IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 12, 2021, accepted January 28, 2021, date of publication February 2, 2021, date of current version February 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056595

Bearing Fault Diagnosis Method Based on Ensemble Composite Multi-Scale Dispersion Entropy and Density Peaks Clustering

AI-SONG QIN[1], HAN-LING MAO[1], QIN HU[2,3], AND QING-HUA ZHANG[2]

[1]School of Mechanical Engineering, Guangxi University, Nanning 530004, China
[2]Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China
[3]Department of Automation, Rocket Force University of Engineering, Xi'an 710025, China

Corresponding author: Han-Ling Mao (maohl79@gxu.edu.cn)

This work was supported in part by the Science and Technology Base and Talents Special Project of Guangxi Province under Grant AD19259002; in part by the National Key Research and Development Program of China under Grant 2018YFC0808600; in part by the National Nature Science Foundation of China under Grant 61933013, Grant 62073091, and Grant 62073090; in part by the Program of Guangdong Science and Technology Innovation Strategy under Grant 2018KJ011; and in part by the Young Innovative Talents Program of Guangdong University under Grant 2019KQNCX088.

ABSTRACT For bearing fault diagnosis, how to effectively extract informative fault information and accurately diagnose faults is still a key problem. To this end, this study proposes a novel bearing fault diagnosis approach based on ensemble composite multi-scale dispersion entropy (ECMDE), local preserving projections and density peaks clustering. Specifically, ECMDEs are developed to capture multi-scale fault features from the raw vibration signals. The goal of ECMDEs is to synthesize different kinds of composite multi-scale dispersion entropies to find more effective fault information. Subsequently, the local preserving projections method is utilized to reduce high-dimensional feature set and extract the effective fault information. Finally, the reduced features are fed into the density peaks clustering method to obtain the fault diagnosis results. Two experimental cases and extensive comparisons are applied to validate the effectiveness and noise robustness of the proposed method. Experimental results demonstrate that the proposed method is capable to reliably extract effective fault information of raw vibration signals and accurately diagnose bearing faults even under low signal-to-noise ratio conditions.

INDEX TERMS Ensemble composite multi-scale dispersion entropy, local preserving projections, density peaks clustering, noise robustness, fault diagnosis.

ACRONYMS AND ABBREVIATIONS

APC	Affinity Propagation Clustering
CMDE	Composite Multi-scale Dispersion Entropy
CMPE	Composite Multi-scale Permutation Entropy
CMSE	Composite Multi-scale Sample Entropy
CWRU	Case Western Reserve University
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DE	Dispersion Entropy
DPC	Density Peaks Clustering
ECMDE	Ensemble Composite Multi-scale Dispersion Entropy
GPKLFD	Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis
LPP	Local Preserving Projections
MDE	Multi-scale Dispersion Entropy
MPE	Multi-scale Permutation Entropy
MSE	Multi-scale Sample Entropy
mRMR	Minimum Redundancy Maximum Relevance
PCA	Principal Component Analysis
RCMDE	Refined Composite Multi-scale Dispersion Entropy
RF	Reduced Feature
SNR	Signal-to-Noise Ratio
t-SNE	t-Distributed Stochastic Neighbor Embedding

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenbao Liu.

CMDE*smean*	CMDEs based on mean value
CMDE*srms*	CMDEs based on root mean square value
CMDE*svar*	CMDEs based on variance value
CMDE*sra*	CMDEs based on root amplitude value
CMDE*smax*	CMDEs based on maximum value

I. INTRODUCTION

Rolling bearings are widely used in rotating machinery and their working condition directly affects the safety and stability of mechanical operation. A rolling bearing fault can lead to unexpected downtime and economic loss. Consequently, it is significant to accurately detect and diagnose the health states of the bearings [1]. It is well known that fault feature extraction is the basis of fault diagnosis. Up to now, time-domain features, frequency-domain features and different types of entropies are the commonly used fault features. Among these fault features, several types of entropies such as sample entropy [2], [3], permutation entropy [4]–[6], fuzzy entropy [7]–[10] and so on [11]–[15], are effective in quantifying the complexity and uncertainty of the time series, and thus they are widely used in mechanical fault diagnosis. Compared to the root mean square, sample entropy can give a clear distinction for cosine phase angle data between healthy and faulty conditions irrespective of operational conditions [16]. However, it is not fast enough for long signals. Permutation entropy has a fast computing speed. However, it discards some information of amplitudes while mainly concerning the order of amplitude values. With regard to fuzzy entropy, it usually produces uncertain and unsatisfactory analysis for short-term data [17]. Recently, a new type of entropy called dispersion entropy (DE) is developed by Rostaghi and Azami [18]. It has been proved to be useful for machinery fault diagnosis [19], [20] and can overcome above-mentioned shortcomings [21]. In [21], a hybrid fault diagnosis method based on improved multi-scale dispersion entropy, mRMR feature selection and extreme learning machine classifier was used to conduct fault diagnosis of rotating machinery. Essentially, the improved multi-scale dispersion entropy is composite multi-scale dispersion entropy (CMDE). In [22], [23], refined composite multi-scale dispersion entropy (RCMDE) was respectively used to analyze the bearing vibration signals and biomedical signals.

However, considering that the CMDE or RCMDE is carried out only on the basis of mean value, it may be insufficient for characterizing the complexity of time series. To solve such problem, in this study a new type of improvement to CMDE, ensemble composite multi-scale dispersion entropy (ECMDE), is proposed to find useful fault information by substituting the mean value of equal-sized time series with several other features, such as the root mean square value, variance value, root amplitude value and maximum value. As such, the multiple ECMDEs can reveal more useful information hidden in the time series. Although the ECMDEs extraction can obtain more informative information for bearing fault diagnosis, it inevitably brings redundant and irrelevant information. Thus, dimensionality reduction method needs to reduce the feature dimension and obtain an appropriate subset of features. Currently, several dimensionality reduction techniques have been applied in the field of mechanical fault diagnosis [24]–[28]. Among these dimensionality reduction methods, local preserving projections (LPP) as manifold learning method can map the original high-dimensional feature set to low-dimensional feature set without any prior knowledge [29]. In the sense of fault diagnosis, the LPP has good performance to discover discriminative low-dimensional fault information. Existing studies have indicated that LPP is beneficial to dimensionality reduction in machinery fault diagnosis [30]–[32]. In this study, LPP is used to generate the discriminative features. Finally, the discriminative features are fed into a density peaks clustering (DPC) algorithm to accomplish fault diagnosis. DPC algorithm is an unsupervised learning approach that can find underlying substructures in a group of unlabeled data points. For traditional clustering methods (e.g. K-means clustering [33], K-medoids clustering [34] and fuzzy c-means clustering [35]), the number of clusters should be determined in advance and the clustering result is largely dependent on the choice of the initial cluster centres. In practical engineering, the number of clusters cannot be known in prior. However, DPC is unsupervised clustering learning method without the need of prior knowledge concerning the number of clusters. It can be able to automatically find the correct number of clusters [36], which is more suitable for fault diagnosis in practical engineering. To our best knowledge, the studies on DPC used for fault diagnosis are very limited. A few exceptions are the works in [37], [38]. In [37], an adaptive DPC algorithm had been used to adaptively choose the cutoff distance and identify the corresponding categories of different bearing and gear faults. In [38], DPC with geodesic distances was used to explore unsupervised feature selection in the fault diagnosis field. In our study, DPC is used to partition a sample dataset into many subgroups such that samples within a group have a high similarity and samples belonging to different groups have a low similarity. To date, very scarce literature has investigated the LPP-based DPC to perform the fault diagnosis.

Based on the above analysis, this study proposes a novel hybrid bearing fault diagnosis strategy by integrating ECMDEs, LPP and DPC. The validity of the proposed method is assessed through two experimental cases. The experiment results illustrate that the ECMDEs can achieve better diagnosis result than the original CMDEs, RCMDEs and other existing multi-scale analysis methods. Besides, considering the measured bearing vibration signals are usually contaminated by the background noise, noise is an inevitable problem in practical applications. To evaluate the robustness against noise of the proposed method, additional Gaussian noises with different signal-to-noise ratios (SNRs) are added to raw vibration signals to analyze the diagnosis performance. Furthermore, in order to further illustrate the effectiveness of the proposed method,

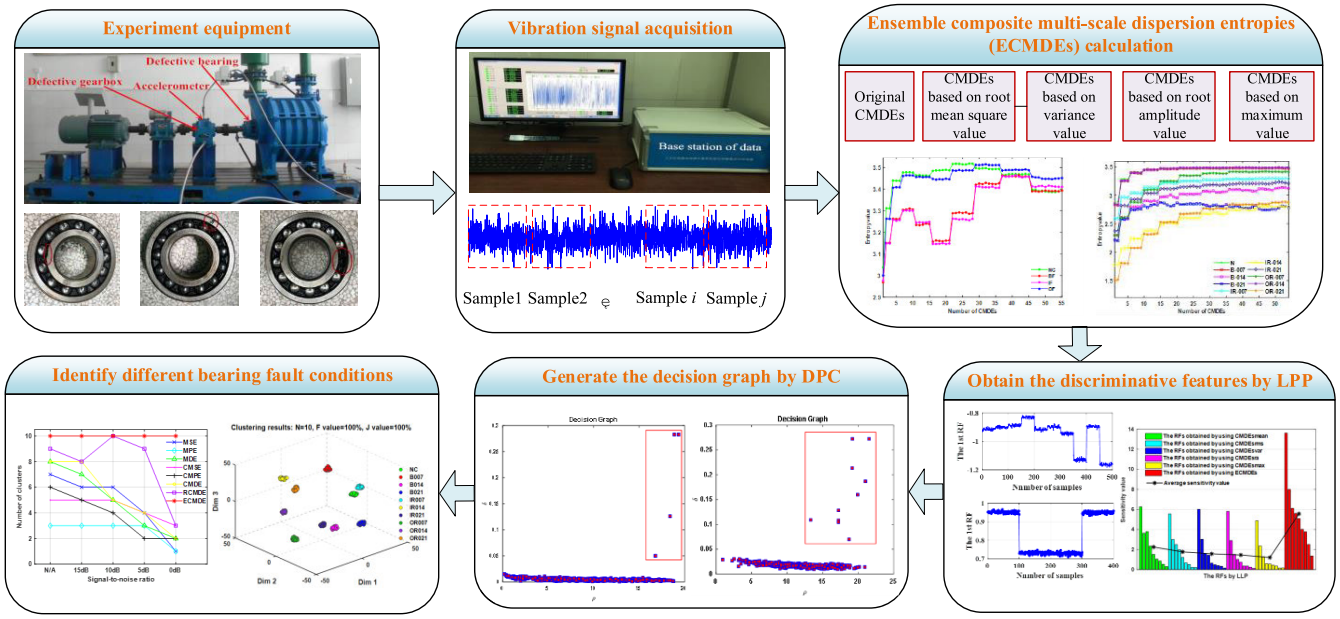


FIGURE 1. Flow chart of the proposed method.

a comparison with other unsupervised clustering methods such as the density-based spatial clustering of applications with noise (DBSCAN) and affinity propagation clustering (APC) method is performed. It is found that the DPC method has superior diagnosis performance than other unsupervised clustering methods under different SNR conditions. Through extensive comparison with state-of-the-art methods and comprehensive noise robustness analysis, it is demonstrated that the proposed method has a better diagnosis performance as well as a stronger robustness performance to noise for bearing fault diagnosis.

