An Integrated Approach for Bearing Health Indicator and Stage Division Using Improved Gaussian Mixture Model and Confidence Value

Mao He and Wei Guo

***Abstract*—Bearing prognosis is of paramount and practical importance for rotating machinery. In this article, we propose an integrated approach for bearing performance degradation assessment. First, we propose an improved clustering algorithm, called the Hellinger distance-based regularized Gaussian mixture model (HRGMM). In this model, the Hellinger distance is incorporated to measure the similarity between probability distributions (PDs) of raw data. The manifold regularized GMM is then enhanced to differentiate bearing performance changes. Second, we construct a new health indicator (HI) that combines the Jensen–Rényi divergence and improved confidence value to normalize the difference in PDs between the test condition and healthy condition. It aims to solve the problems of the insensitivity of HI to the incipient defect and its highly fluctuating behavior as the bearing deteriorates over time. Clustering results of the experimental and real bearings show that the HRGMM successfully distinguishes different health states, damage severities, and working conditions. The results of two run-to-failure bearing tests demonstrate that, for each bearing, the HI based on HRGMM drops in a smooth and monotonic manner and accurately identify the incipient defect as the beginning of the slight degradation. Moreover, it can automatically mark the sharp state declines as the beginning of severe degradation and failure stages so that appropriate warnings can be set for follow-up maintenance.**

***Index Terms*—Bearing, clustering analysis, entropy, Gaussian mixture model (GMM), health indicator (HI), performance degradation assessment (PDA).**

I. Introduction

PROGNOSTICS and diagnosis are widely used tools for machine components and the entire mechanical system to provide early detection of incipient defects and predict their remaining useful life (RUL). Condition monitoring data, also referred to as degradation data, contain adequate health information and related techniques. For crucial components and highly reliable machines, degradation data-based analysis methods may be the only feasible scheme [1] to achieve the goal of prognostics and health management. These methods establish the link between monitoring data and the lifetime; thus, they remain research hotspots for prognostics.

The backbone of industrial rotating machines is bearings and gears, the former of which are the most used and vulnerable components. Their prognostics generally involve two main objectives: performance degradation assessment (PDA) and RUL prediction [2]. Their common focus is an effective health indicator (HI) that not only detects incipient defect(s) in the bearing as early as possible, but also exhibits a monotonic trend over time and minimum fluctuations. This is one of main challenges in reliable PDA because the bearing operates under varying working conditions and its defect growth is stochastic.

There are two major concerns when constructing the HI: how to extract an effective HI and how to evaluate its suitability for RUL prediction. In the literature, many features are extracted from the time, frequency, and time-frequency domains to mine degradation information. Wang *et al*. [3] fused 14 statistics by calculating its Mahalanobis distance (MD) from the healthy state. Rai and Upadhyay [4] used the MD criterion to fuse these features and chose the upward cumulative sum chart to generate a monotonically growing curve. Liu *et al*. [5] used Spearman's coefficient to select some sensitive and robust time-domain features. Zhang *et al*. [6] defined a weighted linear combination of three metrics, correlation, monotonicity, and robustness, that is maximized to select relevant features. Wu *et al*. [7] optimized a weighted linear combination of the above three metrics to select a sensitive feature set and further fused these features by using the dynamic principal component analysis (PCA) and MD. Guo *et al*. [8] defined related-similarity features and other features and utilized the monotonicity and correlation metrics with equal weights to select the most sensitive features and then fused them into an HI through a recurrent neural network. Qiu *et al*. [9] defined a fitness function with metrics of monotonicity, predictability, trendiness, and robustness to select a sensitive feature subset from 56 features and then constructed 1-D HI with weights optimized by the genetic algorithm. Tang *et al*. [10] developed the weighted linear combination of the correlation, monotonicity, and robustness to rank the sensitivity of 62

Manuscript received July 13, 2021; revised September 10, 2021; accepted October 14, 2021. Date of publication October 27, 2021; date of current version May 6, 2022. This work was supported by the National Natural Science Foundation of China under Grant 61833002. Paper no. TII-21-2942. *(Corresponding author: Wei Guo.)*

The authors are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: 576981262@qq.com; guo.w@uestc.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3123060.

Digital Object Identifier 10.1109/TII.2021.3123060

features and imported white Gaussian noise to further enhance their robustness. Yang *et al*. [11] used consistency, monotonicity, and gradualness for dynamically smoothing HI. Razavi-Far *et al*. [12] used the imputation-based oversampling technique for class-imbalanced data. Hassani *et al*. [13] used the accuracy and the distance between test samples and trained hyperplanes to enhance feature extraction of SVM.

In multidimensional features, root mean square (rms) and kurtosis are the most widely used features. Although kurtosis clearly indicates the increase in fluctuation severity, it reduces to the value like normal state even in the failure stage [14]. RMS can reflect the development of faults, but its spurious fluctuation has a great impact on prognosis. Ahmad *et al*. [15] developed a linear rectification technique to handle spurious fluctuations of rms. Pan *et al*. [16] used the linear rectification technology to smoothen the standardized rms and then set the relative rms to suppress sudden change of the rms. Cui *et al*. [17] defined the relative error based on the rms values of raw data and filtered data. However, the window width for smoothing the rms would influence the sensitivity to the initial defect [16], [18]. Antoni and Borghesani [19] defined a family of HIs to track cyclostationary or non-Gaussian symptoms independently.

Clustering algorithms and neural networks are also used to obtain the low-dimensional representation of high-dimensional features. Yu [20] integrated the negative log-likelihood probability with the locality preserving projection to assess bearing degradation level, and he also defined a local/nonlocal manifold regularization in the objective function of the Gaussian mixture model (GMM) to enhance its discrimination capability [21]. Rai and Upadhyay [14] extracted the signal energy values as features and then trained the k-medoids clustering model to determine two cluster centers, corresponding to the normal and failure states. Furthermore, they [22] constructed an HI by using the GMM and Jensen–Rényi divergence (JRD) to quantify the differentiation of defective posterior probability distributions (PDs) from healthy bearings. Several variants of nonnegative matrix factorization (NMF) [23], [24] also provide useful tools for feature extraction in statistical monitoring of non-Gaussian processes, and they preserve the geometric structure of data themselves to improve their clustering capability.

Singh *et al*. [25] first defined a cumulative JRD trend and proposed a confidence value (CV) to normalize the JRD to a specific interval, and then they [26] developed a similarity index and a *K*-means clustering to find prior experience from other bearings having similar degradation pattern to set an appropriate scaling parameter for the CV of the test bearing. Lei *et al*. [27] used the Spearman coefficient and a correlation clustering algorithm to remove redundant features and fused them into an HI using a self-organizing map (SOM). Yang *et al*. [28] used the equal weight combination of the monotonicity and trendability to select a sensitive feature subset and then formed an HI by inputting the SOM. Rai and Kim [29] used the trained probabilistic SOM and Gini–Simpson index to measure the inequality in the PD whose attribute corresponds to the bearing state. Pei *et al*. [30] utilized a continuous deep belief network to extract deep features and then trained the SOM to transform multidimensional features into 1-D HI. Chen *et al*. [31] used a convolution layer in the recurrent neural network to reduce the feature size.

However, many HIs cannot fully meet the above-mentioned criteria [22]. Meanwhile, the variation smoothness in the PD is not considered in some real cases, which makes the estimation of the GMM unstable. Motivated by the above reasons, we propose a novel integrated approach based on the GMM and quantification analysis to obtain an accurate PDA. The main contributions of this article are as follows.

1) An improved clustering algorithm, called the Hellinger distance-based regularized GMM (HRGMM), is presented to distinguish the bearing performances that vary with health states, defect severities, or working conditions. In this model, instead of measuring the similarity between feature samples, the Hellinger distance is introduced to directly measure the similarity between PDs, which then combines with the manifold regularization to improve the clustering performance of the GMM for multimode and nonlinear processes.
2) Based on the HRGMM, we develop a new HI, in which the JRD and tangent function work together to improve the quantification and standardization for bearing PDA. The constructed HI not only shows a monotonically decreasing and relatively smooth degradation trend but also eliminates the effects of unstable performance on the normal CV. It obtains higher scores on the metrics of monotonicity, correlation, and robustness. Moreover, the HI is limited to the range [0, 1], which ensures that the HIs of a group of bearings fall within the same range, and is beneficial to the prognosis in the long run.
3) Using the proposed HI, the degradation process can be divided into three or four health states (i.e., health, slight/severe degradation, and failure) by automatically identifying critical changes in the degradation rate. This makes the monitoring sensitive to incipient defects and state deterioration, and it is suitable for analyzing bearings with multiple failure modes.

The rest of this article is organized as follows. Section II briefly reviews a regularized GMM (RGMM). Section III introduces the integrated approach for bearing PDA. Section IV presents four cases to evaluate the results of the simulation and experiments, as well as for comparison purposes. Finally, Section V concludes this article.

