




Research

# A mixture distributions analysis based feature selection approach for bearing remaining useful life estimation

Fei Huang<sup>1,2</sup>  · Alexandre Sava<sup>2</sup> · Kondo H. Adjallah<sup>2</sup> · Dongyang Zhang<sup>3</sup>

Received: 4 August 2023 / Accepted: 4 October 2023

Published online: 29 October 2023

© The Author(s) 2023 

## Abstract

Feature selection is a difficult but highly important preliminary step for bearings remaining useful life (RUL) estimation. To avoid the weights setting problem in hybrid metric, this work devotes to conduct feature selection by using a single metric. Due to noise and outliers, an existing feature selection metric, called monotonicity, used for estimating bearings RUL, requires data smoothing processing before adequate implementation. Such a smoothing process may remove significant part of meaningful information from data. To overcome this issue, a mixture distribution analysis-based feature selection metric is proposed. Moreover, based on this new metric, a feature selection approach for bearings RUL estimation is proposed. Numerical experiments benchmarking the proposed method and the existing metric monotonicity method on available real datasets highlight its effectiveness.

## Article Highlights

- A feature selection method for bearings remaining useful life estimation
- Retrieval the underlying trend of vibration signal
- Address the drawbacks of smoothing methods that remove meaningful information from data

**Keywords** Feature selection · Mixture distributions · Monotonicity · Remaining useful life · Bearing fault · Degradation monitoring

## 1 Introduction

Bearings are critical common units in rotating machinery, whose failures may lead to severe influences on the overall system performance. Moreover, the bearings' failure may damage other components in the machine. For instance, bearings' failure in high-speed wind turbines [1] could initiate the degradation of other parts of the gearbox, with expensive replacement costs. Authors in [2] have estimated the gearbox cost of a 5 MW wind turbine to be

more than USD 600 000. Thus, monitoring bearing degradation and estimating their remaining useful life (RUL) are essential for rotating machinery preventive maintenance activities [3, 4].

The bearings RUL estimation approach comprises data acquisition, feature extraction, and prognosis with remaining useful life estimation. Vibration signal data acquisition commonly uses accelerometers [1], and vibration signals often incorporate noises. Besides, bearings degrade according to a stochastic process, where degradation and

✉ Fei Huang, defdabian@foxmail.com | <sup>1</sup>Huaiyin Institute of Technology, 1 Meicheng East Road, Huaian 223003, Jiangsu, China. <sup>2</sup>LCOMS, Université de Lorraine, 57078 Metz, France. <sup>3</sup>Xi'an Novastar Tech Co., Ltd, 2 Yunshui Rd, Xi'an 710077, Shaanxi, China.



failure information are embedded in the vibration signal's raw data. It is thus helpful to extract features incorporating bearings' condition information. The extracted features would carry degradation and failure information, among which some are redundant inevitably [5]. Feature selection allows identifying the most relevant original features subset, while discarding redundant ones [6]. Through feature selection, one can estimate the RULs with enough accuracy based on appropriate features [7–12].

The vibration signal-based features can be obtained and presented in six domains, i.e., time-domain, frequency-domain, time–frequency-domain, state-space, phase-space, and complexity [13]. Every single feature describes the condition from a particular aspect; and carries partial information on the degradation condition. A hybrid metric is adopted in [14] to conduct feature selection, which is the weighted combination of three metric (monotonicity, robustness, and prognosability). However, obtaining an effective hybrid metric to achieve an appropriate weights setting is a big challenge. Hence, this work is devoted to conduct feature selection by using unique metric.

As having no self-healing ability, the degradation of mechanical systems or components is monotonically gradual over the whole lifespan. Therefore, showing a monotonic continuously increasing or decreasing evolution over bearings' entire lifespan is required for an ideal degradation feature.