The main contributions of this article are summarized as follows. ECMDEs as new multi-scale fault features are proposed to perform fault diagnosis. Extensive comparison results illustrate that the proposed ECMDEs can extract more effective fault information of raw vibration signals than multiple existing multi-scale fault features. Besides, two experimental cases, including bearing datasets from Case Western Reserve University (CWRU) and Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis (GPKLFD), are applied to fully evaluate the effectiveness and noise robustness of the proposed method. Compared with multiple multi-scale analysis methods and several existing clustering methods, the experimental results demonstrated that the proposed method has better diagnosis performance and stronger noise robustness even under low SNR conditions.

The rest of this article is organized as follows: In Section II, novel hybrid bearing fault diagnosis method is introduced. In Section III, some basic definitions and theories are presented. In Section IV, two experimental studies and data analysis are performed, and extensive comparative studies are conducted for demonstration. Finally, we draw conclusions in Section V.

II. NOVEL HYBRID BEARING FAULT DIAGNOSIS METHOD

In order to efficiently extract useful fault information and reliably achieve more accurate diagnosis for bearing faults, a novel hybrid bearing fault diagnosis approach based on ECMDEs, LPP and DPC is proposed. The effectiveness and noise robustness of the proposed method is validated by using two experimental cases. The flow chart of the proposed method is illustrated in Figure 1. As shown in Figure 1, the proposed method mainly includes three parts: 1) Part 1: ECMDEs extraction for raw vibration signals under noise-free and different SNR conditions; 2) Part 2: Discriminative features generation using LPP; and 3) Part 3: DPC method for identifying different bearing faults. The implementation process of each part is elaborated as follows.

1) Part 1: ECMDEs extraction for raw vibration signals under noise-free and different SNR conditions

Raw vibration signals under different bearing fault conditions are respectively measured in a motor from the CWRU and a centrifugal multi-level impeller blower from the GPKLFD. Based on the collected original signals, a new type of improvement to the CMDE, namely ECMDE, is used to find more effective fault information. That is, multiple ECMDEs are firstly extracted from the bearing vibration signals. In addition, in order to evaluate the noise robustness of the proposed method, for each bearing dataset, additional Gaussian noise with SNR of 0, 5, 10 and 15 dB is respectively added on original vibration signals. For the vibration signals under each SNR condition, several ECMDEs are extracted to perform fault diagnosis as well.

2) Part 2: Discriminative features generation using LPP

The LPP is applied to generate the discriminative features, which can find more sensitive low-dimensional information hidden in the high-dimensional feature structure. Specifically, adjacency graph is firstly constructed for the original ECMDEs and weights for the graph edges are then calculated. Finally, eigenmaps calculation is conducted. The final achieved features can enhance the discrimination between all classes by improving the between-class scatter and within-class scatter for fault identification.

3) Part 3: DPC method for identifying different bearing faults

Based on the discriminative features generated by LPP, DPC is used to recognize different bearing fault conditions. The discriminative features are used as the inputs of the DPC method. The DPC is finally used to automatically identify bearing fault categories. Several clustering evaluation indexes are used to measure the diagnosis result. Besides, in order to evaluate the noise robustness of the proposed method, additional Gaussian noise signals with different SNRs are added to original signals to test the diagnosis performance. Moreover, for comparison purposes, multiple multi-scale analysis methods and several existing clustering methods have been performed bearing fault diagnosis as well. Extensive comparison results demonstrate that the proposed method has a better diagnosis performance as well as a stronger noise robustness performance for bearing fault diagnosis.

III. RELATED THEORIES

A. ENSEMBLE COMPOSITE MULTI-SCALE DISPERSION ENTROPY

The DE method [14] is a nonlinear dynamic method to characterize the complexity and irregularity of time series. Suppose $x = \{x_i, i = 1, 2, \ldots, N\}$ represents a time series with the length of N. The DE can be described as follows:

Firstly, $x = \{x_i, i = 1, 2, \ldots, N\}$ is mapped into $y = \{y_i, i = 1, 2, \ldots, N\}$, $y_i \in (0, 1)$ by using a normal cumulative distribution function, that is

$$y_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} \exp\left(\frac{-(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t \tag{1}$$

where μ and σ represent expected value and standard deviation, respectively. Through using a linear transform method, each y_i is mapped into an integer from 1 to c. Each member of the mapped signal can be defined as $z_i^c = round\,(cy_i + 0.5)$, where z_i^c denotes the ith member of the classified series and c is the number of classes.

Then, embedding vector $z_i^{m,c}$ is introduced and defined as $z_i^{m,c} = \left\{z_i^c, z_{i+d}^c, \ldots, z_{i+(m-1)d}^c\right\}$, $i = 1, 2, \ldots, N - (M-1)\,d$, where m denotes the embedding dimension and d is the time delay. Suppose $z_i^c = v_0, z_{i+d}^c = v_1, \ldots, z_{i+(m-1)d}^c = v_{m-1}$, the dispersion pattern is defined as $\pi_{v_0 v_1 \ldots v_{m-1}}$. The relative frequency of each potential dispersion patterns can be calculated by:

$$P\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right) = \frac{Num\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right)}{N - (m-1)\,d} \tag{2}$$

where $Num\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right)$ represents the total number of $z_i^{m,c}$ mapped to $\pi_{v_0 v_1 \ldots v_{m-1}}$.

Finally, according to the definition of Shannon entropy, the DE of the time series $x = \{x_i, i = 1, 2, \ldots, N\}$ is defined as

$$DE\,(x, m, c, d) = -\sum_{\pi=1}^{c^m} p\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right) \cdot \ln\left(p\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right)\right) \tag{3}$$

Based on the calculation of the DE, the ECMDE is proposed in this study. The specific process of calculating the ECMDE is as follows. For a given time series $x = \{x_1, x_2, \ldots, x_N\}$ with the length of N, the coarse-grained series at scale factor τ can be respectively calculated by using equations (4-8) to obtain multiple coarse-grained series, where $1 \le k \le \tau$, $1 \le j \le \langle N/\tau \rangle$, $\langle\cdot\rangle$ denotes the rounding operation.

Coarse-grained series based on mean value:

$$z_{k,j}^{(\tau)}|_{\mathrm{mean}} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i \tag{4}$$

Coarse-grained series based on root mean square value:

$$z_{k,j}^{(\tau)}|_{rms} = \sqrt{\frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i^2} \tag{5}$$

Coarse-grained series based on variance value:

$$z_{k,j}^{(\tau)}|_{var} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} (x_i - \bar{\mathrm{x}})^2 \tag{6}$$

Coarse-grained series based on root amplitude value:

$$z_{k,j}^{(\tau)}|_{ra} = \left(\frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} \sqrt{|x_i|}\right)^2 \tag{7}$$

Coarse-grained series based on maximum value:

$$z_{k,j}^{(\tau)}|_{\max} = \max_{(j-1)\tau+k \le i \le j\tau+k-1} (x_i) \tag{8}$$

Based on these coarse-grained series, the DE of each coarse-grained series is calculated as follows.

$$\begin{aligned}
\mathrm{CMDE}\,(x, m, c, d, \tau)|_{\mathrm{mean}} &= \mathrm{DE}\left(z_{k,j}^{(\tau)}|_{\mathrm{mean}}, m, c, d\right) \\
\mathrm{CMDE}\,(x, m, c, d, \tau)|_{\mathrm{rms}} &= \mathrm{DE}\left(z_{k,j}^{(\tau)}|_{\mathrm{rms}}, m, c, d\right) \\
\mathrm{CMDE}\,(x, m, c, d, \tau)|_{\mathrm{var}} &= \mathrm{DE}\left(z_{k,j}^{(\tau)}|_{var}, m, c, d\right) \\
\mathrm{CMDE}\,(x, m, c, d, \tau)|_{\mathrm{va}} &= \mathrm{DE}\left(z_{k,j}^{(\tau)}|_{\mathrm{ra}}, m, c, d\right) \\
\mathrm{CMDE}\,(x, m, c, d, \tau)|_{\max} &= \mathrm{DE}\left(z_{k,j}^{(\tau)}|_{\max}, m, c, d\right)
\end{aligned} \tag{9}$$

Through synthesizing the above different kinds of CMDEs, the proposed ECMDEs are finally obtained. In this study, the parameters m, c, d and τ are set as 2, 6, 1 and 10, respectively, according to the reference [14]. In addition, it is noted that the length of processed data x should be greater than 2000. In this study, the data length of each sample is 2400.

B. LOCAL PRESERVING PROJECTION

Manifold learning method LPP aims to preserve the intrinsic geometry and local structure of a dataset. Given a dataset $\{\boldsymbol{x}_i\}_{i=1}^{N}, \boldsymbol{x}_i \in R^d$, the objective of LPP is to seek a mapping matrix $\mathbf{A}$ to acquire a low-dimensional output data set $\{\boldsymbol{z}_i\}_{i=1}^{N}, \boldsymbol{z}_i \in R^l (l < d)$, where $\boldsymbol{z}_i = \mathbf{A}^{\mathrm{T}}\boldsymbol{x}_i \in R^l$ is the new representation of $\boldsymbol{x}_i$. The objective function for selecting such a map can be defined as [24]:

$$\min \sum_{ij} \| \boldsymbol{z}_i - \boldsymbol{z}_j \|^2 W_{ij} \tag{10}$$

where $\mathbf{W}$ is a weighted matrix with W_{ij} possessing the weight of the edge joining $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Considering that the LPP algorithm is lack of supervision information, the weighted matrix is constructed by adding sample label information in this study. if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ both belong to the cth class, then $W_{ij} = \frac{1}{n_c}$, where n_c is the number of samples in the cth class. Otherwise, $W_{ij} = 0$.

