

3. Note $D$, $S$, and $\theta$ as the sub-vector sets corresponding to $D$, $S$, and $\theta$ in $activeset$, and the analytical solution of unconstrained quadratic optimization problems is computed through $D$, $S$, and $\theta$. Besides, the discrete search is carried out in the closed line segment from $S$ to $S_{new}$, and the sign of the searched point $S'$ is compared with the sign of $S_{new}$ as follows:

a. If the signs of $S'$ and $S_{new}$ are same, then $S = S_{new}$;  

b. If the signs of $S'$ and $S_{new}$ are not same, then $S = S'$.  

The 0 components of $S$ are removed from $activeset$, and update $\theta_i = \text{sign}(S_i)$, and the optimal conditions are checked as follows:

a. For non-zero coefficients: $\partial\| y - DS \|_2^2 / \partial s_i + \beta \text{sign}(s_i) = 0$, $\forall s_i \neq 0$. If the condition a is not satisfied, then go back to step (3). Otherwise, check the condition b.  

b. For zero coefficients: $|\partial\| y - DS \|_2^2 / \partial s_i| \leq \beta$, $\forall s_i = 0$. If the condition b is not satisfied, then go back to step (2). Otherwise, end the iteration process.

**HMM**

A HMM is a finite state statistical structure with a fixed number of states, and it is generally applicable to analyze the non-stationary signals such as speech and time-varying noise. HMM is a double-embedded stochastic process with an underlying stochastic process which is not observable directly, but can be observed only through another set of stochastic process which produces the sequence of observations. HMM could be divided into discrete hidden Markov model (DHMM) and continuous hidden Markov model (CHMM) based on the property of the observations which is discrete or continuous. A DHMM can be described using the following parameters:

1. States. Let $N$ represent the number of states in the model. The states can be described as

![Figure 1. Sparse feature construction based on dictionary learning.](image)
$S = \{S_1, S_2, \ldots, S_N\}$. The individual state at time $t$ is denoted as $q_t$. Apparently, there is $q_t \in S$.

2. **Observation symbols.** Let $M$ denote the number of distinct observation symbols per state. The observation symbols are represented as $V = \{V_1, V_2, \ldots, V_M\}$. The observable value at time $t$ is denoted as $o_t$, $o_t \in V$.

3. **State transition probability distribution** $A = \{a_{ij}\}$

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \quad (1 \leq i, j \leq N) \quad (9)$$

with the following property

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i \quad (10)$$

in which $i$ and $j$ denote the state indices.

4. **Observation symbol probability distribution** $B = \{b_j(k)\}$

$$b_j(k) = P(o_t = V_k | q_t = S_j)(1 \leq j \leq N, \ 1 \leq k \leq M) \quad (11)$$

5. **Initial state distribution** $\pi = \{\pi_i\}$

$$\pi_i = P(q_1 = S_i)(1 \leq i \leq N) \quad (12)$$

In summary, a HMM $\lambda$ is defined by two model parameters, $N$ and $M$, observation symbols, and three sets of probability measures: $A$, $B$, and $\pi$. The model $\lambda$ is expressed as

$$\lambda = (\pi, A, B) \quad (13)$$

In practical applications, the observations encountered are continuous usually. Although the continuous signal can be encoded into discrete points, amounts of valuable information may be lost in this encoding process. In this case, a CHMM is advantage over a DHMM. In a CHMM, the Gaussian mixture model is used usually to fit the probability distribution of the observations

$$b_j(O) = \sum_{m=1}^{M} c_{jm}N(O, \mu_{jm}, \Sigma_{jm})(1 \leq j \leq N) \quad (14)$$

In equation (14), $M$ is the number of Gaussian elements, $c_{jm}$ is the mixture coefficient of $m$th Gaussian element in $j$th state. $\mu_{jm}$ and $\Sigma_{jm}$ are the mean vector and covariance matrix of the mixture coefficient of $m$th Gaussian element in $j$th state. In summary, a CHMM can be described as being shown in equation (15) whose observations probability is mixture Gaussian distribution

$$\lambda = (\pi, A, C, \mu, \Sigma) \quad (15)$$

The parameters can be estimated using the expectation maximum (EM$^3$) algorithm.