The more monotonous the evolution trends in the used features, the more accurate the estimates of the RULs will be [15]. The quantification of the monotonicity degree of a feature has been defined in [16]. In [17], the authors propose using monotonicity as the metric for selecting the appropriate features and setting the length of the monitoring moving window. In [15], the authors used monotonicity as a genetic algorithm's fitness function to select the relevant prognosis features. In [18], the authors use a cumulative strategy to improve the monotonicity of the proposed prognosis features. By jointly maximizing the monotonic property and minimizing the variance in the failure threshold of different units, authors in [19] worked out data fusion-based features for degradation prognosis. However, as mentioned in [15, 16, 18], to reduce the sensitivity to noise and outliers and obtain monotonic features, it is necessary to preprocess the data by smoothing, then evaluate the monotonicity defined in [16]. Smoothing methods highlight the patterns while removing outliers from the raw data. The monotonicity performance for feature selection to health prognosis largely relies on the smoothing methods. The method selection and parameter setting are both involved during the application of smoothing methods. However, capturing the global trend

rather than reflecting local variations are conducted during the smoothing process.

Bearings degradation process could be interpreted as stochastic processes. There are several degradation regimes involved in one specific bearing degradation process. Thus, the assumption is that degradation features, during the whole lifecycle of the specific bearing, are derived from a mixture distribution. Each component distribution of this mixture distribution is associated with one certain degradation regime of the bearing degradation process. Hence, the mixture distribution analysis is introduced in this work to reveal the bearing vibration signal-based feature space.

As the main contributions in this paper: to overcome the drawbacks caused by smoothing methods, we propose a new feature selection metric characterized by a monotonous evolution based on the mixture distribution's weighted assessment of the feature data instead of the smoothed feature. We call this new feature selection metric "distribution monotonicity." Additionally, based on distribution monotonicity, a new feature selection approach is proposed.

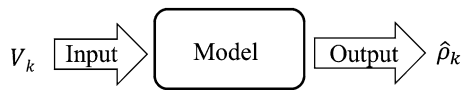
The remainder of this paper is organized as follows. Section 2 recalls the main background for estimating the RUL of the bearings, the existing definition of a feature's monotonicity and its limitation, and lastly, states the problem. Section 3 proposes the new features selection metric called distribution monotonicity. Moreover, a new feature selection approach based on the proposed metric is also proposed in Sect. 3. Then, Sect. 4 provides the performance assessment results of the proposed feature selection metric and method through a case study by using an existing real benchmark datasets from the literature. Lastly, in Sect. 5, this work is concluded and some directions for future research are also discussed.

## 2 Background and problem statement

### 2.1 The RUL modeling

The health prognosis refers to remaining useful life estimation. The general scheme in [20] is the RUL estimation model based on the fuzzy inference system for bearings by using a feature set extracted from vibration signals.

The scheme of health prognosis process is shown in Fig. 1. This scheme does not involve either a fixed bearing failure threshold or any available information on failure modes. Feature data extracted from monitoring vibration signals is used as the model input, denoted by  $V_k$ .  $k = 1, 2, \dots, K$  is the rank of observation. As the degradation processes among bearings are diverse, manually determining bearings failure thresholds could inevitably introduce noise and error. To avoid this issue, the estimated past useful life ratio is adopted as the output of the model, denoted by  $\hat{\rho}_k$ .



**Fig. 1** The schema for the RUL estimation

The RUL model accuracy depends on the quality of features extracted from the vibration signals for health conditions monitoring. Specifically, this paper addresses the problem of the bearing degradation features selection for the RUL estimation.

## 2.2 Existing monotonicity estimation metric

For mechanical components, such as bearings, there is no ability of self-healing during their degradation processes. In other words, the degradation of a bearing throughout its life cycle is monotonic. To achieve an accurate estimation of bearing RUL, vibration signal based features should be selected according to their monotonicity [15].

An existing monotonicity estimation metric  $M_i$ , defined in Eq. (1) [16] can be used to evaluate the degree of monotonicity of the  $i^{\text{th}}$  feature.

$$M_i = \left| \sum_{k=1}^{K-1} \frac{\text{sgn}(v_{(k+1),i} - v_{k,i})}{K-1} \right| \quad (1)$$

where  $v$  is the data point of  $V_k$  and  $i = 1, 2, \dots, l$  is the features label. Typically, the function defined by Eq. (1) takes values in  $[0, 1]$ . It corresponds to the statistic average of the trend of each pair of adjacent data points of the selected feature.