II. Brief Introduction on RGMM

A degradation model based on GMM is to find appropriate low-dimensional representation for high-dimensional features and then quantify the bearing state. For one sample $\mathbf{x}_i$, it is a linear combination of K Gaussian components and is described as

$$P\left(\mathbf{x}_i \left| \Theta \right.\right) = \sum_{k=1}^{K} \pi_k p_k \left(\mathbf{x}_i \left| \theta_k \right.\right) \tag{1}$$

where $\Theta = (\pi_k, \ldots, \pi_K, \theta_1, \ldots, \theta_K)$ represents all parameters of the GMM, θ_k describes Gaussian density function p_k of the kth Gaussian component, i.e., $p_k(\mathbf{x}_i|\theta_k) \sim \mathrm{N}(\mathbf{x}_i|\mu_k, \Sigma_k)$, and π_k is its positive weight, satisfying $\sum_{k=1}^{K} \pi_k = 1$.

For given samples $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, Θ is optimized by the maximum-likelihood principle to make the probability $P(\mathbf{X}|\Theta)$ maximal, and the log-likelihood function is defined as

$$L(\Theta) = \log P(\mathbf{X}|\Theta) = \log \prod_{i=1}^{N} P(\mathbf{x}_i|\Theta)$$

$$= \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k N_k(\mathbf{x}_i|\mu_k, \Sigma_k) \right). \quad (2)$$

Introducing the latent variable $P(c|\mathbf{x}_i)$, the conditional probability contribution of $\mathbf{x}_i$ belonging to the component c, the complete log-likelihood function is defined as follows:

$$\sum_{i=1}^{N} \sum_{k=1}^{K} P(c|\mathbf{x}_i) (\log \pi_k + \log N_k(\mathbf{x}_i|\mu_k, \Sigma_k)). \quad (3)$$

The expectation-maximization (EM) algorithm is then used to find the solutions for the models and obtain the estimation for Θ. Further incorporating geometric knowledge of the discrete PD [32] for an ideal result, the following regularized log-likelihood function is defined:

$$Q = L(\Theta) - \lambda R \quad (4)$$

where λ is a regularization parameter, and the Laplacian regularization term R introduces the spectral graph theory into the objective function of GMM to preserve the manifold structure in the data and is then defined as

$$R = \sum_{i,j=1}^{n} D_{ij} w_{ij} \quad (5)$$

where D_{ij} represents the distance between two distributions $P_i(c)$ and $P_j(c)$, where $P_i(c) = P(c|\mathbf{x}_i)$. In a nearest neighbor graph, w_{ij} is an edge weight between the sample $\mathbf{x}_i$ and one of its p nearest neighbors $N_p(\mathbf{x}_i)$ satisfying

$$w_{ij} = \begin{cases} 1 \text{ if} \mathbf{x}_i \in N_p(\mathbf{x}_j) \text{ or} \mathbf{x}_j \in N_p(\mathbf{x}_i) \\ 0 \qquad\qquad \text{otherwise} \end{cases}. \quad (6)$$

The traditional GMM cannot obtain satisfactory results because it does not consider the case in which the data are supported on a submanifold of the ambient space. Liu *et al.* [32] incorporated the geometric knowledge of the PD into the learning of the GMM and demonstrated the possibility of improving the clustering performance of the GMM. Motivated by this idea, an improved RGMM can be developed by exploiting the nonlinear manifold structure and similarity measure between PDs to smooth the PD along geodesics and enhance the learning capability of the RGMM.

III. Methodology

In this section, an integrated approach is presented for the bearing PDA; its block diagram is shown in Fig. 1. After extracting multidomain features from raw vibration data, an improved GMM, i.e., the HRGMM, is developed to implement the data clustering. Then, a new HI based on the HRGMM is proposed to complete the information quantification and normalization to enable accurate assessment for bearing performance degradation.

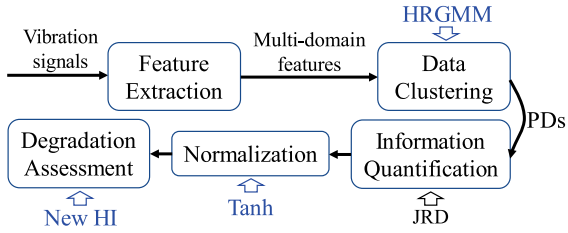

Fig. 1. Block diagram for the proposed clustering and HI construction.

A. Hellinger Distance

When monitoring the health status of a bearing, the probability density function (PDF) preserves the complete probabilistic information of its data; thus, comparing the PDF is more suitable than using only partial low-order or high-order moments. In general, the Euclidean distance and Kullback–Leibler (KL) divergence are used to measure the similarity or distance between two samples or distributions [32], [33]. However, the KL divergence is nonsymmetric. Moreover, its value ranges from 0 to $+\infty$, which makes its measurement difficult to explain [34]. Instead of using the KL divergence, the Hellinger distance, a symmetric and bounded metric, is introduced in this article.

For PDs $P = [p_i]_{i\in[n]}$ and $Q = [q_i]_{i\in[n]}$, supported on $[n]$, their Hellinger distance is defined as

$$h(P, Q) = 1/\sqrt{2} \cdot \left\| \sqrt{P} - \sqrt{Q} \right\|_2. \quad (7)$$

For all PDs, it satisfies that $0 \le h(P, Q) \le 1$, and is a symmetric metric, i.e., $h(P, Q) = h(Q, P)$, satisfying the triangle inequality and has nice properties [35].

B. Hellinger-Distance-Based Regularized GMM (HRGMM)

Using the Hellinger distance, the square of the distance between two distributions P_i and P_j is defined as follows:

$$h^2(P_i, P_j) = \frac{1}{2} \sum_{k=1}^{K} \left(\sqrt{P(k|\mathbf{x}_i)} - \sqrt{P(k|\mathbf{x}_j)} \right)^2 \quad (8)$$

where $P(k|\mathbf{x}_i)$ is the posterior probability that generates the sample $\mathbf{x}_i$ by the kth Gaussian component satisfying

$$\sum_{k=1}^{K} P(k|\mathbf{x}_i) = 1.$$

Refer to the manifold assumption [36] and the smoothness of conditional PDs [33], the regularization term (5) is

rewritten as

$$\mathcal{R} = \sum_{i,j=1}^{n}\left[\frac{1}{2}\sum_{k=1}^{K}\left(\sqrt{P(k|\mathbf{x}_i)}-\sqrt{P(k|\mathbf{x}_j)}\right)^2\right]w_{ij}$$
$$= \sum_{k=1}^{K}\left[\frac{1}{2}\sum_{i,j=1}^{n}\left(\sqrt{P(k|\mathbf{x}_i)}-\sqrt{P(k|\mathbf{x}_j)}\right)^2\right]w_{ij}$$

then its kth item is

$$\begin{aligned}\mathcal{R}_k &= \frac{1}{2}\sum_{i,j=1}^{n}\left(\sqrt{P(k|\mathbf{x}_i)}-\sqrt{P(k|\mathbf{x}_j)}\right)^2 w_{ij}\\ &= \frac{1}{2}\sum_{i,j=1}^{n}\left(P(k|\mathbf{x}_i)+P(k|\mathbf{x}_j)\right)w_{ij}\\ &\quad -\sum_{i,j=1}^{n}\left(\sqrt{P(k|\mathbf{x}_i)}\cdot\sqrt{P(k|\mathbf{x}_j)}\right)w_{ij}\\ &= \sum_{i=1}^{n}\left(\sqrt{P(k|\mathbf{x}_i)}\right)^2 d_{ii}\\ &\quad -\sum_{i,j=1}^{n}\left(\sqrt{P(k|\mathbf{x}_i)}\cdot\sqrt{P(k|\mathbf{x}_j)}\right)w_{ij}\\ &= h_k^T\mathbf{D}h_k - h_k^T\mathbf{W}h_k = h_k^T\mathbf{L}h_k \end{aligned} \tag{9}$$

where the Laplacian matrix $\mathbf{L}=\mathbf{D}-\mathbf{W}$; since the weight matrix $\mathbf{W}$ is symmetric, $\mathbf{D}$ is then a diagonal matrix that satisfies $d_{ii}=\sum_{j=1}^{n} w_{ij}$; and $h_k=(\sqrt{P(k|\mathbf{x}_1)},\cdots,\sqrt{P(k|\mathbf{x}_n)})^T$.

Like the GMM, the latent sample set $\mathbf{Z}=(z_1,\ldots,z_m)$ is used so that the generalized EM algorithm can estimate the actual sample set $\mathbf{X}$. For the complete dataset $\{\mathbf{X},\mathbf{Z}\}$, the regularized likelihood function is rewritten as

$$\begin{aligned}\mathcal{Q} &= \log\left(\mathcal{N}\left(\mathbf{X},\mathbf{Z}|\,\Theta\right)\right)-\lambda\sum_{k=1}^{K}\mathcal{R}_k\\ &= \sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\pi_k\mathcal{N}_k(\mathbf{x}_i|\,\mu_k,\Sigma_k)\right)-\lambda\sum_{k=1}^{K}\mathcal{R}_k.\end{aligned} \tag{10}$$

By maximizing the log-likelihood function and minimizing the regularization term in (10), the incomplete samples, i.e., the Gaussian clusters of actual samples $\mathbf{X}$, can be estimated. The estimation procedure is described here.