The final optimization problem of LPP is

$$\arg\min_{A} \; tr\left(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\mathbf{A}\right), \quad \text{s.t. } \mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}}\mathbf{A} = \mathbf{I} \tag{11}$$

where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]_{d\times N}$ is data matrix in the original feature space, $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$, and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix.

Using the Lagrange multiplier method, the mapping matrix is given by minimum eigenvalue solution:

$$\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\mathbf{A} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}}\mathbf{A} \tag{12}$$

where λ is the eigenvalue and $\mathbf{A}$ represents the corresponding eigenvector.

The first l eigenvectors from $\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_l$ are associated with the l smallest eigenvalue. Finally, the dimensionality reduction is as follows:

$$\boldsymbol{Z} = \mathbf{A}^{\mathrm{T}}\boldsymbol{X}, \quad \mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_l] \tag{13}$$

The reduced vector $\boldsymbol{Z}$ preserves the local variation information regarding the relationship with its neighbors and will be served as the new features. Therefore, LPP seeks a mapping matrix $\mathbf{A}$ to project high-dimensional input data set $\mathbf{X}$ into a low-dimensional data set.

C. DIAGNOSIS PERFORMANCE ASSESSMENT USING DENSITY PEAK CLUSTERING ALGORITHM

The DPC algorithm as an unsupervised clustering method is chosen to analyze and assess the diagnosis performance based on the discriminative features obtained by using manifold learning method LPP. It can identify the clustering centers in an intuitive way and automatically detect the best cluster number. Let the data set $x = \{x_i\}_{i=1}^{N}, x_i \in R^m$ that is to be clustered, where N denotes the number of samples. Based on the distance $d_{ij} = \|x_i - x_j\|_2$ between sample x_i and x_j, two important parameters for each sample are calculated: local density ρ_i and separation distance δ_i.

Local density ρ_i of each sample is defined as follows.

$$\rho_i = \sum_{j=1, j\neq i}^{N} \exp\left(-\frac{d_{ij}^2}{d_c^2}\right) \tag{14}$$

where d_c denotes the cut-off distance. It should be set to make the average number of neighbours for each sample about 1%∼2% of the total number of samples. Suppose $D = \{d_k\}_{k=1}^{N\times N}$ is the set of distance between all samples, which are sorted from small to large as $d_1 < d_2 < \cdots < d_{N\times N}$. In this study the cut-off distance d_c is set by $d_c = d_M$, where $M = [0.02N \times N]$.

Then, separation distance δ_i between the samples of higher density and sample x_i is calculated as follows:

$$\delta_i = \min_{j:\rho_j > \rho_i} d_{ij} \tag{15}$$

For the sample that has the highest local density, we simply let

$$\delta_i = \max_{j\neq i} \delta_j \tag{16}$$

Finally, for each sample x_i, (ρ_i, δ_i) can be calculated. All (ρ_i, δ_i) are drawn in the two-dimensional map, which is the decision graph. In the decision graph, the sample with the relatively higher local density and relatively larger distance is selected as the cluster centre. A more detail description concerning the DPC algorithm can be found in literature [32], [34].

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, two experimental cases are applied to test the fault diagnosis performance of the proposed method. The first experimental case is based on a bearing dataset from CWRU, and the second experimental case is based on a bearing dataset from GPKLFD. For each experimental case, comparative studies among multiple multi-scale analysis methods and several existing clustering methods are performed. Besides, Gaussian noises with different SNRs are added to original vibration signals to test the robustness of the proposed method.

A. CASE 1: BEARING DATASET FROM CWRU

Bearing dataset collected from the driven end of a motor in 1hp (1772 rpm) from CWRU is used to valid the effectiveness and noise robustness of the proposed method. The bearing vibration signals under four different working conditions: normal condition (N), Ball fault (B), inner race fault (IR) and outer race fault (OR), are employed. Furthermore, for B, IR and OR conditions, vibration signals of three severity levels

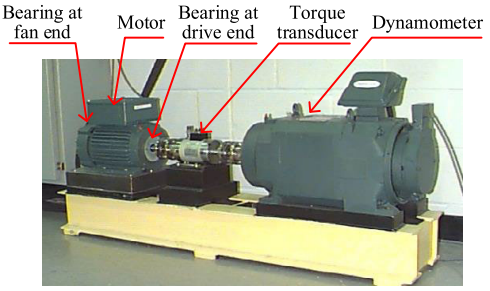


FIGURE 2. A motor platform.

(0.007, 0.014 and 0.021 mils) are included. As such, a total of 10 different fault conditions are tested. For convenience, these 10 working conditions are expressed as N, B-007, B-014, B-021, IR-007, IR-014, IR-021, OR-007, OR-014, and OR-021, respectively. Besides, the sampling frequency of the signals is 12 kHz, and the sample length is 2400. Figure 2 shows the experimental platform device [39].

1) ECMDES EXTRACTION

The raw vibration signals are then processed using the ensemble composite multi-scale dispersion entropy method, and the scale factor is set to 10 in this study. As such, a feature set containing 271 ECMDEs is obtained. Specifically, it consists of five different types of CMDEs, including 55 CMDEs based on mean value (CMDE*smean*), 54 CMDEs based on root mean square value (CMDE*srms*), 54 CMDEs based on variance value (CMDE*svar*), 54 CMDEs based on root amplitude value (CMDE*sra*) and 54 CMDEs based on maximum value (CMDE*smax*). Through synthesizing different kinds of CMDEs, the proposed ECMDEs are finally obtained. The goal of ECMDEs is to find more effective fault information. It is noted that our proposed method is based on the 271 ECMDEs. Besides, the 55 CMDE*smean* are the original CMDEs. Figure 3(a)-(e) respectively show the mean curves of the extracted CMDEs based on mean value, root mean square value, variance value, root amplitude value and maximum value for different bearing conditions. It can be seen that different kinds of CMDEs have different distinguishability for various bearing conditions. ECMDEs are good feature representations that can provide more effective fault information than the original CMDEs.

2) CLUSTERING RESULTS AND ANALYSIS

Based on the extracted ECMDEs, LPP is used to find the essential low-dimensional embedding manifold. After applying the LPP, several reduced features (RFs) are selected as the sensitive features for clustering analysis. In order to illustrate the superiority of the proposed method, five different kinds of CMDEs based on mean value, root mean square value, variance value, root amplitude value and maximum value with LPP and DPC, are respectively adopted for comparing their clustering performance. In this experiment, using the LPP method the extracted ECMDEs are reduced to nine features, as illustrated in Figure 4.

In order to quantify the distinguishability of the RFs obtained by using ECMDEs and different kinds of CMDEs for different fault types, Sensitivity values based on the distance ratios of between-class and within-class distances are calculated. The sensitivity values of the RFs are presented in Figure 5, where the x-axis represents the RFs, and the y-axis is the sensitivity value for each RF. From the Figure 5, it can be seen that the average sensitivity value of the RFs obtained by using ECMDEs is higher than that of the RFs obtained by using each kind of CMDE, indicating that the ECMDEs has better distinguishability than each kind of CMDE.

Based on the RFs extracted by LPP, DPC method is then used to conduct bearing fault identification. A representation called decision graph is introduced to help one find cluster centres. The decision graph is a plot of the separation distance as a function of the local density for each sample. In the decision graph, only the samples of relatively high density and high distance are the cluster centres. The decision graph generated by DPC of using ECMDEs is shown in Figure 6(a). The samples in the red rectangle are the cluster centres clearly displayed in the decision graph. For the proposed method, it can be seen from Figure 6(a) that the number of clusters is ten, which is equal to the actual number of fault categories. Additionally, in order to quantitatively evaluate the clustering results, two measures, F score (F) and Jaccard index (J), are used for evaluating the clustering performance. They are computed as follows:

$$F = 2 * (Recall * Precision) \big/ (Recall + Precision) \tag{17}$$

$$J = TP \big/ (TP + FN + FP) \tag{18}$$

$$precision = TP \big/ (TP + FP) \tag{19}$$

$$recall = TP \big/ (TP + FN) \tag{20}$$

where TP is the true positive representing the number of pair points having same class labels belonging to the same clusters, FP is the false positive representing the number of pair points having different class labels belonging to the same clusters, FN is the false negative representing the number of pair points having same class labels belonging to different clusters, and TN is the true negative representing the number of pair points having different class labels belonging to the different clusters. Note that the higher the F value and J value are, the better the clustering result is. Through using the DPC method, the detailed clustering result of the proposed method is obtained, as shown in Figure 6(b). For the proposed method, the F value and J value are both equal to 100%, indicating that all samples are correctly assigned to the corresponding labels. In order to illustrate the superiority of the proposed method, the clustering results using the DPC for the five different types of CMDEs are respectively shown in Figures 7-11. For Figures 7(a)-11(a), there are eight samples, five samples, four samples four samples and two samples, identified as cluster centres, respectively. After finding the cluster centres, the remaining samples are automatically