### The flow chart of the proposed method

The flow chart of the proposed method is shown in Figure 2 which contains main four basic steps as follows:

1. **State dictionary set construction.** Apply dictionary learning method on the training vibration data of the different running states of rolling bearing, and the sub-dictionary $D_k$ corresponding to the different running states are obtained. Then, fuse each sub-dictionary $D_k$ and the device state dictionary set $D = \{D_1, D_2, \ldots, D_c\}$ is obtained.

2. **Sparse feature extraction.** Extract the sparse feature based on $D = \{D_1, D_2, \ldots, D_c\}$ and each group of signals is decomposed into the sum of a series of sub-components, that is, $L = \{L_1, L_2, \ldots, L_c\}$. Besides, the obtained sparse feature is normalized and last obtained sparse feature is marked as $F = \{F_1, F_2, \ldots, F_c\}$.

3. **The training of HMM model.** A HMM model is trained for each state of data and the trained HMM models are saved in the model base, that is, $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_c\}$.

4. **On-line diagnosis.** On-line diagnosis of the test samples using the trained HMM models in the model base. First, extract the feature of the test samples using the state dictionary set and the feature sequence $O = \{o_1, o_2, \ldots, o_T\}$ is generated. Then, calculate the probability of test samples under different models, that is, $P(O|\lambda_i)$, and the state of the test samples is decided by the biggest $P(O|\lambda_i)$.

### Experiment

The rolling element bearing compound fault experiment is carried out in the section to verify the effectiveness of the proposed method. Figure 3 is the test rig, and NU205 is the used rolling bearing type in the experiment. The corresponding parameters of NU205 are presented in Table 1. Four running states of test bearings are implemented, respectively: normal (N),
outer race and ball compound fault (OB), outer and inner race compound fault (OI), and outer race and inner race and ball compound fault (OIB). Figure 4 (a)–(c) is the processed faults on inner race, out race, and rolling element of the test bearing, respectively, and the three kinds of compound faults are realized by their different combinations. The right-end bearing supporting the right end of the shaft is detachable for the convenient replacement of the bearing in the test processes. During the experiment, the outer ring is fixed and the inner ring rotates synchronously with the shaft. The acceleration sensor is installed near the test bearing and is used to collect the peak value of corresponding vibration signal. Set the sampling frequency as $f_s = 8192$ Hz and the rotating frequency of the shaft is $f_r = 13.3$ Hz. Equations (16)–(18) are used to calculate the characteristic frequencies of inner race fault, outer race fault, and rolling element

$$f_i = \frac{Z}{2} \left( 1 + \frac{d}{D} \cos \beta \right) f_r$$  \hspace{1cm} (16)$$

$$f_o = \frac{Z}{2} \left( 1 - \frac{d}{D} \cos \beta \right) f_r$$  \hspace{1cm} (17)$$

$$f_b = \frac{D}{d} \left[ 1 - \left( \frac{d \cos \beta}{D} \right)^2 \right] f_r$$  \hspace{1cm} (18)$$

In equations (16)–(18), $Z$ is the number of rolling elements, $d$ is rolling element diameter, $D$ is the pitch diameter, and $\beta$ is the contact angle. The values of $f_i, f_o$, and $f_b$ are 95.38, 64.61, and 5.38 Hz, respectively, through calculation.

The time-domain waveforms of the test bearings’ four states with their corresponding envelope

Table 1. Parameters of the test rolling element bearing.

| Type   | Ball number | Ball diameter (mm) | Pitch diameter (mm) | Contact angle | Motor speed (r/min) |
|--------|-------------|--------------------|---------------------|---------------|---------------------|
| NU205  | 12          | 7.5                | 39                  | 0             | 800                 |

Figure 2. Flow chart of the proposed method.

Figure 3. Test rig.
Figure 4. Processed faults on the each parts of the test bearing: (a) inner race fault, (b) rolling element fault, and (c) outer race fault.