1) E-Step: at the tth iteration, using the current parameter values Θ^{t-1} ($t>1$) obtained in the previous iteration or the initialized values Θ^0 ($t=1$) obtained by the K-means algorithm, the posterior probabilities $P(k|\mathbf{x}_i,\Theta^{t-1})$ ($k=1,\ldots,K$) are estimated as

$$P\left(k|\,\mathbf{x}_i,\Theta^{t-1}\right)=\frac{\pi_k^{t-1}\mathcal{N}_k\left(\mathbf{x}_i|\,\mu_k^{t-1},\Sigma_k^{t-1}\right)}{\sum_{k=1}^{K}\pi_k^{t-1}\mathcal{N}_k\left(\mathbf{x}_i|\,\mu_k^{t-1},\Sigma_k^{t-1}\right)} \tag{11}$$

and the corresponding log-likelihood function is rewritten as

$$\begin{aligned}\mathcal{Q}\left(\Theta,\Theta^{t-1}\right) &= E\left[\log\left(\mathcal{N}\left(\mathbf{X},\mathbf{Z}|\,\Theta\right)\right)\big|\mathbf{X},\Theta^{t-1}\right]\\ &= \sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\mathcal{N}\left(\mathbf{x}_i|\,\mu_k,\Sigma_k\right)\right)\\ &\quad\times P\left(k|\,\mathbf{x}_i,\Theta^{t-1}\right)\\ &= \sum_{i=1}^{n}\sum_{k=1}^{K}\log\left(\pi_k\right)P\left(k|\,\mathbf{x}_i,\Theta^{t-1}\right)\\ &\quad+\sum_{i=1}^{n}\sum_{k=1}^{K}\log\left(\mathcal{N}\left(\mathbf{x}_i|\,\mu_k,\Sigma_k\right)\right)\\ &\quad\times P\left(k|\,\mathbf{x}_i,\Theta^{t-1}\right).\end{aligned} \tag{12}$$

Further incorporating the regularization term, the object is

$$\mathcal{Q}\left(\Theta,\Theta^{t-1}\right)=\mathcal{Q}\left(\Theta,\Theta^{t-1}\right)-\lambda\sum_{k=1}^{K}R_k \tag{13}$$

that is to say, the optimization is to maximize $\mathcal{Q}(\Theta,\Theta^{t-1})$ and minimize $\sum_{k=1}^{K}R_k$ separately to find $\mathcal{Q}$.

2) M-Step: using the Newton–Raphson updates, the posterior probabilities

$$\begin{aligned}\sqrt{P\left(k|\,\mathbf{x}_i\right)} &\leftarrow (1-\gamma)\sqrt{P\left(k|\,\mathbf{x}_i\right)}\\ &+\gamma\frac{\sum_{j=1}^{n}w_{ij}\sqrt{P\left(k|\,\mathbf{x}_j\right)}}{\sum_{j=1}^{n}w_{ij}}\end{aligned} \tag{14}$$

where γ is the step parameter and satisfies $0\leq\gamma\leq1$. After that, the Gaussian parameters Θ^t is updated as follows:

$$\pi_k^t=\frac{1}{n}\sum_{i=1}^{n}P\left(k|\,\mathbf{x}_i\right) \tag{15}$$

$$\mu_k^t=\frac{\sum_{i=1}^{n}\mathbf{x}_iP\left(k|\,\mathbf{x}_i\right)}{\sum_{i=1}^{n}P\left(k|\,\mathbf{x}_i\right)} \tag{16}$$

$$\Sigma_k^t=\frac{\sum_{i=1}^{n}P\left(k|\,\mathbf{x}_i\right)\left(\mathbf{x}_i-\mu_k^t\right)\left(\mathbf{x}_i-\mu_k^t\right)^T}{\sum_{i=1}^{n}P\left(k|\,\mathbf{x}_i\right)}. \tag{17}$$

Completing the above two steps, the current value of the log-likelihood function is calculated according to the following equation:

$$Q=\sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\pi_kN_k(\mathbf{x}_i|\,\mu_k,\Sigma_k)\right)-\lambda\sum_{k=1}^{K}\mathcal{R}_k. \tag{18}$$

By continuously decreasing the step parameter γ, the regularization term $\mathcal{R}_k$ gradually decreases until its value can no longer decrease. As a result, the conditional probability function is smooth and the maximization of the log-likelihood function can be achieved by updating the Gaussian parameters. At the end of the iteration, the RGMM with the corresponding parameters Θ^t turn all samples into K clusters. The proposed HRGMM algorithm is summarized in Table I. The optimal

TABLE I
HELLINGER DISTANCE-BASED REGULARIZED GMM (HRGMM) ALGORITHM

Input: n samples $X = (x_1, x_2, \ldots, x_n)$, the number K for Gaussian components, the number p for the nearest neighbors, the regularization parameter λ, a small value δ for the termination condition.
Output: parameters Θ^t of HRGMM and the posterior probabilities $P(k\|x_i, \Theta^t)$ $(k = 1, \ldots, K)$.
Step:
0: Set the initial value of the step parameter $\gamma = 0.9$.
1: Construct a nearest neighbor graph with weight matrix W.
2: Initialize the Gaussian parameters Θ^0 by using the K-means algorithm.
3: Initialize the sequence number t for the iteration, $t \leftarrow 1$.
4: While (true)
5: E-step: the posterior probabilities are computed by using (11).
6: M-step: the posterior probabilities until convergence is smoothed by using (14).
7: The parameters of the HGMM are updated by using (15), (16), and (17).
8: The regularized log-likelihood value is evaluated by using (18).
9: If $\mathcal{L}(\Theta^t) < \mathcal{L}(\Theta^{t-1})$, $\gamma \leftarrow 0.9\gamma$, go to Step 6;
10: If $\mathcal{L}(\Theta^t) - \mathcal{L}(\Theta^{t-1}) \le \delta$, break,
11: $t \leftarrow t+1$;
12: Use the HRGMM with the parameters Θ^t to obtain K clusters.

number of Gaussian components for one dataset is estimated by the Bayesian information criterion (BIC) [37].

C. New HI Based on HRGMM

For a discrete PD $\mathbf{p} = (p_1, p_2, \ldots, p_K)$, satisfying $\sum_{i=1}^{K} p_i = 1$, its tails are easily affected by the bearing defect, and yet the Shannon entropy is insensitive to such changes [22]. Consequently, the Rényi entropy is used to quantify the information content in the PD and is expressed as

$$R_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \sum_{i=1}^{K} p_i^\alpha, \alpha > 0 \text{ and } \alpha \neq 1 \tag{19}$$

where the exponent α manages the sensitivity of the entropy toward changes in shape of the PD and its value is suggested to be 0.5 [22].

Supplying new feature samples $\mathbf{X}$ from one bearing over the lifetime to the HRGMM, the divergence serves to express the dissimilarities between the PDs. Let $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ be n PDs on X, and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ be a weight vector satisfying $b_j \geq 0$ and $\sum_{j=1}^{n} b_j = 1$. Then, the JRD [38] is

$$\mathrm{JRD}_\alpha^{\mathbf{b}}[\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n] = R_\alpha\left(\sum_{j=1}^{n} b_j \mathbf{p}_j\right) - \sum_{j=1}^{n} b_j R_\alpha(\mathbf{p}_j) \tag{20}$$

where uniform weights $(1/n)$ is set in this article. For the bearing, its health condition $\mathbf{p}_0$ is used as a reference. If the bearing is healthy, there is no clear difference between $\mathbf{p}_0$ and a testing PD. Once the fault exists, its changes in the distribution would alter the value of JRD, which is the purpose of applying JRD [39]. Setting uniform weights can make the JRD more sensitive to the change in the defect severity, while setting nonuniform weights may smooth this change [25], which is illustrated later.

To standardize the quantification in the PDA, a modified CV is defined by introducing the function tanh. As one of popular activation functions in neural networks, this function has smooth, asymptotic, and monotonic properties. Here, it is used to normalize the JRD value and is expressed as

$$\tanh(\mathrm{JRD}) = \frac{e^{(\mathrm{JRD})} - e^{-(\mathrm{JRD})}}{e^{(\mathrm{JRD})} + e^{-(\mathrm{JRD})}}. \tag{21}$$

Then, the CV is defined by

$$\mathrm{CV} = 1 - \tanh^2(\mathrm{JRD}). \tag{22}$$

The CV value of a bearing ranges from 0 to unity, where "1" indicates a state of perfect health and "0" indicates its failure. The successive decline of the CV implies that the health condition will change and the bearing will continue deteriorate. The rate of CV decline also reflects the progression of the bearing defect, which can be used for health state division. The modified CV is not affected by the scaling parameter specified based on prior experience or similar bearings [25], [26].