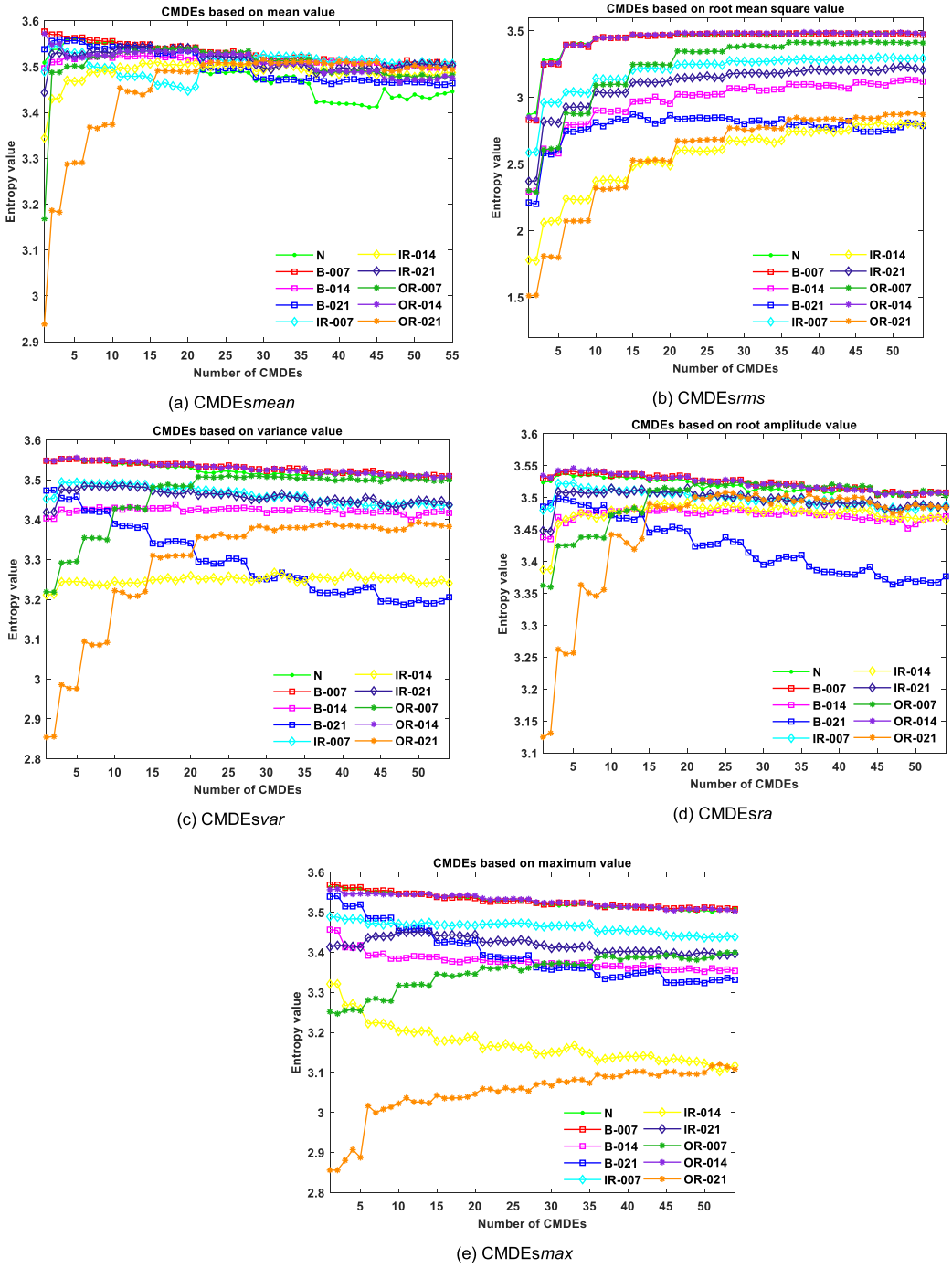


FIGURE 3. Data distribution of different types of CMDEs for the various bearing conditions.

assigned to the corresponding cluster centres. The sample clustering results of these five compared methods are shown in Figures 7(b)-11(b). For the ECMDEs and different kinds of CMDEs, the detailed quantitative clustering results are shown in Table 1.

It can be found from Table 1 that only for the proposed method, the number of the identified clusters is equal to the actual number of bearing fault categories. In addition, the F value and J value of the proposed method have the highest values among these compared methods. For the CMDEs*mean*, the F value and J value are 76.9% and 62.5%, respectively. This clustering result is not ideal. For the CMDEs*rms*, CMDEs*var* and CMDEs*ra*, the F value and J value are low, indicating the clustering results are poor. Especially, the clustering method based on the CMDEs*max* has the worst clustering result. The number of the identified

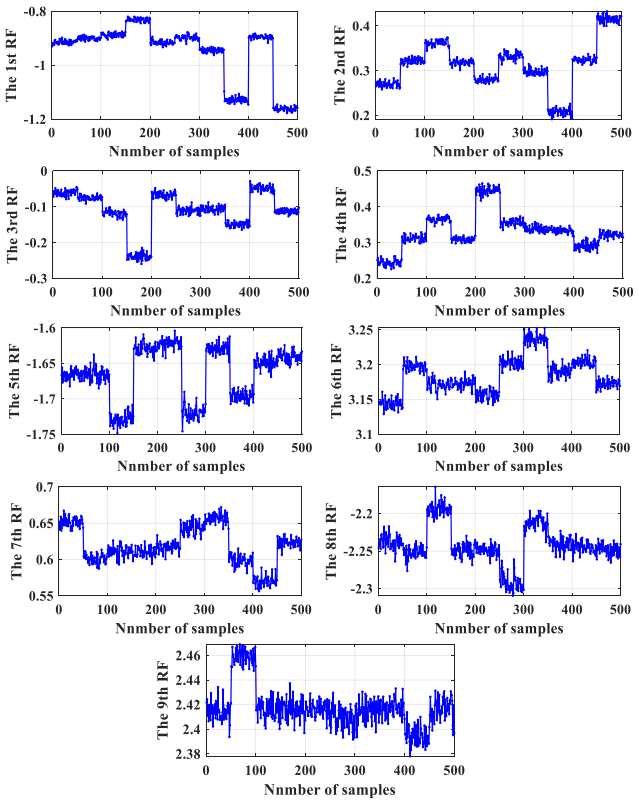


FIGURE 4. Nine RFs generated by the LPP method for ECMDEs.

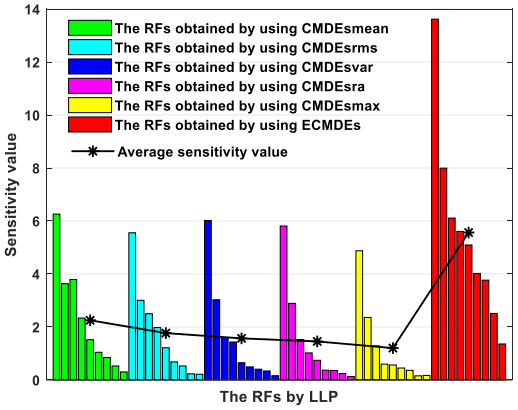


FIGURE 5. Sensitivity values of the RFs obtained by using ECMDEs and different kinds of CMDEs.

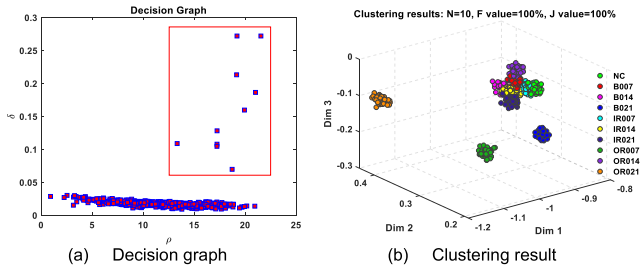


FIGURE 6. Clustering result of using the DPC for ECMDEs.

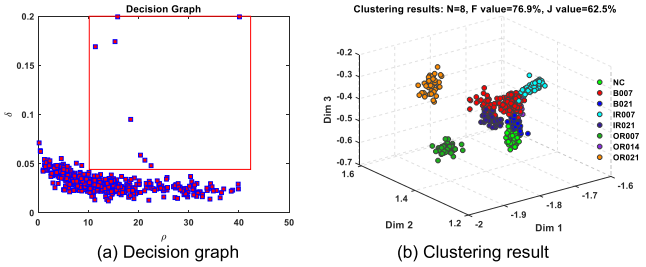


FIGURE 7. Clustering result of using the DPC for CMDEs*mean*.

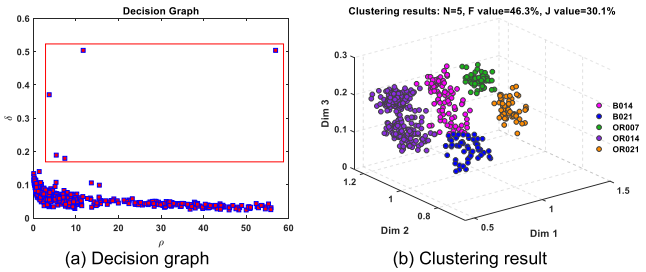


FIGURE 8. Clustering result of using the DPC for CMDEs*rms*.

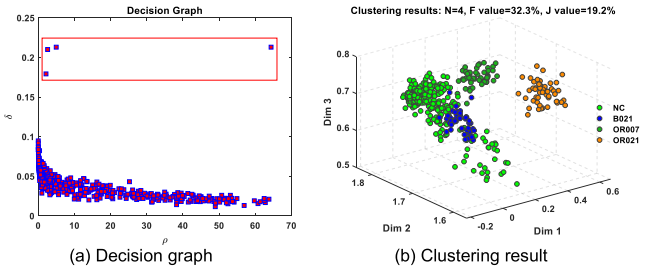


FIGURE 9. Clustering result of using the DPC for CMDEs*var*.

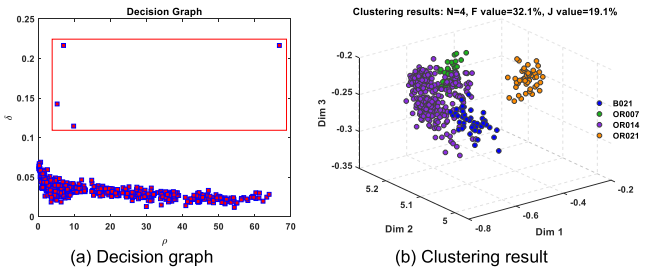


FIGURE 10. Clustering result of using the DPC for CMDEs*ra*.

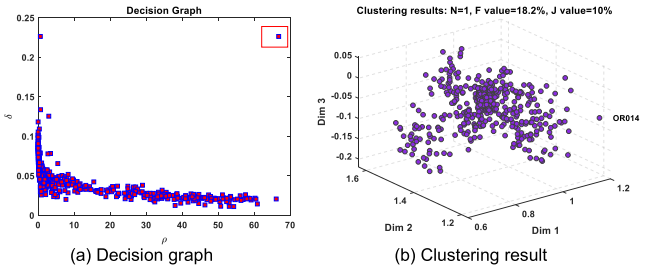


FIGURE 11. Clustering result of using the DPC for CMDEs*max*.

clusters is one. The F value and J value are only 18.2% and 10%, respectively. These quantitative clustering results reveal that the proposed method has better diagnosis performance than the compared methods.

TABLE 1. Quantitative Clustering Results Using ECMDEs and Different Kinds of CMDEs.

Different methods	Number of the identified clusters (True number of fault categories)	F value	J value
CMDE*smean*	8(10)	76.9%	62.5%
CMDE*srms*	5(10)	46.3%	30.1%
CMDE*svar*	4(10)	32.3%	19.2%
CMDE*sra*	4(10)	32.1%	19.1%
CMDE*smax*	1(10)	18.2%	10%
ECMDEs	10(10)	100%	100%

TABLE 2. Quantitative Clustering Results of ECMDEs Using Different Dimension Reduction Methods.