Figure 5. Time-domain waveforms of test bearings’ fours states with their corresponding envelope demodulation spectral: (a) time-domain waveform of N, (b) envelope demodulation spectral of the signal shown in (a), (c) time-domain waveform of OB, (d) envelope demodulation spectral of the signal shown in (c), (e) time-domain waveform of OI, (f) envelope demodulation spectral of the signal shown in (e), (g) time-domain waveform OIB, and (h) envelope demodulation spectral of the signal shown in (g).
Figure 6. Learned basis atoms for each condition: (a1)–(a4) for Normal; (b1)–(b4) for OB; (c1)–(c4) for OI; and (d1)–(d4) for OIB.

Figure 7. Energy distribution of each class of bearing data on each sub-dictionary: (a) N, (b) OB, (c) OI, and (d) OIB.
demodulation spectral are shown in Figure 5. Figure 5 (a) and (b) is the time-domain waveform with it envelope demodulation spectral of N. It is evident that very little amount of impulsion signal arises in the time-domain waveform of N. In Figure 5(d), though the fault characteristic frequency (FCF) of outer race is extracted, the FCF of ball is not extracted. In Figure 5(f) and (h), the spectral lines are chaotic from which the compound fault features of OI and OIB could not identified clearly.

Ten groups of samples of each state are selected randomly as training samples and are used to learn the corresponding sub-dictionary $D_i$. The original signal of bearing’s four states are analyzed by SISC, and the parameters are selected as follows: the atomic length is 256, the overlap rate is 0.25, the sparsity is 1, and the number of base atoms is 4. Then, the sub-dictionary of the four states are obtained, and use $D_1$ to represent the sub-dictionary of N, $D_2$ to represent the sub-dictionary of OB, $D_3$ to represent the sub-dictionary of OI, and $D_4$ to represent the sub-dictionary of OIB. The device state dictionary set $D = [D_1, D_2, D_3, D_4]$ is obtained by fusing the four of them. The learned basis for each state using SISC is presented in Figure 6.

The sub-dictionary has much better adaptability to the samples in the corresponding state since each of

---

Figure 8. Diagnosis results using different feature set: (a) SFV, (b) TDSV features, and (c) WPE features.
them is obtained from the training samples in the corresponding state. In other words, the sub-dictionaries are more easily activated when the test samples are expressed by the state dictionary set. The test sample could be decomposed into a series of sub-components \( \{l_1, l_2, l_3, l_4\} \) using the sparse feature construction method as introduced in the previous section after obtaining the dictionary set. Besides, the normalized sparse feature \( \{E_1, E_2, E_3, E_4\} \) is also obtained, and values of sparse features represent the energy distribution of samples in each sub-dictionary. Figure 7 is the energy distribution of each class of bearing data on each sub-dictionary.

Each group of training samples are divided into 10 segments, and the length of each segment was 0.08 s. Each segment of the signal is sparsely decomposed, and sparse features are extracted in the dictionary. A 4*10 feature vector sequence can be obtained from each training sample group. HMM is trained using the feature sequences of different states, and four HMM models \( \{\lambda_{NC}, \lambda_{ORB}, \lambda_{IRF}, \lambda_{REF}\} \) corresponding to the bearing’ four states are obtained. Sparse features are extracted from each group of test samples and are input into each trained HMM models. Likelihood probability is calculated by Viterbi algorithm in turn, and the state of the test samples is determined by the maximum value of likelihood probability. There are total of 80 sets of test samples: No. 1–20 come from N state, No. 21–40 come from OB state, No. 41–60 come from OI state, and No. 61–80 come from OIB state. Figure 8 shows the diagnosis results using different feature extraction vector: (a) sparse feature vectors (SFVs); (b) time-domain statistical feature vectors (TDSVs) such as AMP, P-P, and so on; and (c) wavelet packet energy (WPE). Misclassified samples are marked with black circles: there are total two N samples and one OIB sample misclassified as shown in Figure 8(a). There are total 9 samples misclassified as shown in Figure 8(b), and there are total 17 samples misclassified in Figure 8(c). It is evident that the SFV has better classification result than the other two feature extraction vectors.