The presence of noise and outliers inevitably influences the data quality used for the feature extraction. An example is provided in [15], where the value of monotonicity is low when the function in Eq. (1) is directly applied, without smoothing, to the root mean square (RMS), a time-domain feature. In practice, it is necessary to adequately smooth the data before assessing the monotonicity [16]. There is a variety of smoothing methods, such as Savitzky-Golay filtering [21], adaptive LOESS filtering [22], and spline smoothing [23]. Smoothing methods highlight the trend patterns by removing outliers from data. Meanwhile, smoothing the same feature using different smooth methods or the same smoothing method with different parameter settings will result in different monotonicity values derived from Eq. (1). Thus, the performance of the monotonicity metric for feature selection largely depends on smoothing method selection and parameter setting.

In summary, there are two drawbacks with this existing monotonicity estimation metric due to the smoothing

implementation. First, they are exclusive to capture the global trend rather than reflect local variations. A significant part of meaningful information could be removed during the smoothing implementation. Second, the monotonicity assessment for feature selection largely relies on the smoothing methods choice and their parameter setting. Hence, this work aims to work out a selection method of meaningful features for accurate bearings RUL estimation based on a monotonicity assessment method without smoothing.

## 3 Feature selection for bearing RUL estimation

### 3.1 Distribution Monotonicity

To overcome the drawbacks of the monotonicity metric defined in Eq. (1), we suggest considering each data point for a feature monotonicity assessment. The sample data  $V$  denotes features which are extracted from a vibration signal collected throughout the whole life cycle of one bearing. The sample data  $V$  is a multivariate time series of multiple features. Since the bearing degradation evolution is complex, it is assumed that the sample data  $V$  follows a finite and countable multivariate mixture distribution. Let  $f(V_k)$  denote the probability of  $V_k$ , given the multivariate mixture distribution of features.

The new monotonicity assessment function, called distribution monotonicity, is denoted by  $DM$ .  $DM$  is a vector of features weighted trends rather than the features statistical trend. Elements of  $DM$  are defined in Eq. (2):

$$DM_i = \left| \sum_{k=2}^K \frac{f(V_k) \text{sgn}(v_{k,i} - v_{(k-1),i})}{K-1} \right| \quad (2)$$

Thus, the impact term  $f(V_k)$  is introduced into the monotonicity analysis, which is associated with each measure to characterize the impact of each data point on the monotonicity of a given feature. Hence, distribution monotonicity defined in Eq. (2) analyzes not only the features statistical monotonicity but also the underlying distribution attribute of each data point. In other words, distribution monotonicity assesses the essence of monotonicity to features.

### 3.2 Multivariate mixture distribution analysis in the feature space

A multivariate mixture distribution is illustrated by Eq. (3) with a constraint in Eq. (4), which is a weighted mixture of multivariate normal distributions [24]. Each single

component distribution defined in Eq. (5) is associated with a single regime component in samples space [24]. Specifically, it can make sense that each single component distribution is associated with a single degradation stage in the bearings features space. In addition, the probability of membership is defined in Eq. (6) with the Bayesian rule [24].

$$f(V_k) = \sum_{j=1}^J w_j g_j(V_k; C_j, H_j) \tag{3}$$

$$\sum_{j=1}^J w_j = 1 \tag{4}$$

$$g_j(V_k; C_j, H_j) = \frac{1}{(2\pi)^{l/2} |H_j|^{1/2}} e^{(-\frac{1}{2}(V_k - C_j)H_j^{-1}(V_k - C_j)^T)} \tag{5}$$

$$P(\theta_{R_j} | V_k) = \frac{P(\theta_{R_j})P(V_k | \theta_{R_j})}{P(V_k)} = \frac{w_j g_j(V_k; C_j, H_j)}{f(V_k)} \tag{6}$$

where  $H_j$  and  $C_j$  are respectively the covariance matrix and the mean vector associated with the  $j^{th}$  regime  $\theta_{R_j}$ .