Different methods	Number of the identified clusters (True number of fault categories)	F value	J value
ECMDEs+DPC	3(10)	35.7%	21.7%
ECMDEs+PCA+DPC	3(10)	35.7%	21.7%
ECMDEs+t-SNE+DPC	7(10)	71.6%	55.7%
the proposed method	10(10)	100%	100%

In order to clearly illustrate the clustering results, a data visualization tool, t-distributed stochastic neighbour embedding (t-SNE) [40], is used to visualize the distribution of samples clustered by DPC algorithm. Figure 12 shows the t-SNE visualization result for the ECMDEs and different kinds of CMDEs. It can be seen from Figure 11 that the samples clustered by DPC based on ECMDEs can be completely separated.

In order to illustrate the necessity of dimension reduction using LPP, Table 2 gives the diagnosis performance comparison among ECMDEs without using dimension reduction method and with using different dimension reduction methods, such as principal component analysis (PCA) and t-SNE. It can be observed that only for the proposed method, the number of the identified clusters is equal to the actual number, and the F value and J value are 100%. However, for the ECMDEs without using dimension reduction method and with using PCA or t-SNE, clustering results are very poor.

Furthermore, in order to further illustrate the superiority of the proposed method, other seven multi-scale analysis methods, such as multi-scale sample entropy (MSE) [2], multi-scale permutation entropy (MPE) [5], multi-scale dispersion entropy (MDE) [41], composite multi-scale sample entropy (CMSE) [42], composite multi-scale permutation entropy (CMPE) [6], composite multi-scale dispersion entropy (CMDE) [21], and refined composite multi-scale dispersion entropy (RCMDE) [22], are respectively combined with LPP and DPC to diagnose bearing faults. Detailed clustering results are displayed in Table 3. It can be observed that only for the proposed method, the number of the identified clusters is equal to the true number of bearing fault categories. Besides, the proposed method has the highest F value and J value among these compared methods. This result illustrates that the proposed method can accurately achieve the fault sample clustering. According to the comparison results

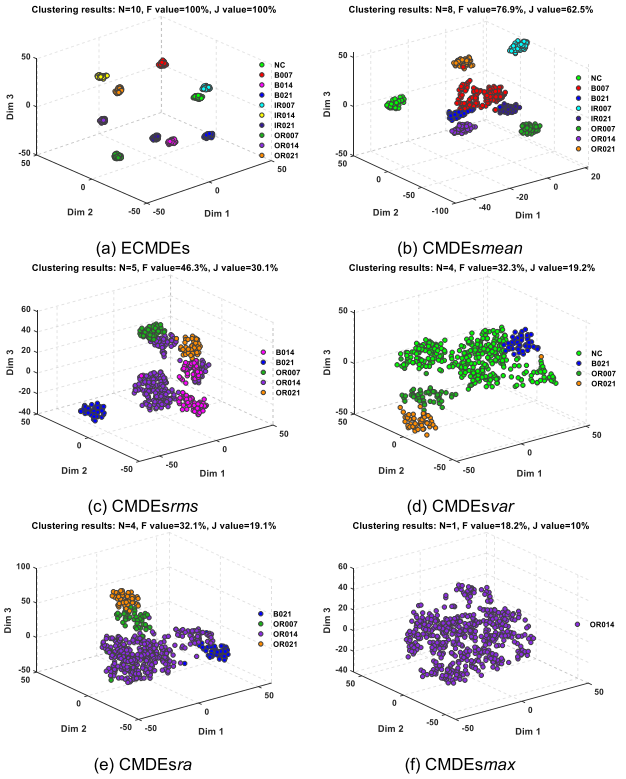


FIGURE 12. The t-SNE visualization of clustering results.

TABLE 3. Quantitative Clustering Results of Different Multi-Scale Analysis Methods.

Different multi-scale analysis methods	Number of the identified clusters (True number of fault categories)	F value	J value
MSE	7(10)	73.8%	58.5%
MPE	3(10)	38.4%	23.8%
MDE	8(10)	74.7%	59.6%
CMSE	5(10)	54%	36.9%
CMPE	6(10)	49.9%	33.2%
CMDE	8(10)	76.9%	62.5%
RCMDE	9(10)	87%	77%
the proposed method	10(10)	100%	100%

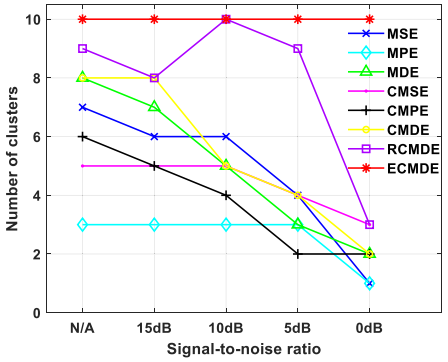
in Table 3, it can be concluded that the proposed method can greatly improve the diagnosis performance.

3) ROBUSTNESS ANALYSIS OF THE PROPOSED METHOD

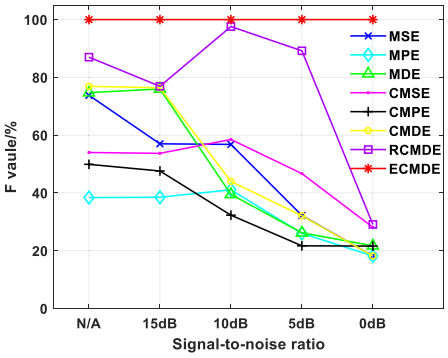
Usually, in practical applications, a variety of background noises often accompany with the collected vibration signals and different levels of noise have various negative impacts on the diagnosis results. In order to explore the robustness performance of the proposed method in different environmental noise, additional Gaussian noise with SNR of 0, 5, 10 and 15 dB is respectively added on the original vibration signals. The definition of SNR is defined as follows.

$$SNR\,(dB) = 10log_{10}(P_{signal}/P_{noise})$$

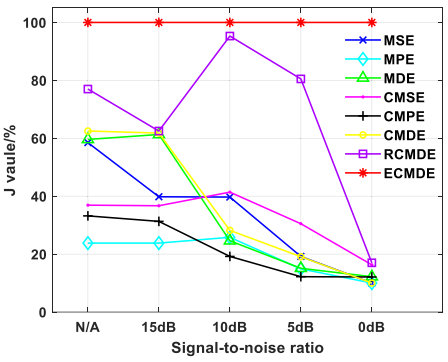
where P_{signal} and P_{noise} denote the powers of the original signal and the additional Gaussian noise, respectively. The



(a) Number of clusters

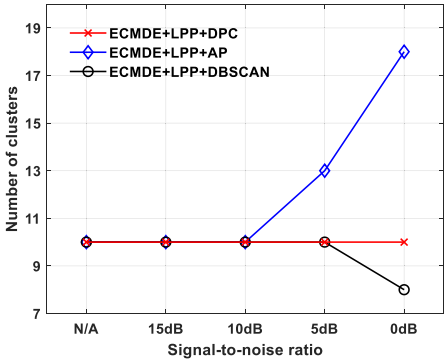


(b) *F* value

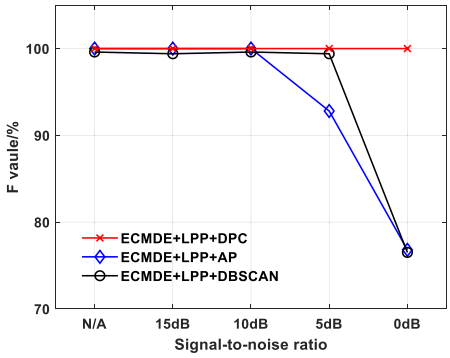


(c) *J* value

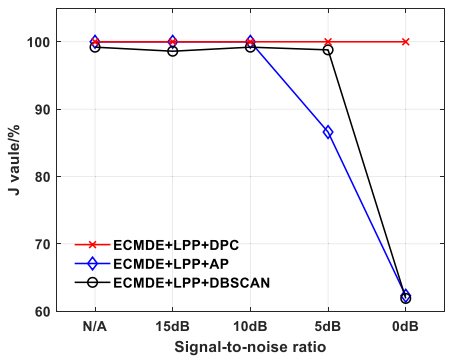
FIGURE 13. Clustering results comparison among different multi-scale analysis methods under noise-free and different SNR conditions (SNR = 0, 5, 10 and 15 dB).



(a) Number of clusters



(b) *F* value



(c) *J* value

FIGURE 14. Clustering results comparison among different clustering methods under noise-free and different SNR conditions (SNR = 0, 5, 10 and 15 dB).

higher SNR, the more useful information and the less noise of signal. The clustering results under noise-free condition and various SNR levels (SNR = 0, 5, 10 and 15 dB) are shown in Figure 13, where the N/A represents the noise-free condition. It can be observed that for the proposed method under noise-free condition and various SNR levels, the number of the identified clusters are all correct, and the F value and J value are significantly higher than that of the compared methods. However, for the compared methods under noise-free condition and various SNR levels, the number of the identified clusters is almost false. Besides, the F value and J value are significantly decreased when the SNR is decreased from 15dB to 0dB. For example, for the compared method based on MSE with LPP and DPC, when the SNR is decreased from 15 dB to 0 dB, the F value has fallen from 73.8% to 18.2%, and the J value has fallen from 58.5% to 10%. Similarly, for other compared methods, the F value and J value are decreased in various degrees as well when the SNR is decreased from 15 dB to 0 dB. That is, when the noise is strong, the clustering results of these compared methods become poor. However, the proposed method performs well even with the decreasing of the SNR. This result indicates that the proposed method is more robust against noise than the compared methods.

In addition, in order to further illustrate the noise robustness of the proposed method, a comparison with other unsupervised clustering methods such as the density-based spatial clustering of applications with noise (DBSCAN) [43] and affinity propagation clustering (APC) [44] is also performed. Figure 14 depicts the clustering results under

TABLE 4. The Parameters of the Used Rolling Bearing.

Bearing type	Inside diameter	Outside diameter	Thickness	Number of the roller
NSK1310	50 mm	110 mm	27mm	13

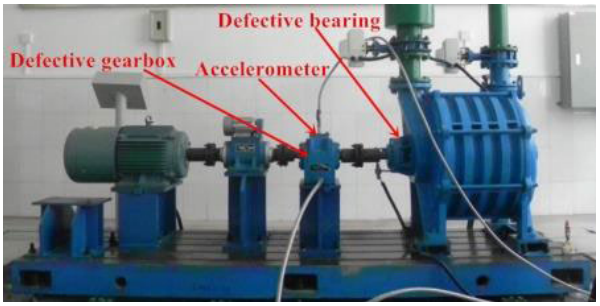


FIGURE 15. Developed real test bed.

noise-free condition and various SNR levels. It can be found from Figure 14 that clustering result obtained by using the DPC method is better than that of obtained by using the DBSCAN and APC. The proposed method can obtain good clustering effect even though the SNR of the measured data is low, showing that the proposed method is robust against noise. However, when the noise signal becomes strong, the clustering effects of using the APC and DBSCAN are declined. This result proves that the DPC method has superior clustering performance than other two clustering methods DBSCAN and APC. Meanwhile, it suggests that the proposed method has a strong robustness to identify different bearing fault conditions under various SNRs.