The K-nearest-neighbor (KNN) and BP neural network algorithm are used to analyze the above three kinds of different features to compare their analysis results, and the diagnosis results are given in Table 2. BP neural network adopts three-layer network structure, in which the number of nodes in the input layer is consistent with the characteristic dimension, while the numbers of nodes in the hidden layer and the output layer are 10 and 4, respectively. The diagnostic rates of KNN, BP, and HMM were 90%, 92.5%, and 96.25%, respectively, when the same sparse feature is used as input, and the diagnostic accuracy of HMM is the highest. It can also find that HMM has the best diagnostic effect when other eigenvectors are used. By comparing different features under the same classifier, it is found that the diagnostic rate corresponding to SFV is significantly higher than that of TDSV feature and WPE feature.

### Conclusion

Identification of rolling bearing fault patterns, especially for the compound faults, has attracted notable attention and is still a challenge in fault diagnosis. As a relative signal-processing method, sparse decomposition could capture the implicit characteristics of vibration signal when fault arises in rotating machinery, and it has great application potential in composite fault diagnosis of rolling bearing. In this article, a method of constructing sparse features based on dictionary learning is proposed. This method learns each state of the device to get the feature dictionary, then fuses all the feature dictionaries to form the state dictionary set, and then the sparse features of the different compound faults of rolling bearing are learned based on the obtained state dictionary set. Furthermore, the learned sparse features are used as training and test input of HMM, and satisfactory classification results are obtained at last. The concrete flow chart of the proposed method is given, and the validity of the proposed method is verified by compound fault experiment of rolling bearings. Besides, in order to highlight the superiority of sparse feature, the classification results of sparse feature are compared with those of time-domain feature and energy feature of wavelet packet. The comparison results show that sparse feature has high classification accuracy and stability.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this

---

**Table 2. Comparisons with other feature set and classifiers.**

| Feature class | KNN (%) | BP (%) | HMM (%) |
|---------------|---------|--------|---------|
| SFV           | 90      | 92.5   | 96.25   |
| TDSV          | 77.5    | 83.75  | 88.75   |
| WPE           | 80      | 77.5   | 78.75   |

KNN: K-nearest-neighbor; HMM: hidden Markov model; SFV: sparse feature vector; TDSV: time-domain statistical feature vector.
References

1. Qiu H, Lee J, Lin J, et al. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *J Sound Vib* 2006; 289: 1066–1090.

2. Selesnick IW. Wavelet transform with tunable Q-Factor. *IEEE T Signal Proces* 2011; 59: 3560–3575.

3. Wu ZH and Huang NE. Ensemble empirical mode decomposition: a noise assisted data analysis method. *Adv Adapt Data Anal* 2009; 1: 1–41.

4. Antoni J. The spectral kurtosis: a useful tool for characterizing non-stationary signals. *Mech Syst Signal Pr* 2006; 20: 282–307.

5. Antoni J. Fast computation of the kurtogram for the detection of transient faults. *Mech Syst Signal Pr* 2007; 21: 108–124.

6. Manjurul MM and Kim JM. Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vectors machines. *Reliab Eng Syst Safe* 2019; 184: 55–66.

7. Hao YS, Song LY, Ren BY, et al. Step-by-step compound faults diagnosis method for equipment based on majorization-minimization and constrained SCA. *IEEE-ASME T Mech* 2019; 24: 2477–2487.

8. Huang RY, Liao YX, Zhang SH, et al. Deep decoupling convolutional neural network for intelligent compound fault diagnosis. *IEEE Access* 2019; 7: 1848–1858.

9. Wang HC, Chen J and Dong GM. Feature extraction of rolling bearing’s early weak fault based on EEMD and tunable Q-factor wavelet transform. *Shock Vib* 2017; 19: 462–483.

10. Lei YG, He ZJ and Zi YY. Application of a novel intelligent method to compound fault diagnosis of locomotive roller bearings. *J Vib Acoust* 2008; 130: 034501.

11. Singh SK, Kumar S and Dwivedi JP. Compound fault prediction of rolling bearing using multimedia data. *Multimed Tools Appl* 2017; 76: 18771–18788.

12. Purushotham V, Narayanan S and Prasad SAN. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT&E Int* 2005; 38: 654–664.