IV. EXPERIMENTAL RESULTS AND COMPARISONS

In this section, four cases are used to validate and compare the performance of the proposed HRGMM and HI: Case 1 for clustering performance, and Cases 2–4 for the PDA description of the HI. Case 2 uses simulated datasets, whereas Cases 3 and 4 use experimental datasets with different defects.

A. Case 1: Verification of the HRGMM

The HRGMM is compared with the following five popular clustering algorithms:

1) K-means clustering;
2) original GMM;
3) PCA-GMM in which multifeatures are processed by PCA and only most variance information (e.g., 75%) [20] is reserved as the input of the GMM;
4) locally consistent GMM (LCGMM) [32], in which the KL divergence measures the similarity between distributions for comparison with the HRGMM using the Hellinger distance;
5) generalized nonnegative matrix factorization (GNMF) [23], i.e., NMF with a regularization term to obtain a better feature representation that maintains the local geometry of data.

Fig. 2 shows the clustering performances of the HRGMM and the five other algorithms for the benchmark of two half-moons data. Only the proposed HRGMM can clearly separate two clusters, whereas other methods present the confusion of two clusters in different degrees.

Furthermore, two metrics, i.e., accuracy (AC) and normalized mutual information (NMI) [33], are calculated to evaluate the clustering performances of these six methods

$$\mathrm{AC} = \frac{\sum_{i=1}^{N} \delta\left(s_i, \mathrm{map}\left(r_i\right)\right)}{N} \tag{23}$$

where N is the total number of samples, and

$$\delta\left(s_i, \mathrm{map}\left(r_i\right)\right) = \begin{cases} 1 \text{ if } s_i = \mathrm{map}\left(r_i\right) \\ 0 \quad \text{others} \end{cases} \tag{24}$$

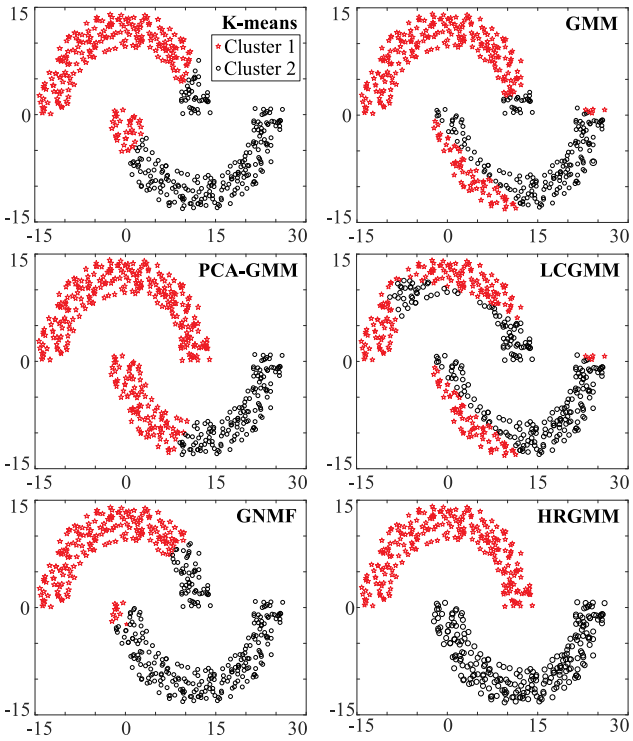

Fig. 2. Comparison of six clustering methods for half-moons data.

where s_i and r_i are real and identified labels, respectively, and map(r_i) is the permutation mapping function to map the label r_i to the equivalent label s_i using the Kuhn-Munkres algorithm

$$\mathrm{NMI}(C, C') = \sum_{c_i \in C, c'_j \in C'} p(c_i, c'_j) \cdot \log_2 \left(\frac{p(c_i, c'_j)}{p(c_i) \cdot p(c'_j)} \right) \tag{25}$$

where C and C' denote the sets of real and obtained clusters, respectively; and $p(c_i)$ and $p(c'_j)$ are probabilities that an arbitrarily selected sample belongs to the clusters c and c', and $p(c_i, c'_j)$ is the joint probability that the sample belongs to two clusters.

The first three datasets come from experiments conducted by Lessmeier *et al.* [40] from Paderborn University. The test rig is shown in Fig. 3(a). Vibration data from ball bearing 6203 were accumulated at a sampling rate of 60 kHz and each collection lasts 4 s. Each bearing rotates at 900 r/min with a load torque of 0.7 N·m and a radial force of 1000 N.

The last four experiments were conducted on the wheelset bearing test rig, mainly including an axle and its two supporting bearings shown in Fig. 3(b), which is used for a train vested in the Qingdao Sifang Institute. Vibration data were collected by the accelerometer mounted on the bearing at 12 O'clock. Artificial damages were generated on bearing components. Bearing data under different vertical loads and train speeds were obtained at the sampling frequency of 12.8 kHz.

Raw vibration data are used to extract 11 features, including six statistical features in time domain (rms, kurtosis, skewness, crest factor, peak-to-peak, and variance), and three features in frequency domain (frequency center, rms frequency, and standard deviation frequency). Moreover, a wavelet "db5" from Daubechies family wavelets in five levels was used to process data and the first two-level wavelet energy indexes are chosen as the last two features. The number p of the nearest neighbors and regularization parameter λ are set to 2 and 0.1, respectively.

Each result was obtained by repeating the experiment 20 times. Tables II and III show the experimental settings and evaluation results on data clustering. The proposed HRGMM show the best performance on all experimental datasets. It is difficult to rank the clustering performance of the other five methods.

In Experiment 1, for bearings with the same damage extent and three health states, the HRGMM successfully group

TABLE II
COMPARISONS OF SIX CLUSTERING METHODS FOR BEARINGS [40] IN ALTS

Experiment	Method	NMI	AC
Experiment 1: Bearing state/code: Health (K001), OR defect (KA04), IR defect (KI17) Damage extent: 1	K-means	0.64	0.75
	GMM	0.56	0.75
	PCA-GMM	0.64	0.84
	LCGMM	0.79	0.91
	GNMF	0.63	0.31
	HRGMM	0.81	0.92
Experiment 2: Bearing code: KB23, KB24, and KB27 Defect component: OR+IR Damage extents: 1, 2 and 3	K-means	0.40	0.58
	GMM	0.37	0.57
	PCA-GMM	0.40	0.57
	LCGMM	0.39	0.59
	GNMF	0.38	0.46
	HRGMM	0.43	0.67
Experiment 3: Bearing code: KI16, KI18, KI21 Defect component: IR Damage extents: 1, 2 and 3	K-means	0.36	0.58
	GMM	0.61	0.83
	PCA-GMM	0.60	0.82
	LCGMM	0.60	0.83
	GNMF	0.21	0.54
	HRGMM	0.64	0.84

TABLE III
COMPARISONS OF SEVEN CLUSTERING METHODS FOR WHEELSET BEARINGS

Experiment	Method	NMI	AC
Experiment 4: Defect component: OR, IR, and ball Train speed = 90km/h, Vertical load = 56kN	K-means	0.30	0.53
	GMM	0.29	0.51
	PCA-GMM	0.58	0.50
	LCGMM	0.58	0.67
	GNMF	0.58	0.50
	HRGMM	0.64	0.81
Experiment 5: Defect component: OR, IR, and ball Train speed = 120km/h, Vertical load = 146kN	K-means	0.50	0.66
	GMM	0.51	0.67
	PCA-GMM	0.50	0.67
	LCGMM	0.51	0.68
	GNMF	0.58	0.49
	HRGMM	0.71	0.83
Experiment 6: Defect component: IR Train speeds: 60, 90, and 120km/h, Vertical load = 56kN	K-means	0.46	0.66
	GMM	0.36	0.66
	PCA-GMM	0.59	0.79
	LCGMM	0.58	0.68
	GNMF	0.39	0.67
	HRGMM	0.70	0.82
Experiment 7: Defect component: Ball Train speeds: 60, 90, and 120km/h Vertical load = 56kN	K-means	0.50	0.68
	GMM	0.50	0.67
	PCA-GMM	0.58	0.50
	LCGMM	0.66	0.82
	GNMF	0.50	0.63
	HRGMM	0.71	0.84

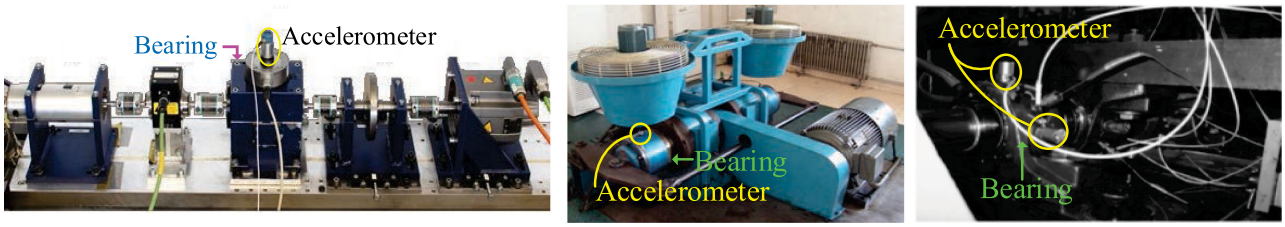

Fig. 3. Experimental test rigs for collecting bearing vibration data. (a) Test rig for Experiments 1–3 in Case 1 [40]. (b) Test rig for Experiments 4–6 in Case 1. (c) Test rig for Cases 3 and 4 [42].

them into three clusters with an accuracy of 0.92. Although the combination of the outer race (OR) and inner race (IR) defects is not easily detected, different damage extents were identified in Experiments 2 and 3 with accuracies of 0.67 and 0.84, respectively. It indicates that the HRGMM can accurately identify the bearing state.