B. CASE 2: BEARING DATASET FROM GPKLFD

Bearing dataset collected from a centrifugal multi-level impeller blower from GPKLFD is used to valid the effectiveness and noise robustness of the proposed method. The bearing vibration signals under four different working conditions: normal condition (NC), Ball fault (BF), inner race fault (IR) and outer race fault (OR), are employed. For convenience, these four fault conditions are expressed as NC, BF, IR and OR, respectively. The centrifugal multi-level impeller blower is mainly composed of an electric motor, shaft, gearbox, bearing and an 11 kW five-stage centrifugal fan, as shown in Figure 15. The rolling bearing used in the experiment is NSK1310, and more parameters concerning the NSK1310 are described in Table 4. Figure 16 shows the base station of data, and the defective bearings are illustrated in Figure 17. The sampling frequency of the signals is 1 kHz, and the sample length is 2400. For each bearing condition, 100 groups of vibration data are measured.

1) ECMDES EXTRACTION

The raw vibration signals are then processed using the ensemble composite multi-scale dispersion entropy method, and the scale factor is set to 10 in this study. As such, a feature set containing 271 ECMDEs is obtained. Specifically,

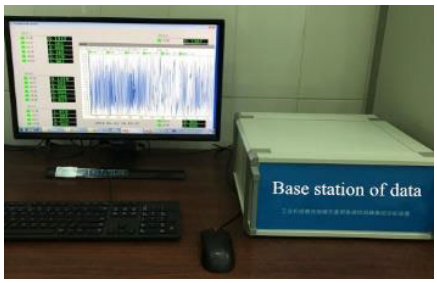


FIGURE 16. Base station of data.

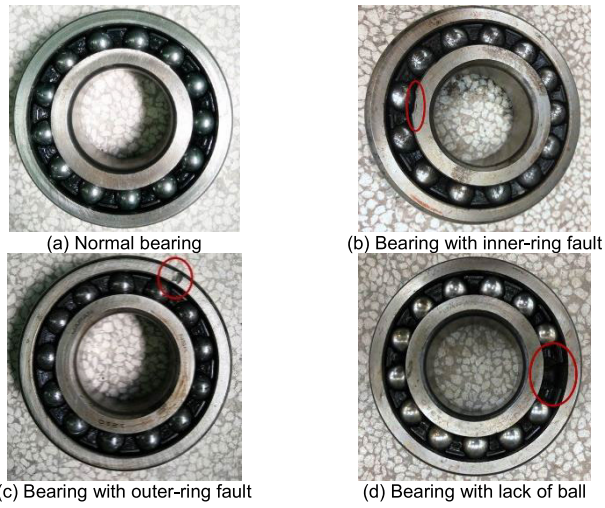


FIGURE 17. Faulty components.

it consists of five different types of CMDEs, including 55 CMDE*smean*, 54 CMDE*srms*, 54 CMDE*svar*, 54 CMDE*sra* and 54 CMDE*smax*. The goal of ECMDEs is to synthesize different kinds of CMDEs to find more effective fault information. Figure18(a)-(e) respectively show the mean curves of the extracted CMDEs based on mean value, root mean square value, variance value, root amplitude value and maximum value for different bearing conditions. It can be seen that different kinds of CMDEs have different distinguishability for the various bearing conditions. ECMDEs are good feature representations that can provide more effective fault information than the original CMDEs.

2) CLUSTERING RESULTS AND ANALYSIS

Similar to Case 1, based on the extracted ECMDEs, LPP is used to find the essential low-dimensional embedding manifold. In this experiment, the extracted ECMDEs are reduced to three features using LPP method, as illustrated in Figure 19. These three RFs are selected as the sensitive features for clustering analysis. In order to illustrate the superiority of the proposed method, five different kinds of CMDEs based on mean value, root mean square value, variance value, root amplitude value and maximum value with LPP and DPC, are respectively adopted for comparing their clustering performance.

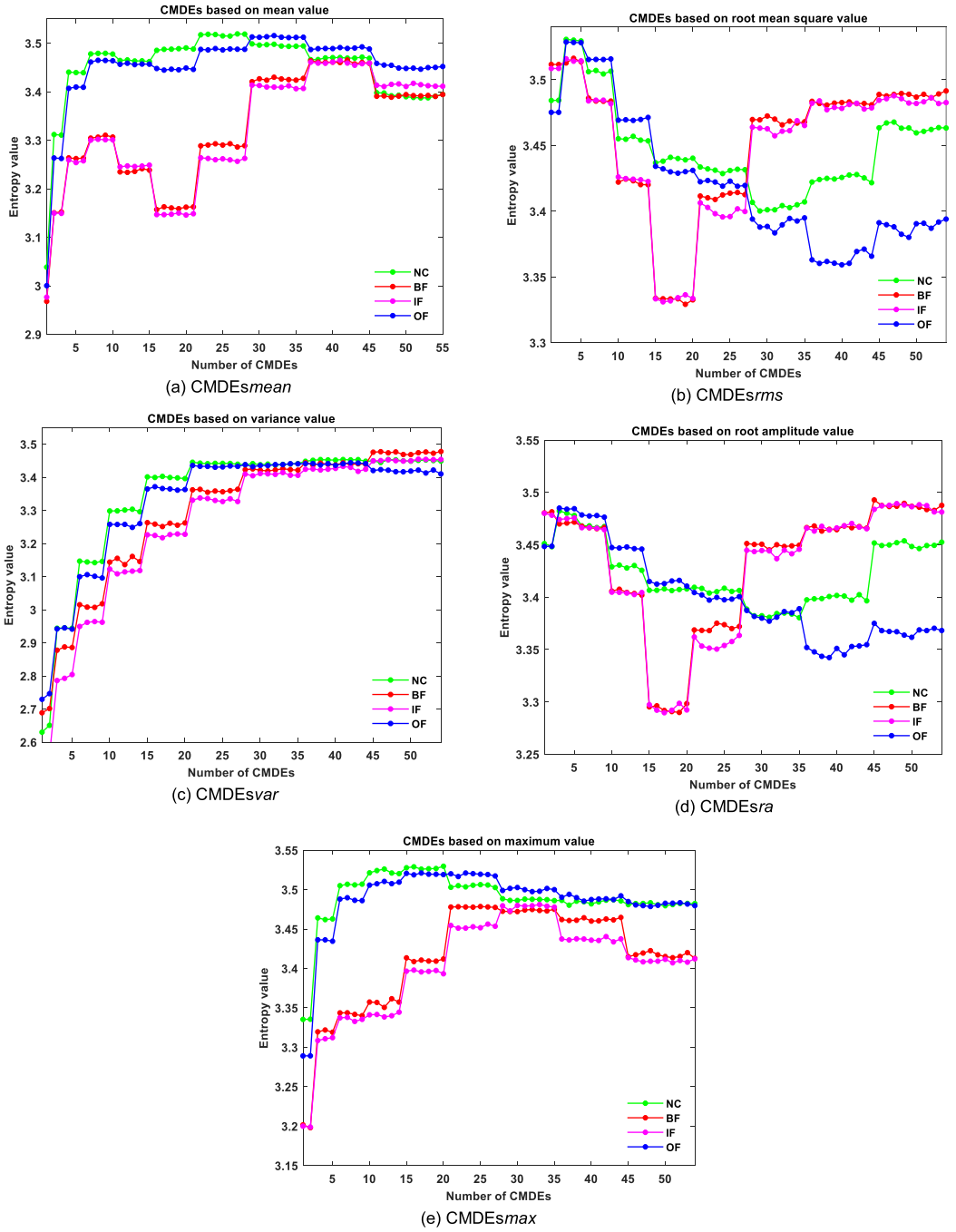


FIGURE 18. **Data distribution of different kinds of CMDEs for the various bearing conditions.**

In order to quantify the distinguishability of the RFs obtained by using ECMDEs and different kinds of CMDEs for bearing fault types, the sensitivity values based on the distance ratios of between-class and within-class distances are calculated. The sensitivity values of the RFs are presented in Figure 20, where the x-axis represents the RFs, and the y-axis is the sensitivity value for each RF. From the Figure 20, it can be seen that the average sensitivity value of the RFs obtained by using ECMDEs is higher than that of the RFs obtained by using each kind of CMDE, indicating that the ECMDEs has better distinguishability than each kind of CMDE.

Based on the RFs extracted by LPP, DPC method is then used to conduct bearing fault diagnosis. The decision graph generated by DPC for ECMDEs is shown in Figure 21(a). The samples in the red rectangle are the cluster centres clearly displayed in the decision graph. For the proposed method, it can be seen from Figure 21(a) that the number of the clusters is

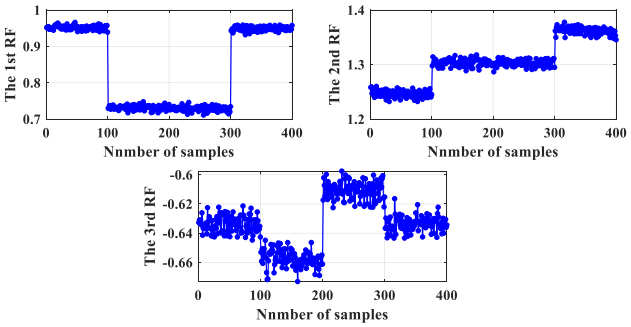


FIGURE 19. Three RFs generated by the LPP method for ECMDEs.