As the definition in Eq. (3),  $f(V_k)$  could be considered as a function with parameters  $w_j$ ,  $H_j$ , and  $C_j$ . Hence, the maximum-likelihood estimation problem of parameter  $w_j$  could be formulated with the Lagrange multiplier method as in Eq. (7) [24].

$$\begin{aligned} \hat{w}_j &= \underset{w_j}{\operatorname{argmax}} \left( \sum_{k=1}^K \ln f(V_k) - \lambda \sum_{j=1}^J (w_j - 1) \right) \\ &= \underset{w_j}{\operatorname{argmax}} \left( \sum_{k=1}^K \ln \sum_{j=1}^J w_j g_j(V_k; C_j, H_j) - \lambda \sum_{j=1}^J (w_j - 1) \right) \end{aligned} \tag{7}$$

where  $\lambda$  is the Lagrange multiplier.

The part in the brackets of Eq. (7) is convex due to  $f(V_k)$  the mixture of multivariate normal distributions. Hence, the Eq. (8) is satisfied while  $w_j$  takes the maximum-likelihood value  $\hat{w}_j$ . Based on the derivation in [24] on Eq. (8) using Eq. (6),  $\hat{w}_j$  defined in Eq. (9) is the maximum-likelihood estimated values of the parameter  $w_j$ .

$$\frac{\partial (\sum_{k=1}^K \ln \sum_{j=1}^J w_j g_j(V_k; C_j, H_j) - \lambda (\sum_{j=1}^J w_j - 1))}{\partial w_j} = 0 \tag{8}$$

$$\hat{w}_j = \frac{1}{K} \sum_{k=1}^K P(\theta_{R_j} | V_k) \tag{9}$$

### 3.3 Probability of membership estimation

To obtain estimated value of  $P(\theta_{R_j} | V_k)$  in Eq. (9), the subtractive clustering method is conducted on a  $K \times (l + 1)$

matrix [25]. The  $K \times (l + 1)$  matrix is a collection of all the input and output data samples. All the parameters setting of subtractive clustering follows the work in [20].

Through subtractive clustering, the centroid of the  $j^{th}$  cluster in the input space is obtained as in Eq. (10). While, the degree of membership  $\mu_j(V_k)$  is also obtained, which is considered as the potentiality measurement of data point  $V_k$  for being in the  $j^{th}$  cluster of sample space.

$$C_j = [c_{j,1} \ c_{j,2} \ \dots \ c_{j,l}]^T \tag{10}$$

where  $j = 1, 2, \dots, J$  is the label of the clusters.

According to definition in the subtractive clustering, the degree of membership  $\mu_j(V_k)$  is subject to the following conditions:

$$\mu_j(V_k) \in [0, 1], \forall j, k; \sum_{j=1}^J \mu_j(V_k) = 1, \forall k; 0 < \sum_{k=1}^K \mu_j(V_k) < K, \forall j.$$

Compared with the definition of the probability of the membership function in Eq. (6),  $\mu_{j,k}$  can be considered as the estimated value of the probability of the membership  $P(\theta_{R_j} | V_k)$ . Hence, Eq. (9) is rewritten as Eq. (11).

$$\hat{w}_j = \frac{1}{K} \sum_{k=1}^K \hat{P}(\theta_{R_j} | V_k) = \frac{1}{K} \sum_{k=1}^K \mu_j(V_k) \tag{11}$$

### 3.4 Estimation procedure of function $f(V_k)$

The following describes the procedure for the impact term  $f(V_k)$ .

Step 1. The centroid vector  $C_j$  is obtained through subtractive clustering.

Step 2. The covariance matrix  $H_j$  is formed as in Eq. (12) following the setting in [20].

$$H_j = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_l^2 \end{bmatrix} \tag{12}$$

Note that: in fact, feature correlations are not totally irrelevant. However, this section aims to estimate the impacts respectively corresponding to each data point of every single feature (variable). Thus, the correlation between features is ignored. Hence, the covariance matrix  $H_j$  is defined as a diagonal matrix.