For wheelset bearings, the HRGMM also outperforms the other five methods. In Experiment 4, the accuracy is increased by $(0.81-0.53)\times 100\%/0.81 = 35\%, 37\%, 38\%, 17\%$, and 38% compared with the *K*-means, GMM, PCA-GMM, LCGMM, and GNMF algorithms, respectively. Improvements are also observed in Experiments 5, 6, and 7, in which varying speeds and loads are considered. The above results based on raw data processing indicate that the HRGMM is promising for the applications of real bearings. Hence, the proposed HI based on this model is further used to assess the bearing degradation.

B. Case 2: Simulated Degradation

Time-domain signals are generated to emulate the bearing degradation process and the expressions are shown as follows: Unnumber equation shown at the bottom of this page, where y_s represents the signal due to the shaft rotation, s_n is the noise, and y_1, y_2, and y_3 mimic the impulses from a faulty bearing with increasing severity.

In the simulation, every 2048 data points form one data file; the complete signal consisted of 60 data files. As shown in Fig. 4(a), Stages I, II, III, and IV were used to mimic the health, early fault, and increasing severity. Three features in time domain, including the rms, kurtosis, and skewness, were chosen to form one feature sample. The number of Gaussian components was six during training. Only health data were used to construct the HRGMM model. The CV curve using the proposed HI is shown in Fig. 4(b).

Cases 2–4 demonstrate how the proposed HI improves the bearing PDA by comparison with the following three methods.

1) *RMS and kurtosis:* The former reflects the energy increase, the latter monitors the incipient defect, and the 3σ interval [41] distinguishes between the normal and abnormal states.
2) *GMM-JRD:* A method based on feature extraction with preprocessing by ensemble empirical mode decomposition, GMM, and JRD with a small scaling parameter [22].
3) *NLLPEWMA:* Negative log-likelihood probability-based exponentially weighted moving average statistic. The threshold is determined by the kernel density estimation for the beginning of slight degradation [20].

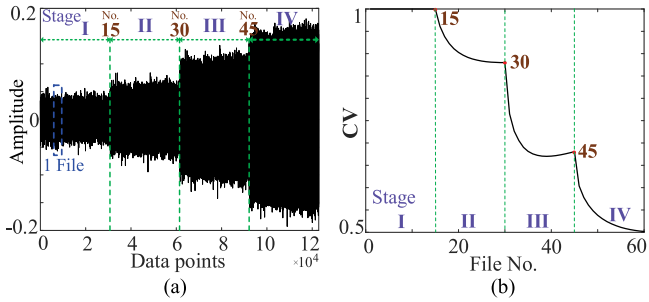

Fig. 4. Simulated bearing degradation in four stages: signals in (a) and the degradation trend described by the proposed HI in (b). (a) Simulated signals. (b) Degradation trend.

These methods were applied to simulated datasets, and the results are shown in Fig. 5. The comparison of Fig. 5(b) with Figs. 4(b) and 5(a) shows that the proposed HI and rms exhibit monotonic trends and are sensitive to the state change. The fluctuation of kurtosis makes it less identifiable at the beginning of Stage II. The result of the GMM-JRD in Fig. 5(c) was unexpected. For the NLLPEWMA in Fig. 5(d), the trend is also insensitive to state changes.

C. Case 3: Bearing 1 in Test 2

Experimental datasets in Cases 3 and 4 were provided by Qiu *et al.* [42] from the University of Cincinnati. Four double row bearings were installed on one shaft of the test rig shown

$$\begin{cases} y_h = y_s + s_n = 0.01 \cdot \sin(2\pi \cdot 30 t_{\text{time}}) + s_n \\ y_{f1} = y_h + y_1 = y_h + 0.025 \cdot \exp(0.01 \cdot t_{\text{time}}) \times \sin(2\pi \cdot 50 t_{\text{time}} + 2) \\ y_{f2} = y_h + y_2 = y_h + 0.055 \cdot \exp(0.02 \cdot t_{\text{time}}) \times \sin(2\pi \cdot 50 t_{\text{time}} + 2) \\ y_{f3} = y_h + y_3 = y_h + 0.105 \cdot \exp(0.035 \cdot t_{\text{time}}) \times \sin(2\pi \cdot 50 t_{\text{time}} + 2) \end{cases}$$

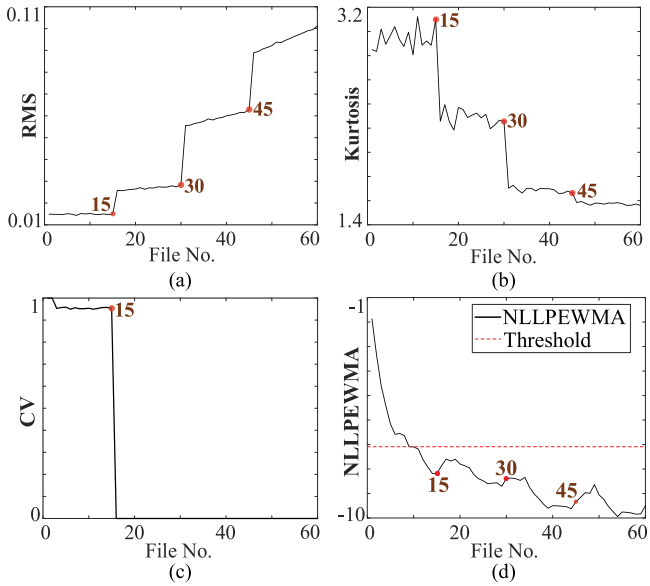

Fig. 5. Degradation trends described by four HIs for simulated datasets. (a) RMS. (b) Kurtosis. (c) GMM-JRD. (d) NLLPEWMA.

in Fig. 3(c) and performed run-to-failure tests under constant rotation speed of 2000 r/min and a radial load of 6000 lbs. Each collection sets the sampling rate of 20 kHz and obtained 20480 data points. The failures occurred after exceeding their designed lifetime.

Two additional HIs, i.e., relative rms (RRMS) [16] and SOM-MQE [27], [43], were compared with the proposed HI. RRMS is obtained by standardization and linear rectification to reduce stochastic fluctuations of rms. SOM-MQE stands for the SOM to fuse multiple features into a HI named minimum quantization error (MQE).

In Case 3, 984 data files from Bearing 1 were analyzed. At the end of the experiment, the OR defect was identified. In theory, the characteristic defect frequency (CDF) f_{OR} of this bearing is 236.4 Hz. From each vibration dataset, 11 features mentioned in Section IV-A are extracted. The first 50 samples are used as the health reference of the model.

The BIC [37] is used to provide a reliable estimation to select the appropriate number K of Gaussian components for the analyzed dataset. The minimum BIC signifies the optimal K value, maximizing the variability explained with the fewest possible clusters [37], [42]. As shown in Fig. 6(a), the optimal number of Gaussian components in this case is 4.

As for the weights in (20), their selection has a significant influence on the behavior of the JRD. The weight b_j is calculated by using the following exponential function [44]:

$$b_j = C \cdot \exp\left(-\frac{\eta}{n} j\right) \tag{26}$$

where η is the sensitivity parameter, and C is the normalizing constant. The degradation curves when setting uniform weights in Figs. 7(a) and 11(a) are compared with some examples of setting nonuniform weights shown in Fig. 8(a) and (b).

The experimental results and observations are analyzed as follows.

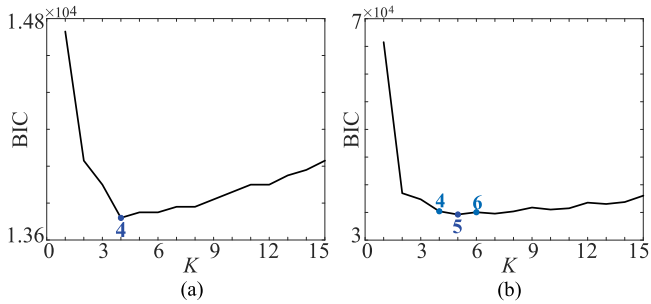

Fig. 6. BIC curves for K Gaussian components in Cases 3 and 4.