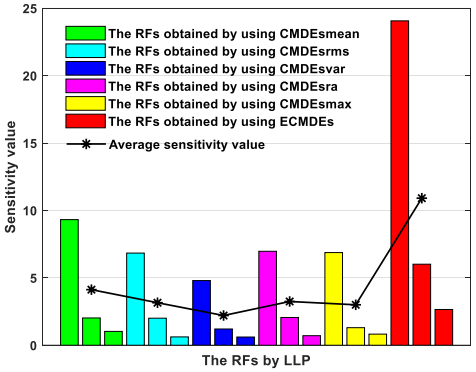
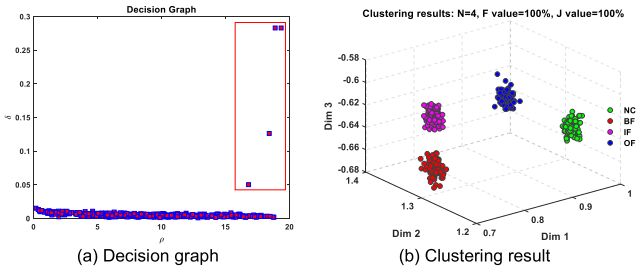
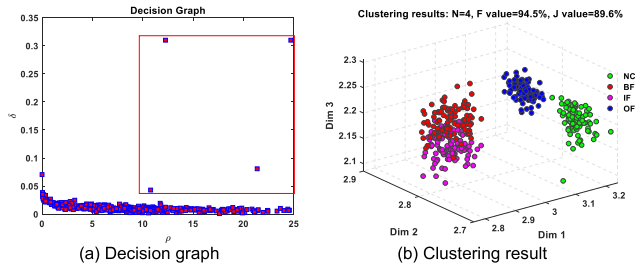


FIGURE 20. Sensitivity values of the RFs obtained by using ECMDEs and different kinds of CMDEs.



(a) Decision graph (b) Clustering result

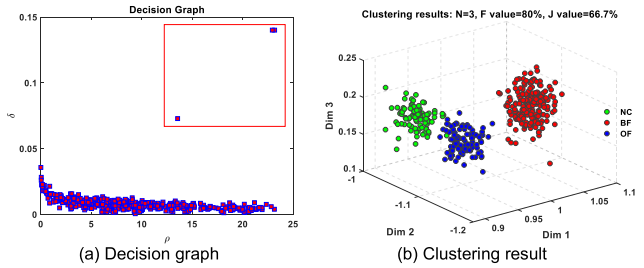
FIGURE 21. Clustering result of using the DPC for ECMDEs.



(a) Decision graph (b) Clustering result

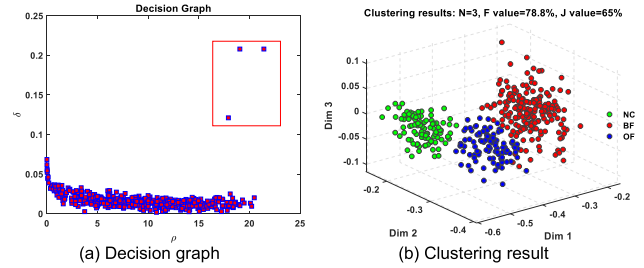
FIGURE 22. Clustering result of using the DPC for CMDEs*mean*.

four, which is equal to the actual number of fault categories. Through using the DPC algorithm, the detailed sample clustering result of the proposed method is obtained, as shown in Figure 21(b). The F value and J value are both equal to 100%, indicating that the samples in each cluster clustered by DPC are the samples in real situations. Furthermore, in order to



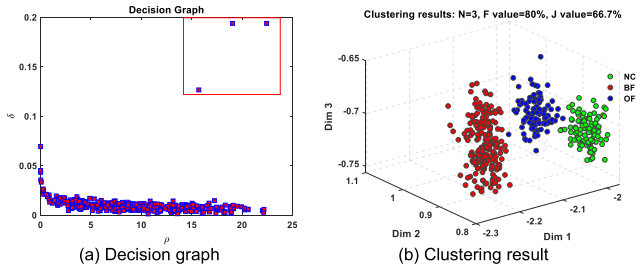
(a) Decision graph (b) Clustering result

FIGURE 23. Clustering result of using the DPC for CMDEs*rms*.



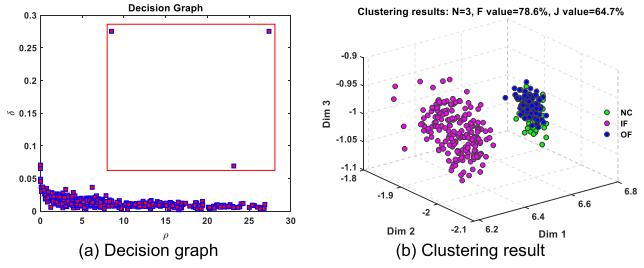
(a) Decision graph (b) Clustering result

FIGURE 24. Clustering result of using the DPC for CMDEs*var*.



(a) Decision graph (b) Clustering result

FIGURE 25. Clustering result of using the DPC for CMDEs*ra*.



(a) Decision graph (b) Clustering result

FIGURE 26. Clustering result of using the DPC for CMDEs*max*.

illustrate the superiority of the proposed method, the clustering results using the DPC for different types of CMDEs are respectively shown in Figures 22-26. For Figures 22(a)-26(a), there are four samples, three samples, three samples three samples and three samples, identified as cluster centres, respectively. After finding the cluster centres, the remaining samples are automatically assigned to the corresponding cluster centres. The clustering results of these compared methods are shown in Figures 22(b)-26(b). For the ECMDEs and the five different kinds of CMDEs, the detailed quantitative clustering results are shown in Table 5.

TABLE 5. Quantitative Clustering Results Using ECMDEs and Different Kinds of CMDEs.

Different methods	Number of the identified clusters (True number of fault categories)	F value	J value
CMDE*smean*	4(4)	94.5%	89.6%
CMDE*srms*	3(4)	80%	66.7%
CMDE*svar*	3(4)	78.8%	65%
CMDE*sra*	3(4)	80%	66.7%
CMDE*smax*	3(4)	78.6%	64.7%
ECMDEs	4(4)	100%	100%

TABLE 6. Quantitative Clustering Results Using ECMDEs of Different Dimension Reduction Methods.

Different methods	Number of the identified clusters (True number of fault categories)	F value	J value
ECMDEs+DPC	2(4)	66.5%	49.8%
ECMDEs+PCA+DPC	3(4)	76.6%	62.1%
ECMDEs+ t-SNE+DPC	3(4)	79.3%	65.7%
the proposed method	4(4)	100%	100%

It can be found from Table 5 that only for the proposed method, the number of the identified clusters is equal to the actual number of bearing fault categories. In addition, the F value and J value of the proposed method have the highest value among these compared methods. For the CMDE*smean*, the F value and J value are 94.5% and 89.6%, respectively. This clustering result is not ideal. For the CMDEs based on root mean square value, variance value and root amplitude value, the F value and J value are low as well, indicating the clustering results are poor. These quantitative clustering results reveal that the proposed method has better diagnosis performance than the compared methods. In this experiment, since the dimension of the reduced features is three, it is unnecessary to use the t-SNE tool for data visualization.

Table 6 gives the diagnosis performance comparison among ECMDEs without using dimension reduction method and with using different dimension reduction methods, such as PCA and t-SNE. It can be observed that only for the proposed method, the number of the identified clusters is equal to the actual number, and the F value and J value are 100%. However, for the ECMDEs without using dimension reduction method and with using PCA or t-SNE, clustering results are very poor.

In order to further illustrate the superiority of the proposed method, other seven multi-scale analysis methods are respectively combined with LPP and DPC to diagnose bearing faults. Detailed clustering results are displayed in Table 7. It can be found that for the MPE, CMPE, CMDE and the proposed method, the number of the identified clusters is equal to the actual number of bearing fault categories. However, the proposed method has the highest F value of 100% and J value of 100% among these compared methods. The clustering effect of using CMPE is second only to that of the proposed method. Through using the CMPE, the F value is 99.8% and J value is 99.5%. According to the comparison results in Table 7, it can be concluded that the proposed method can significantly improve the clustering effect.

TABLE 7. Quantitative Clustering Results of Different Multi-Scale Analysis Methods.

Different multi-scale analysis methods	Number of the identified clusters (True number of fault categories)	F value	J value
MSE	3(4)	79%	65.3%
MPE	4(4)	95%	90.5%
MDE	3(4)	78.8%	65%
CMSE	3(4)	79.2%	65.6%
CMPE	4(4)	99.8%	99.5%
CMDE	4(4)	94.5%	89.6%
RCMDE	3(4)	79.8%	66.4%
the proposed method	4(4)	100%	100%

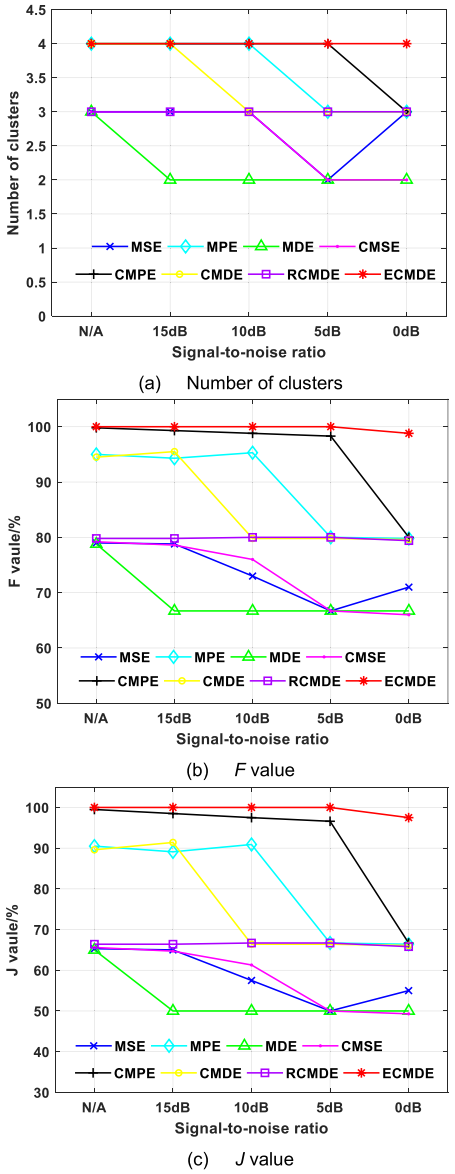


FIGURE 27. Clustering results comparison among different multi-scale analysis methods under noise-free and different SNR conditions (SNR = 0, 5, 10 and 15 dB).