Step 3. Parameter  $w_j$  is obtained using formula (11).

Step 4. Substitute the previous parameters into Eq. (5) to obtain  $g_j(V_k; C_j, H_j)$ . Then, substitute  $g_j(V_k; C_j, H_j)$  into Eq. (3) to obtain  $f(V_k)$ .

### 3.5 The algorithm of feature selection approach

Step1: Extract features from vibration signals collected over the training bearings lifespan.

Step 2: Calculate the distribution monotonicity value of each feature in every training bearing.

Step 3: Calculate the average values of the distribution monotony of the same feature in all training bearings.

Step 4: Order the features according to the decreasing average value of the distribution monotonicity.

Step 5: Select the number of first sorted features to feed the RUL estimation model. (The number of selected features is fixed according to model complexity requirements. The complexity of the model increases with the number of input features.)

## 4 Case study

### 4.1 Benchmark datasets

We used the IEEE PHM 2012 benchmark dataset (bearing type: Ball bearing) to assess the effectiveness of the distribution monotonicity proposed in this paper for bearings health prognosis. By using the PRONOSTIA testbed [26], the IEEE PHM 2012 dataset is collected under three different load conditions with the same type of bearings. There is no information about failure modes. And no fixed failure threshold is provided. In this paper, two ball bearings with the tags (1–1, 1–2) were used as training bearings under load condition 1, and one ball bearing (1–3) was used as a test bearing under the same load condition.

Condition 1 operates at 1800 rpm with a 4000 N radial load. Vibration signal data is collected over an entire

run-to-failure period for each bearing. The data sampling rate is 25.6 kHz, and the recording duration is 0.1 s with 10 s sample intervals. Two accelerometers collect the vibration signals in the vertical and horizontal directions. For simplicity, we consider only the vibration signals in the horizontal direction.

### 4.2 Feature selection

We selected features with fewer or no parameters to decline the influence of parameter selection during the features extraction process in this case study. Thus, Table 1 displays 24 features extracted from the vibration signals collected for each training bearing over an entire run-to-fail period. Following the procedure introduced in Sect. 3, we applied respectively the monotonicity metric with smoothing preprocessing, and the distribution monotonicity without smoothing preprocessing, to implement the feature selection. As its advance in processing even spacing data, Savitzky-Golay filter is adopted for the smoothing preprocessing [27].

### 4.3 Assessment

Respectively, methods based on clustering algorithms, neural networks, and metaheuristics are the two main categories of identification methods implemented in the T-S FIS [33]. The identification method consisting of the subtractive clustering and least square estimation (FSC-LSE) could be a representative of the first category [25]. The method composed by the least square estimation and backpropagation gradient descent (BPGD-LSE) is belong to the second category [34]. For convenience, in

**Table 1** The list of features

Root-mean-square (RMS) [28]	Mean (MN) [28]	Band power (BP) [28]	Occupied bandwidth of the power spectral density (OBP) [28]
Standard deviation (STD) [28]	Maximum-to-minimum difference (MM) [28]	Median frequency (MF) [28]	Signal to noise and distortion ratio of power spectral density (SNP) [28]
Variance (VA) [28]	Mean frequency (MNF) [28]	Spectral entropy (SE) [29]	Spurious free dynamic range of the one-sided power spectrum (SFP) [28]
Kurtosis (KU) [28]	Spurious free dynamic range (SF) [28]	Mean frequency of a power spectral density estimation (MEP) [28]	approximate entropy (AE) [30]
Peak-magnitude-to-RMS ratio (PMR) [28]	Signal to noise and distortion ratio (SN) [28]	Median frequency of a power spectral density estimation (MFP) [28]	Largest Lyapunov exponent (LLE) [31]
Skewness (SK) [28]	Occupied bandwidth in terms of the sample rate (OB) [28]	Band power of the power spectral density estimation (BPP) [28]	Correlation dimension (CD) [32]

the sequel, we shall denote A the FSC-LSE based method, B the BPGD-LSE based method.