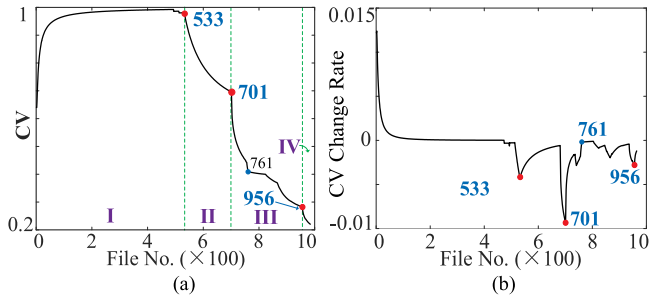

Fig. 7. CVs using the proposed HI to describe the degradation of Bearing 1. (a) CVs. (b) Change rate of CV.

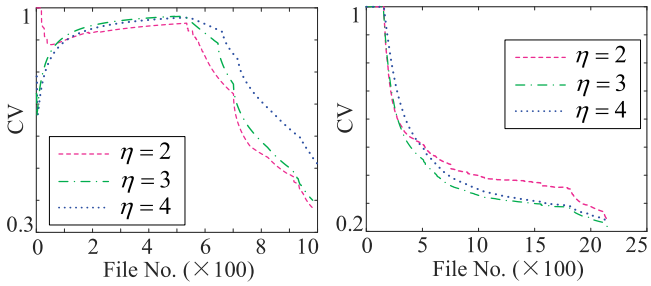

Fig. 8. Examples for setting nonuniform weights to JRD in Cases 3 and 4.

1) Identification of the Incipient Defect: In Fig. 7(a), a sharp drop occurred at File 533 indicates the change in the bearing state. In Fig. 9(a), no CDF of Bearing 1 is detected at File 532, whereas the identified f_{OR} of 230.7 Hz and its harmonics are clearly observed in the low-frequency range of the envelope spectrum of Files 533, as shown in Fig. 9(b). Therefore, the incipient OR defect is confirmed at the time corresponding to File 533.

Using the rms, RRMS, and SOM-MQE, a less obvious change occurs at File 533 in Fig. 10(a), (e), and (f), whereas the kurtosis in Fig. 10(b) does not reach the threshold until File 647, corresponding to the identified time for incipient defect, which is much later than the real one. RRMS and SOM-MQE show similar trends as the rms, but RRMS fractionally smoothens its fluctuations.

When using the GMM-JRD in Fig. 10(c) and the NLLPEWMA in Fig. 10(d), the identified defects occurred at

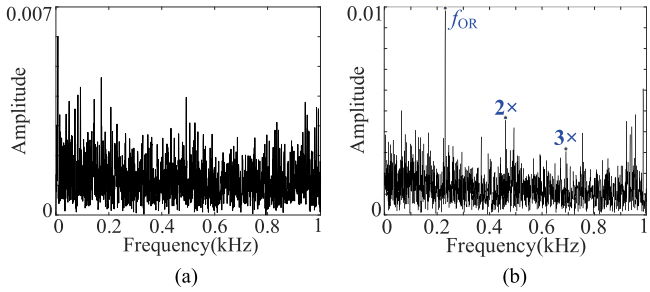

Fig. 9. Envelop spectra in 0–1 kHz of two data files for Bearing 1 in Case 3. (a) File 532. (b) File 533.

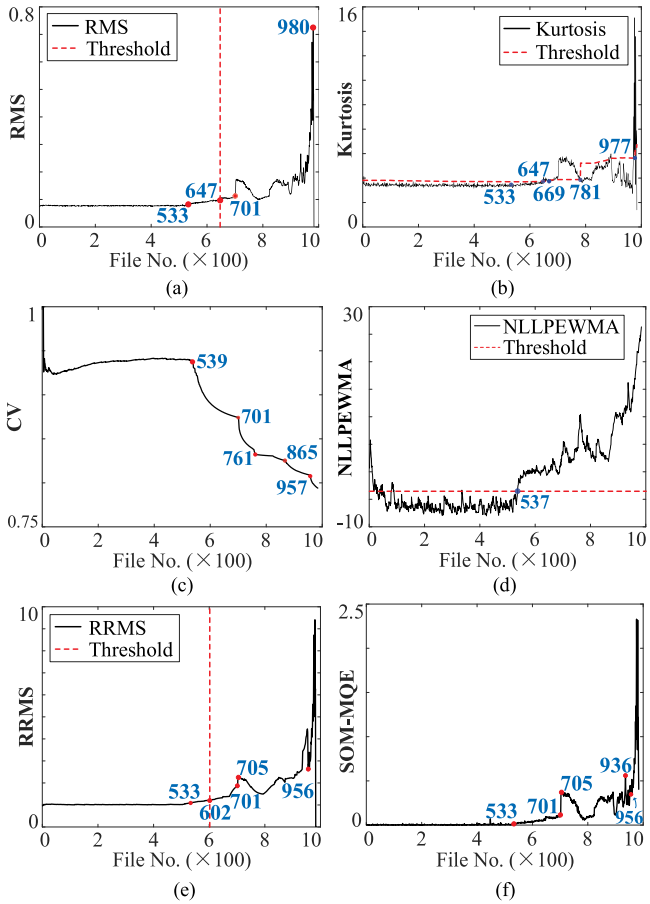

Fig. 10. Degradation trends described by six HIs in Case 3. (a) RMS. (b) Kurtosis. (c) GMM-JRD. (d) NLLPEWMA. (e) RRMS. (f) SOM-MQE.

Files 539 and 537, respectively, which are slightly closer to the real case and the result shown in Fig. 7(a).

2) Division of Bearing Stages: Compared with the HIs in Fig. 10, the proposed HI based on HRGMM (HRGMM-HI) in Fig. 7(a) provides a clearer quantification of the bearing degradation process from health to failure. The CV change rate is shown in Fig. 7(b). Three larger changes are automatically identified at Files 533, 701, and 956, where the decreasing rate changes to an increasing rate. These represent the occurrence of rapid deterioration and are identified as boundaries between adjacent health stages, whereas the deceleration at File 761

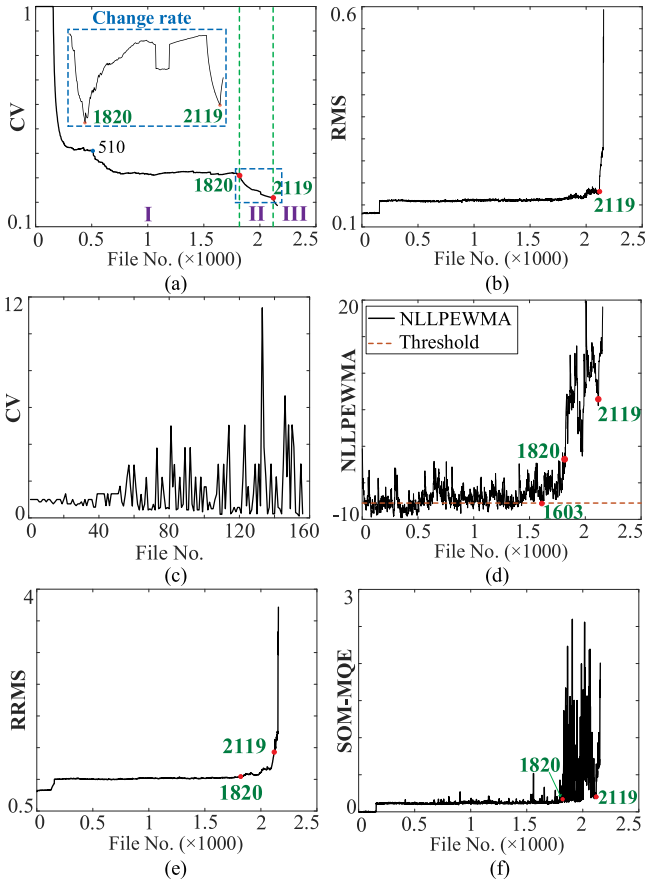

Fig. 11. Degradation trends described by six HIs in Case 4. (a) HRGMM-HI. (b) RMS. (c) GMM-JRD. (d) NLLPEWMA. (e) RRMS. (f) SOM-MQE.

indicates a slow evolving state and it is not a critical change for bearing deterioration.

In Fig. 7(a), after a short break-in period, the CV curve is flat before File 533; thus, the bearing is designated as healthy. Starting from File 533, the CV suddenly drops; this is viewed as the first change from health to slight degradation, which is validated by the results in Fig. 9. Afterward, the CV curve gradually declines until the second change at File 701, which is shown in Fig. 10(a), (c), (e), and (f). Therefore, the period from Files 533 to 701 is the slight degradation stage, which is very important in the bearing life. At this stage, despite the occurrence of bearing defects, the bearing or machine being setup continues to function properly; hence, there is sufficient buffer time to perform effective maintenance and logistical scheduling to extend the bearing life. In other words, this is an early warning for the bearing operation.