3) ROBUSTNESS ANALYSIS OF THE PROPOSED METHOD

Similar to Case 1, additional Gaussian noise with SNR of 0, 5, 10 and 15 dB is respectively added to the original vibration signals. The clustering results under noise-free and different SNR conditions are shown in Figure 27, where the N/A

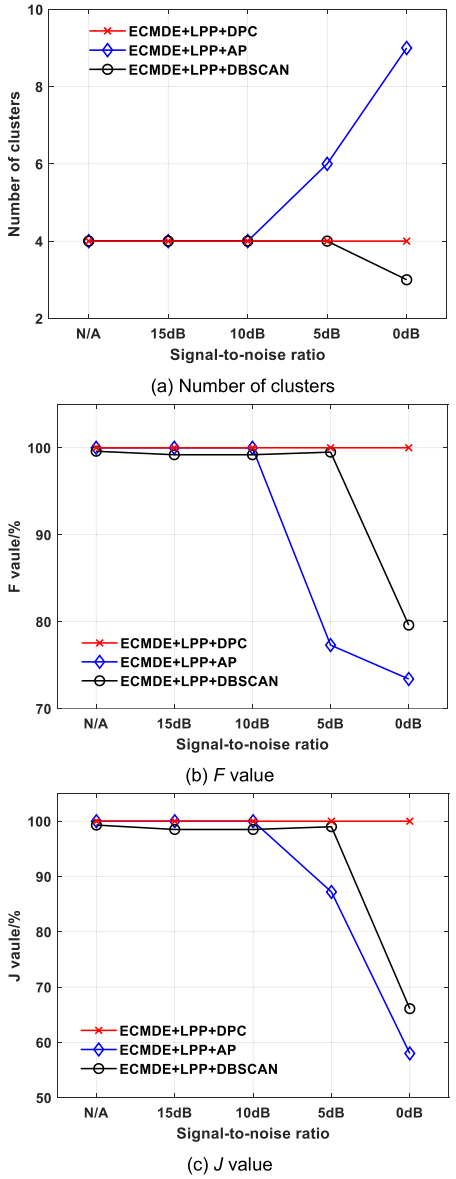


FIGURE 28. Clustering results comparison among different clustering methods under noise-free and different SNR conditions (SNR = 0, 5, 10 and 15 dB).

represents the noise-free condition. It can be observed that for the proposed method under noise-free condition and various SNR levels, the number of the identified clusters are all correct, and the F value and J value are significantly higher than that of the compared methods. However, for the compared methods, the F value and J value are significantly decreased when the SNR is decreased from 15 dB to 0 dB. The clustering effect of using CMPE is second only to that of using ECMDE. Using the CMPE under noise-free and three SNR conditions (SNR = 5, 10 and 15 dB), the correct number of clusters and higher F value and J value are obtained. However, when the SNR is 0 dB, the F value and J value are only 80% and 66.7%, respectively. In addition, using the MPE and CMDE under noise-free and higher SNR conditions, the correct number of clusters and higher F value and J value are obtained as well. However, for the MPE under SNRs <= 5 dB, the F value and J value are only 80% and 66.7%. For the CMDE under SNRs <= 10 dB, the F value and J value are only 79.8% and 66.4%, respectively. Furthermore, for other kinds of multi-scale fault features, the clustering effect under noise-free and different SNR conditions is always unsatisfactory. When the noise is strong, the clustering results of the compared methods become poor. However, the proposed method performs well even with the decreasing of the SNR. This result indicates that the proposed method is more robust against noise than the compared methods. The proposed method can not only obtain very good clustering results with noise free signals but also are highly robust to noisy signals.

Furthermore, a comparison with other unsupervised clustering methods such as such as the DBSCAN and APC is performed to illustrate the noise robustness of the proposed method. Figure 28 depicts the clustering results using different unsupervised clustering methods under noise-free various SNR conditions. It can be found from Figure 28 that clustering result obtained using the DPC method is better than that of obtained using the DBSCAN and APC. The proposed method can obtain good clustering result even though the SNR of the measured data becomes low, showing that the proposed method is robust against noise. However, when the noise signal becomes strong, the clustering effects of using the DBSCAN and APC are declined. This result validates that the DPC method has superior clustering performance than other two clustering methods DBSCAN and APC in low SNRs. Meanwhile, it suggests that the proposed method has a strong robustness to identify different bearing faults under different SNR conditions.

V. CONCLUSION

A novel hybrid fault diagnosis approach by integrating ECMDEs, LPP and DPC is proposed in this study. ECMDEs as new fault features are proposed for bearing fault diagnosis. Based on the multiple ECMDEs, the DPC with the manifold learning method LPP is explored to identify the bearing fault conditions. Extensive comparison results illustrate that the proposed ECMDEs have better diagnosis performance than the multiple existing multi-scale fault features. Besides, considering noise is an inevitable problem in practical applications, the noise robustness of the proposed method has been carefully analyzed and discussed in details. Through comparing with several state-of-the-art methods, the proposed method can obtain a satisfactory diagnosis result even when the SNR of the measured data is low, indicating that the proposed method has a stronger noise robustness performance. Two experimental cases have validated that the proposed method is capable to reliably extract more effective information of raw signal and accurately diagnose bearing faults.

REFERENCES

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[2] Y. Li, K. Feng, X. Liang, and M. J. Zuo, "A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy," *J. Sound Vib.*, vol. 439, pp. 271–286, Jan. 2019.
[3] M. Han and J. Pan, "A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings," *Measurement*, vol. 76, pp. 7–19, Dec. 2015.
[4] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines," *Mech. Syst. Signal Process.*, vol. 29, pp. 474–484, May 2012.
[5] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, and C.-C. Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine," *Entropy*, vol. 14, no. 8, pp. 1343–1356, Jul. 2012.
[6] J. Zheng, H. Pan, S. Yang, and J. Cheng, "Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 99, pp. 229–243, Jan. 2018.
[7] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. Signal Process.*, vol. 85, pp. 746–759, Feb. 2017.
[8] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, pp. 35042–35056, 2018.
[9] H. Zhao, M. Sun, W. Deng, and X. Yang, "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 14, pp. 1–21, 2017.
[10] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism Mach. Theory*, vol. 78, pp. 187–200, Aug. 2014.
[11] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, Jul. 2017.
[12] H. Zhao, H. Liu, J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4165–4172, Jul. 2020.
[13] X. Wang, S. Si, and Y. Li, "Multiscale diversity entropy: A novel dynamical measure for fault diagnosis of rotating machinery," *IEEE Trans. Ind. Informat.*, early access, Sep. 11, 2020, doi: 10.1109/tii.2020.3022369.
[14] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review," *IEEE Access*, vol. 6, pp. 66723–66741, 2018.
[15] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy based fault classification using the case western reserve university data: A benchmark study," *IEEE Trans. Rel.*, vol. 69, no. 2, pp. 754–767, Jun. 2020.
[16] K. Feng, K. Wang, Q. Ni, M. J. Zuo, and D. Wei, "A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions," *J. Sound Vib.*, vol. 408, pp. 190–209, Nov. 2017.
[17] Y. Li, M. Xu, R. Wang, and W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *J. Sound Vib.*, vol. 360, pp. 277–299, Jan. 2016.
[18] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016.
[19] M. Rostaghi, M. R. Ashory, and H. Azami, "Application of dispersion entropy to status characterization of rotary machines," *J. Sound Vib.*, vol. 438, pp. 291–308, Jan. 2019.
[20] J. Zheng, H. Pan, Q. Liu, and K. Ding, "Refined time-shift multiscale normalised dispersion entropy and its application to fault diagnosis of rolling bearing," *Phys. A, Stat. Mech. Appl.*, vol. 545, May 2020, Art. no. 123641.
[21] X. Yan and M. Jia, "Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection," *Knowl.-Based Syst.*, vol. 163, pp. 450–471, Jan. 2019.
[22] H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2872–2879, Dec. 2017.
[23] W. Zhang and J. Zhou, "A comprehensive fault diagnosis method for rolling bearings based on refined composite multiscale dispersion entropy and fast ensemble empirical mode decomposition," *Entropy*, vol. 21, no. 680, pp. 1–19, 2019.
[24] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
[25] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[26] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[27] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, Oct. 2018.
[28] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 153–160.
[29] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *J. Sound Vib.*, vol. 335, pp. 367–383, Jan. 2015.
[30] J. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mech. Syst. Signal Process.*, vol. 25, no. 7, pp. 2573–2588, Oct. 2011.
[31] L. Wu, B. Yao, Z. Peng, and Y. Guan, "Fault diagnosis of roller bearings based on a wavelet neural network and manifold learning," *Appl. Sci.*, vol. 7, no. 2, pp. 1–10, 2017.
[32] H. Zheng, R. Wang, J. Yin, Y. Li, H. Lu, and M. Xu, "A new intelligent fault identification method based on transfer locality preserving projection for actual diagnosis scenario of rotating machinery," *Mech. Syst. Signal Process.*, vol. 135, Jan. 2020, Art. no. 106344.
[33] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.
[34] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, Mar. 2009.
[35] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, Jan. 1984.
[36] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[37] Y. Wang, Z. Wei, and J. Yang, "Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 105–115, Jan. 2019.
[38] Q. Hu, Q. Zhang, X.-S. Si, A.-S. Qin, and Q.-H. Zhang, "Fault diagnosis based on multi-scale redefined dimensionless indicators and density peak clustering with geodesic distances," *IEEE Access*, vol. 8, pp. 84777–84791, 2020.
[39] (Jul. 20, 2012). *Case Western Reserve University Bearings Data Set*. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file
[40] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[41] H. Shang, F. Li, and Y. Wu, "Partial discharge fault diagnosis based on multi-scale dispersion entropy and a hypersphere multiclass support vector machine," *Entropy*, vol. 21, no. 1, pp. 81–93, 2019.
[42] S.-D. Wu, C.-W. Wu, S.-G. Lin, C.-C. Wang, and K.-Y. Lee, "Time series analysis using composite multiscale entropy," *Entropy*, vol. 15, no. 3, pp. 1069–1084, Mar. 2013.
[43] M. Ester, H. P. Kriegel, J. Sander, and X. W. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, vol. 96, no. 34, 1996, pp. 226–231.
[44] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

AI-SONG QIN received the B.Sc. degree in automation and the M.Sc. degree in control theory and control engineering from the Taiyuan University of Technology, China, in 2010 and 2013, respectively. She is currently pursuing the Ph.D. degree with the Department of Mechanical Engineering, Guangxi University, China. Her research interests include the fault diagnosis of rotating machinery, clustering analysis, and signal processing.

HAN-LING MAO received the Ph.D. degree in mechanical engineering from Zhejiang University, China, in 1995. He is currently a Professor and a Ph.D. Supervisor with the School of Mechanical Engineering, Guangxi University, China. His research interests include NDT, faults diagnosis, and vibration control.

QIN HU received the master's degree in control theory and control engineering from the Guangdong University of Technology, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Automation, Rocket Force University of Engineering, China. His research interests include the fault diagnosis of rotating machinery, signal processing, and pattern recognition.

QING-HUA ZHANG received the M.Sc. degree in industrial automation and the Ph.D. degree in control theory and control engineering from the South China University of Technology, China, in 1995 and 2004, respectively. His research interests include the condition monitoring and fault diagnosis of rotating machinery, and applications of intelligent algorithms.