In this paper, the effectiveness of the distribution monotonicity for bearings health prognosis, is evaluated with FIS prognosis models identified respectively by two identification methods (e.g., method A and B). The parameters setting, involved in methods A and B, follows the work in [20]. Note that: to avoid considering the influence of the smoothing method choice and its parameters selection, the selected features imported into the model are without smoothing preprocessing.

We use the relative root mean square error (RRMSE) (13), and an average of ARRME (ARRMSE) (14) to evaluate the models' performance.

$$RRMSE_s = \sqrt{\frac{1}{K} \sum_{k=1}^K \left( \frac{\rho_k - \hat{\rho}_k}{\rho_k} \right)^2} \tag{13}$$

$$ARRMSE = \frac{1}{S} \sum_{s=1}^S RRMSE_s \tag{14}$$

where  $s = 1, 2, \dots, S$  is the number of selected features. The features are selected from whole features set in descending order by corresponding monotonicity assessment value.

### 4.4 Numerical results

Based on training bearings of the IEEE PHM 2012 datasets, features sorted by monotonicity are shown in Fig. 2. Features sorted by distribution monotonicity are shown in Fig. 3.

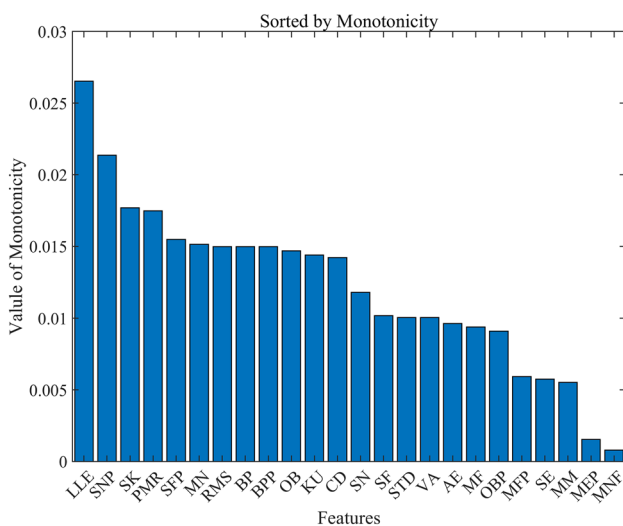


Fig. 2 Features sorted by monotonicity (related to IEEE PHM 2012 datasets)

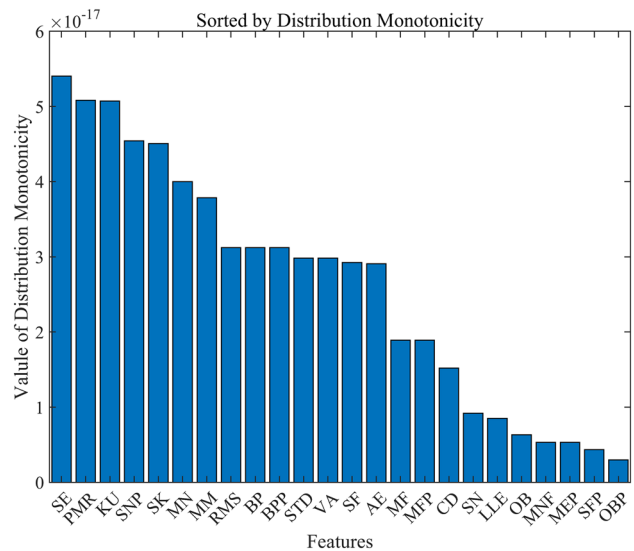


Fig. 3 Features sorted by distribution monotonicity (related to IEEE PHM 2012 datasets)

Table 2 Results with method A

Method	Monotonicity	Distribution monotonicity
RRMSEs	s=1	1.6724
	s=2	1.4622
	s=3	1.3482
	s=4	1.3146
	s=5	1.1705
	s=6	1.1800
	s=7	1.1320
	s=8	1.1400
	s=9	1.1400
	s=10	0.7969
	s=11	0.8098
	s=12	1.3936
	s=13	1.4022
	s=14	1.4013
	s=15	1.4428
	s=16	3.6605
	s=17	3.6794
	s=18	0.5498
	s=19	4.3952
	s=20	3.8718
	s=21	2.5158
	s=22	2.2833
	s=23	3.2170
	s=24	3.3953
ARRMSE	1.9323	1.3651