An obvious drop occurs at File 701. The CV decreases significantly from 701 to 761, which demonstrates that the bearing is going from bad to worse. Although the decrease of the CV from 761 slows down, it is evident that the bearing performance continues to deteriorate. At the last change at File 956, the CV begins to decline dramatically until the end of lifetime. Therefore, two change points at Files 701 and 956 are viewed as the beginning of the severe degradation and failure stages of the

bearing, which is also clear in the change rate shown in Fig. 7(b). The severe degradation stage is fairly short, approximately 26% of the bearing life; an urgent alarm must be set before the failure. Therefore, the degradation process of Bearing 1 is automatically divided into four stages: I—health, II—slight degradation, III—severe degradation, and IV—failure.

In addition, as shown in Fig. 8(a), with the increase of the sensitivity parameter in (26), abrupt changes of the CV are more difficult to be found, and the use of nonuniform weights makes the HI less sensitivity to the bearing state changes.

Although the GMM-JRD shows a similar trend, its parameter settings may lead to unstable performance, e.g., the results in Figs. 5(c) and 11(c). For other HIs, there are no clear boundaries to distinguish bearing stages for prognosis.

In [2], the kurtosis with signal preprocessing shows a sharp increase at File 703 but a minor change at File 533 and a nonmonotonic trend after 703.

D. Case 4: Bearing 3 in Test 1

In this case, the experimental data from Bearing 3 [42] were analyzed. The number of data files was 2156. At the end of the experiment, the IR defect occurred in this bearing, with a theoretically CDF f_{IR} of 297 Hz. Eleven features mentioned above are extracted from raw datasets and uniform weights are also set for the JRD. According to the BIC curve in Fig. 6(b), the optimal number of Gaussian components is 5. After individually applying the proposed HI and other HIs to these datasets, their degradation assessment results are shown in Fig. 11.

The experimental results reveal some interesting points.

1) For this bearing, the proposed HI in Fig. 11(a), rms in Fig. 11(b), NLLPEWMA in Fig. 11(d), and RRMS in Fig. 11(e), show monotonic trends with the increasing defect severity of the bearing. The GMM-JRD in Fig. 11(c) does not show stable performance in describing the degradation process because the scaling parameter close to zero leads to infinite CV values. SOM-MQE does not show a monotonic trend but can assist in identifying an abnormal state.
2) The rms is at a low and stable level until a sudden jump occurs at File 2119; the RRMS eliminates this obvious change.

For the NLLPEWMA, its degradation curve fluctuates around the threshold before File 1603, and then jumps to another level after File 1820. The SOM-MQE in Fig. 11(f) exhibits erratic fluctuations after around File 1820. However, none of these provides a clear indication of incipient defects and health stages.

With the HRGMM-HI, its degradation curve shows a slow change after the break-in period and remained relatively stable between Files 510 and 1820. After File 1820, an obvious drop indicates the state change from health to degradation. Another sharp drop occurs at File 2119. Both can be presented in the change rate shown in the subgraph of Fig. 11(a) and can be used for stage division. Therefore, the bearing lifetime is divided into three stages: health, degradation, and failure.

3) Envelope spectral analysis is applied to two data files to detect the incipient defect of Bearing 3; the results are shown in Fig. 12. A small but clear impulse can be found at 294.2Hz at File 1820. A much more obvious impulse at the same frequency is also found in File 1857, where the identified frequency is close to the theoretical frequency of 297 Hz.

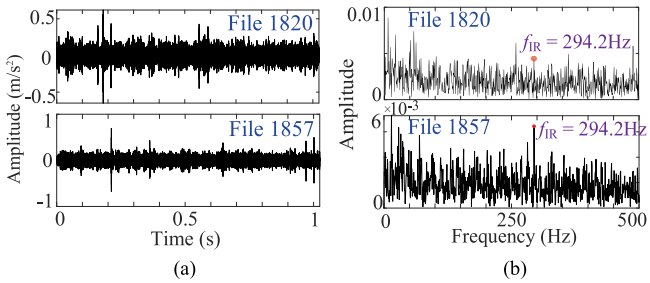

Fig. 12. Temporal and spectral waveforms of Files 1820 and 1857 for Bearing 3 in Case 4. (a) Temporal waveforms. (b) Envelope spectra in 0–500 Hz.

For the same bearings in Cases 3 and 4, Cheng *et al.* [45] identified their incipient defects at Files 534 and 1810, which are close to the above results, while their trends are nonmonotonic.

E. Evaluation of HIs

Lei *et al.* [46] mentioned that the HI has a great influence on the complexity and accuracy of prognostics. There are many metrics to evaluate and compare the HI. In this article, three popular metrics, including monotonicity "Mon"[46], correlation "Corr"[47], and robustness "Rob"[48], are expressed as follows:

$$\mathrm{Mon}(\mathbf{s}) = \frac{1}{K}\left|\mathrm{No.of}\, d/ds > 0 - \mathrm{No.of}\, d/ds < 0\right| \tag{27}$$

$$\mathrm{Corr}(\mathbf{s}) = \frac{\left|\sum_{i=1}^{L} s_i - \bar{s}\left(i - \sum_{i=1}^{L} i/L\right)\right|}{\sqrt{\sum_{i=1}^{L}(s_i-\bar{s})^2 \sum_{i=1}^{L}\left(i - \sum_{i=1}^{L} i/L\right)^2}} \tag{28}$$

$$\mathrm{Rob}(\mathbf{s}(t_i)) = \exp\left(\frac{-\mathrm{std}\left(\mathbf{s}_R(t_i)\right)}{\left|\mathbf{s}_T(\mathrm{start}) - \mathbf{s}_T(\mathrm{end})\right|}\right) \tag{29}$$

where $\mathbf{s} = \{s_i\}_{i=1:L}$ is L HI values with s_i representing the HI value at time t_i; meanwhile, it can be decomposed into a trend part $\mathbf{s}_T(t_i)$ and a random part $\mathbf{s}_R(t_i)$, describing the certainty and uncertainty of degradation, respectively; $d/ds = s_{i+1} - s_i$ denotes the difference of two adjacent HI values; No. of $d/ds > 0$ and No. of $d/ds < 0$ represent the numbers of positive differences and negative differences, respectively; and $\bar{s}$ denotes the mean value of all HI values. Each metric range is [0, 1], and a larger score means a better performance of HI. The evaluation results of the proposed HI and the other HIs are shown in Table IV.

Monotonicity measures whether the degradation process shows a monotonic increasing/decreasing trend. For the HI, its fluctuation would decrease its monotonicity score [46]. The correlation represents the degree of linear correlation between the HI and the sampling time t. The fluctuation of the HI weakens its linear correlation with time. Robustness denotes the tolerance of $\mathbf{s}(t_k)$ to the exception value.

Main observations for above comparisons are as follows.

TABLE IV
EVALUATION METRICS ON SIX HIS FOR CASES 3 AND 4

Case	Method	Monotonicity	Correlation	Robustness
3	RMS	0.125	0.623	0.817
	NLLPEWMA	0.036	0.813	0.946
	GMM-JRD	0.111	0.857	0.976
	RRMS	0.231	0.682	0.815
	SOM-MQE	0.134	0.626	0.863
	HRGMM-HI	0.239	0.870	0.973
4	RMS	0.065	0.466	0.991
	NLLPEWMA	0.026	0.706	0.953
	GMM-JRD	NAN	NAN	NAN
	RRMS	0.141	0.501	0.993
	SOM-MQE	0.033	0.622	0.924
	HRGMM-HI	0.245	0.748	0.992

1) The proposed HI has higher scores on three metrics in two cases, indicating its superiority for bearing PDA.
2) The RRMS developed from the rms generates a more smoothed degradation curve but shares similar scores as the rms on these metrics. Although NLLPEWMA and SOM-MQE do not achieve a high monotonicity score, they show a slightly better correlation and robustness than the rms and RRMS. GMM-JRD is sensitive to its parameter setting.

V. CONCLUSION

In this article, we presented an integrated approach using the improved GMM and a new HI for bearing performance degradation.

First, the HRGMM improves the clustering performance, where the Hellinger distance accurately measures the similarity between distributions in local manifold structure and the modified RGMM enhances the identification of intrinsic changes. The results of the experimental and real bearings show that the HRGMM can clearly distinguish varying states, such as defects, degree of damage, and working conditions.

Second, a new HI based on the HRGMM exhibited satisfactory performance by consistently depicting the bearing degradation over its entire lifetime. The results of two test-to-failure experiments validate that the HI shows a monotonic and relatively smooth trend, and has stable degradation assessment performance, which is essential for bearing prognostics. Moreover, it improves the comprehensibility of quantification indication. Using the proposed HI, the health state division automatically marks the degradation process from health to slight/sever degradation to failure. It also brings about sensitivity to the incipient defect, multiple stages, and a limited range [0, 1], which are beneficial in the selection of the failure threshold and prediction models.

The proposed HRGMM and HI can be further applied to other machine components, and offer an applicable method for the continuous monitoring of machines. In addition, machine components usually suffer from various failure modes resulting in the change in degradation trends. Thus, our future work will focus on the multistage prediction model with adaptively adjusted parameters.