13. Abbasian S, Raftsanjani A, Farshidianfar A, et al. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mech Syst Signal Pr* 2007; 21: 2933–2945.

14. Zhou SH, Qian SL, Chang WB, et al. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. *Sensors* 2018; 18: 1934.

15. Rodriguez N, Cabrera G, Lagos C, et al. Stationary wavelet singular entropy and kernel extreme learning for bearing multi-fault diagnosis. *Entropy* 2017; 19: 541.

16. Rodriguez N, Pablo A, Lida B, et al. Combining multiscale wavelet entropy and kernelized classification for bearing multi-fault diagnosis. *Entropy* 2019; 21: 152.

17. Wang HC, Chen J and Dong GM. Feature extraction of rolling bearing’s early weak fault based on EEMD and tunable Q-factor wavelet transform. *Mech Syst Signal Pr* 2014; 48: 103–119.

18. Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms. *Adv Neural Inf Process Syst* 2007; 19: 801–808.

19. Liu LN, Liu CL and Huang YX. Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mech Syst Signal Pr* 2011; 25: 558–574.

20. Rahmani M and Akbarizadeh G. Unsupervised feature learning based on sparse coding and spectral clustering for segmentation of synthetic aperture radar images. *IET Comput Vis* 2014; 9: 629–638.

21. Samadi F, Akbarizadeh G and Kaabi H. Change detection in SAR image using deep belief network: a new training approach based on morphological images. *IET Image Process* 2019; 13: 2255–2264.

22. Zalpour M, Akbarizadeh G and Alaei-Shen N. A new approach for oil tank detection using deep learning features with control false alarm rate in high-resolution satellite imagery. *Int J Remote Sens* 2020; 41: 2239–2262.

23. Moghadam AE, Akbarizadeh G and Kaabi H. Automatic detection and segmentation of blood vessels and pulmonary nodules based on a line tracking method and generalized linear regression model. *Signal Image Video P* 2019; 13: 457–464.

24. Tirandaz Z, Akbarizadeh G and Kaabi H. PolSAR image segmentation based on feature extraction and data compression using Weighted Neighborhood Filter Bank and Hidden Markov random field-expectation maximization. *Measurement* 2020; 153: 107432.

25. Ahmadi N and Akbarizadeh G. Iris tissue recognition based on GLDM feature extraction and hybrid MLPNN-ICA classifier. *Neural Comput Appl* 2020; 32: 2267–2281.

26. Sharifzadeh F, Akbarizadeh G and Kavian YS. Ship classification in SAR images using a new hybrid CNN–MLP classifier. *J Indian Soc Remote* 2019; 47: 551–562.

27. Kourodi A, Hajii M, Harkat MF, et al. Hidden Markov model based principle component analysis for intelligent fault diagnosis of wind energy converter systems. *Renew Energ* 2020; 150: 598–606.

28. Xin G, Hamzaou N and Antoni J. Semi-automated diagnosis of bearing faults based on a hidden Markov model of the vibration signals. *Measurement* 2018; 127: 141–166.

29. Sadhu A, Prakash G and Narasimhan S. A hybrid hidden Markov model towards fault detection of rotating components. *J Vib Control* 2017; 23: 3175–3195.
30. Wang HC. Fault diagnosis of rolling element bearing based on wavelet Kernel principle component analysis-coupled hidden Markov. *J Vibroeng* 2017; 19: 5992–6006.
31. Wang HC, Li LW, Gong XY, et al. Blind source separation of rolling element bearing’ single channel compound fault based on shift invariant sparse coding. *J Vibroeng* 2017; 19: 1809–1822.
32. Elad M and Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE T Image Process* 2006; 15: 3736–3745.
33. Hong XB, Zhou JX, He YK. Damage detection of anchored region on the messenger cable based on matching pursuit algorithm. *Mech Syst Signal Pr* 2019; 130: 221–247.
34. Zheng HB, XZ Chen and Li ZY. The application of basis de-noising of gear fault diagnosis. *J Vib Meas Diagn* 2003; 23: 128–130.
35. Rabiner LR. A tutorial on hidden Markov models and selected application in speech recognition. *Proc IEEE* 1989; 77: 257–286.