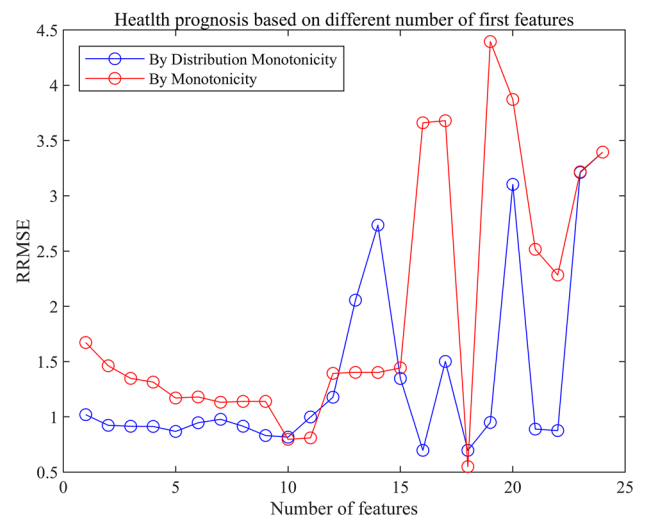
**Table 3** Results with method B

Method		Monotonicity	Distribution monotonicity
RRMSEs	s=1	1.6724	1.0625
	s=2	1.4537	0.9955
	s=3	1.3123	0.9749
	s=4	1.2335	0.9800
	s=5	1.1529	0.9586
	s=6	1.1888	0.8953
	s=7	1.1646	0.9305
	s=8	1.4206	0.9721
	s=9	1.2700	0.8540
	s=10	2.3664	0.8154
	s=11	3.1644	0.6847
	s=12	1.3427	1.3232
	s=13	3.4546	1.0775
	s=14	0.7297	0.7789
	s=15	2.3659	0.4808
	s=16	2.0887	1.0105
	s=17	1.3981	1.1108
	s=18	6.9399	0.8867
	s=19	1.7031	0.5584
	s=20	13.5209	10.5667
	s=21	9.9597	0.5672
	s=22	10.4508	10.2013
	s=23	1.2960	1.8199
	s=24	0.9711	0.9711
ARRMSE		3.0675	1.7282

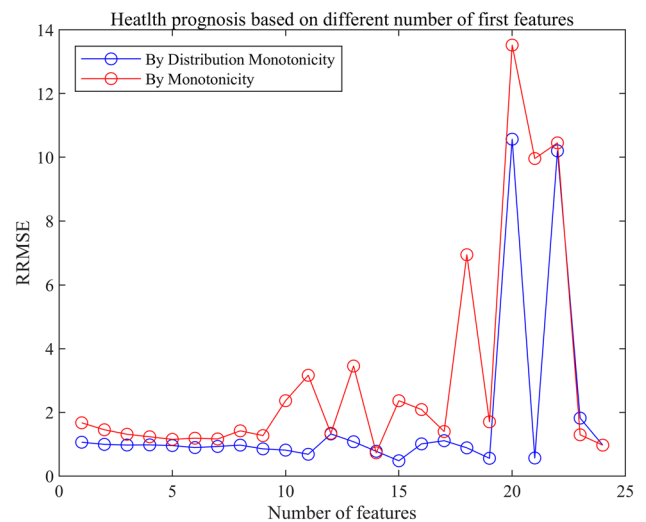
In Tables 2, 3, Figs. 4 and 5, the results of the health prognosis of test bearings based on different numbers of features are shown, respectively selected by using the monotonicity and the distribution monotonicity.