REFERENCES

[1] Z. X. Zhang, X. S. Si, C. H. Hu, and Y. G. Lei, "Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods," *Eur. J. Oper. Res.*, vol. 271, no. 3, pp. 775–796, Dec. 2018.

[2] D. Wang, K. L. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.

[3] Y. Wang, Y. Z. Peng, Y. Y. Zi, X. H. Jin, and K. L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 924–932, Jun. 2016.

[4] A. Rai and S. H. Upadhyay, "The use of MD-CUMSUM and NARX neural network for anticipating the remaining useful life of bearings," *Measurement*, vol. 111, pp. 397–410, Dec. 2017.

[5] X. J. Liu, P. Song, C. Yang, C. B. Hao, and W. J. Peng, "Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1549–1558, Feb. 2018.

[6] B. Zhang, L. J. Zhang, and J. W. Xu, "Degradation feature selection for remaining useful life prediction of rolling element bearings," *Qual. Rel. Eng. Int.*, vol. 32, no. 2, pp. 547–554, 2016.

[7] J. Wu, C. Y. Wu, S. Cao, S. Wing Or, C. Deng, and X. Y. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 529–539, Jan. 2019.

[8] L. Guo, N. P. Li, F. Jia, Y. G. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.

[9] G. Q. Qiu, Y. K. Gu, and J. J. Chen, "Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model," *Measurement*, vol. 150, Jan. 2020, Art. no. 107097.

[10] J. Tang *et al.*, "Rolling bearing remaining useful life prediction via weight tracking relevance vector machine," *Meas. Sci. Technol.*, vol. 32, no. 2, Feb. 2021, Art. no. 024006.

[11] F. Yang, M. S. Habibullah, and Y. Shen, "Remaining useful life prediction of induction motors using nonlinear degradation of health index," *Mech. Syst. Signal Process.*, vol. 148, Feb. 2021, Art. no. 107183.

[12] R. Razavi-Far, M. Farajzadeh-Zanjani, and M. Saif, "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2758–2769, Dec. 2017.

[13] H. Hassani, J. Zarei, M. Mehdi Arefi, and R. Razavi-Far, "zSlices-based general type-2 fuzzy fusion of support vector machines with application to bearing fault detection," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7210–7217, Sep. 2017.

[14] A. Rai and S. H. Upadhyay, "Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering," *Mech. Syst. Signal Process.*, vol. 93, pp. 16–29, Sep. 2017.

[15] W. Ahmad, S. A. Khan, and J. M. Kim, "A hybrid prognostics technique for rolling element bearings using adaptive predictive models," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1577–1584, Feb. 2018.

[16] Z. Z. Pan, Z. Meng, Z. J. Chen, W. Q. Gao, and Y. Shi, "A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106899.

[17] L. L. Cui, X. Wang, H. Q. Wang, and J. F. Ma, "Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2858–2867, Jun. 2020.

[18] M. M. Yan, X. G. Wang, B. X. Wanga, M. X. Chang, and I. Muhammada, "Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model," *ISA Trans.*, vol. 98, pp. 471–482, Mar. 2020.

[19] J. Antoni and P. Borghesani, "A statistical methodology for the design of condition indicators," *Mech. Syst. Signal Process.*, vol. 114, pp. 290–327, Jan. 2019.

[20] J. B. 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.

[21] J. B. Yu, "Process monitoring through manifold regularization-based GMM with global/local information," *J. Process Control*, vol. 45, pp. 84–99, Sep. 2016.
[22] A. Rai and S. H. Upadhyay, "An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen–Rényi divergence," *Appl. Soft Comput.*, vol. 71, pp. 36–50, Oct. 2018.
[23] X. Xiu, J. Fan, Y. Yang, and W. Q. Liu, "Fault detection using structured joint sparse nonnegative matrix factorization," *IEEE Trans. Instrum. Meas.*, vol. 70, Mar. 2021, Art. no. 3513011.
[24] J. Wodecki, P. Kruczek, and A. Bartkowiak, "Novel method of informative frequency band selection for vibration signal using nonnegative matrix factorization of spectrogram matrix," *Mech. Syst. Signal Process.*, vol. 130, pp. 585–596, Sep. 2019.
[25] J. Singh, A. K. Darpe, and S. P. Singh, "Bearing damage assessment using Jensen–Rényi divergence based on EEMD," *Mech. Syst. Signal Process.*, vol. 87, pp. 307–339, Mar. 2017.
[26] J. Singh, A. K. Darpe, and S. P. Singh, "Bearing remaining useful life estimation using an adaptive data-driven model based on health state change point identification and K-means clustering," *Meas. Sci. Technol.*, vol. 31, no. 8, Aug. 2020, Art. no. 085601.
[27] Y. G. Lei, N. P. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1314–1326, Sep. 2016.
[28] H. B. Yang, Z. Sun, G. D. Jiang, F. Zhao, and X. S. Mei, "Remaining useful life prediction for machinery by establishing scaled-corrected health indicators," *Measurement*, vol. 163, Oct. 2020, Art. no. 108035.
[29] A. Rai and J. M. Kim, "A novel health indicator based on the Lyapunov exponent, a probabilistic self-organizing map, and the Gini-Simpson index for calculating the RUL of bearings," *Measurement*, vol. 164, Nov. 2020, Art. no. 108002.
[30] H. Pei *et al.*, "An adaptive prognostics method for fusing CDBN and diffusion process: Application to bearing data," *Neurocomputing*, vol. 421, pp. 303–315, Jan. 2021.
[31] Y. H. Chen, G. L. Peng, Z. Y. Zhu, and S. J. Li, "A novel deep learning method based on attention mechanism for bearing remaining useful life prediction," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105919.
[32] J. L. Liu, D. Cai, and X. F. He, "Gaussian mixture model with local consistency," in *Proc. 24th AAAI Conf. Artif. Intell., 22nd Innov. Appl. Artif. Intell. Conf.*, 2010, pp. 512–517.
[33] X. F. He, D. Cai, Y. L. Shao, H. J. Bao, and J. W. Han, "Laplacian regularized Gaussian mixture model for data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1406–1418, Sep. 2011.
[34] C. Li, B. Huang, and F. Qian, "Hellinger distance based probability distribution approach to performance monitoring of nonlinear control systems," *Chin. J. Chem. Eng.*, vol. 23. no. 12, pp. 1945–1950, Dec. 2015.
[35] I. R. Harris and A. Basu, "Hellinger distance as a penalized log likelihood," *Commun. Statist. - Simul. Comput.*, vol. 23, no. 4, pp. 1097–1113, 1994.
[36] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
[37] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
[38] Y. He, A. B. Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1211–1220, May 2003.
[39] A. Rai and J. M. Kim, "A novel health indicator based on information theory features for assessing rotating machinery performance degradation," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6982–6994, Sep. 2020.
[40] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2016, pp. 1–7.
[41] N. P. Li, Y. G. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.
[42] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on roller bearing prognostics," *J. Sound Vib.*, vol. 289, no. 4/5, pp. 1066–1090, 2006.
[43] X. H. Jin, Z. J. Que, Y. Sun, Y. J. Guo, and W. Qiao, "A data-driven approach for bearing fault prognostics," *IEEE Trans. Ind. Appl.*, vol. 55, no. 4, pp. 3394–3401, Jul./Aug. 2019.
[44] B. Dolenc, P. Boškoski, and Đ. D. Juričić, "Change detection based on entropy indices with application to bearing faults," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 1438–1443, 2015.
[45] Y. W. Cheng, H. P. Zhu, J. Wu, and X. Y. Shao, "Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 987–997, Feb. 2019.
[46] Y. G. Lei, N. P. Li, L. Guo, N. B. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
[47] F. Xu, Z. L. Huang, F. F. Yang, D. Wang, and K. L. Tsui, "Constructing a health indicator for roller bearings by using a stacked auto-encoder with an exponential function to eliminate concussion," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106119.
[48] Q. J. Tong and J. Z. Hu, "Bearing performance degradation assessment based on information-theoretic metric learning and fuzzy c-means clustering," *Meas. Sci. Technol.*, vol. 31, no. 7, Jul. 2020, Art. no. 075001.

Mao He received the B.S. degree from Northeast Petroleum University, Daqing, China, in 2014, and the M.S. degree from the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2021, both in mechanical engineering.

His research interests include feature selection and clustering analysis, degradation assessment, and remaining useful life prediction.

Wei Guo received the B.S. degrees in automation and English from the Taiyuan University of Technology, Taiyuan, China, in 2002, the M.S. degree in control theory and control engineering from the Dalian University of Technology, Dalian, China, in 2005, and the Ph.D. degree in system engineering and engineering management from the City University of Hong Kong, Hong Kong, in 2012.

She is currently an Associate Professor with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her research interests include condition monitoring and fault diagnosis, mechanical signal processing, intelligent diagnosis and control, and remaining useful life prediction.