Based on the results shown in Tables 2, 3, Figs. 4 and 5:

- (1) When the number of features exceeds 11, the RUL prediction performance fluctuates wildly. Therefore, we suggest using the first 5 to 11 features for the bearing monitoring.
- (2) Compared with using monotonicity, fewer number of first features are needed to obtain the results with acceptable accuracy by using distribution monotonicity.
- (3) Compare with using monotonicity, the values of ARRMSE are smaller with the same type of model by using distribution monotonicity.
- (4) For a model identified by a particular method, an increase in the variety of the input features is not always accompanied by an increase in its accuracy. It can make sense that some input features might contain inaccurate information resulting in a drop in model accuracy.



**Fig. 4** The results based on method A related to IEEE PHM 2012 datasets



**Fig. 5** The results based on method B related to IEEE PHM 2012 datasets

## 5 Conclusions

Feature selection for bearing degradation monitoring is a difficult but highly important preliminary step. The performance of existing monotonicity-based feature selection approach for bearing RUL estimation highly relies on the data smoothing preprocessing implementation. However, a significant part of meaningful information could be removed during the smoothing implementation. Moreover, the choice of various smoothing methods and their parameters setting can lead to uncertainty in the performance of existing monotonicity-based feature selection approach.

To avoid these drawbacks, we proposed a mixture distribution analysis-based metric for feature selection for bearing RUL estimation, called distribution monotonicity. This metric involves the mixture distribution probability associated with each measure to characterize the impact of each data point on the monotonicity of a given feature. Hence, the proposed metric analyzes not only the features statistical monotonicity but also the underlying distribution attribute of each data point. In other words, distribution monotonicity assesses the essence of monotonicity to features. Moreover, it has no requirement of smoothing implementation. Further, a distribution monotonicity-based feature selection approach is also proposed.

The performance of the proposed metric and feature selection approach were assessed using a benchmark data from the literature. Compared with using monotonicity with smoothing preprocessing, fewer number of features are needed to obtain the results with acceptable accuracy by using distribution monotonicity. In addition, the values of ARRME are smaller with the same type of model by using distribution monotonicity.

The numerical results show that mixture distribution analysis-based monotonicity evaluation, namely distribution monotonicity, and the proposed feature selection approach are promising for the feature selection for bearings' RUL estimation.

Grounded on the results obtained, our future works will investigate the appropriate method for determining the number of selected features setting with an automatic feature selection approach.

**Author contributions** FH wrote the manuscript and prepared figures 1-5. AS and KHA reviewed and edited the main manuscript text. DZ prepared data. All authors reviewed the manuscript.

**Data availability** The IEEE PHM 2012 datasets are available at <http://www.femto-st.fr/ieee-PHM2012-data-challenge>

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Saidi L, Ben AJ, Bechhoefer E, Benbouzid M (2017) Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR. *Appl Acoust* 120:1–8. <https://doi.org/10.1016/j.apacoust.2017.01.005>
2. Sheng S, Link H, LaCava W, van Dam J, McNiff B, Veers P, Keller J, Butterfield S, Oyague F (2011) Wind turbine drivetrain condition monitoring during GRC phase 1 and phase 2 testing. National Renewable Energy Lab.(NREL), Golden, CO (United States)
3. Upadhyay N, Metsebo J, Kankar PK, Nbandjo N (2018) An Improved Theoretical Model of Unbalanced Shaft-Bearing System for Accurate Performance Prediction of Ball Bearing Due to Localized Defects. *Iran J Sci Technol Trans Mech Eng* 42:293–309. <https://doi.org/10.1007/s40997-017-0098-9>
4. Shoor S, Singh M (2022) Detecting Crack Propagation Rate in Taper Roller Bearing Using FEA Analysis and Signal Processing Techniques. *Iran J Sci Technol Trans Mech Eng* 46:1129–1139. <https://doi.org/10.1007/s40997-022-00486-0>
5. Luo M, Li C, Zhang X, Li R, An X (2016) Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings. *ISA Trans* 65:556–566. <https://doi.org/10.1016/j.isatra.2016.08.022>
6. Zheng J, Pan H, Tong J, Liu Q (2022) Generalized refined composite multiscale fuzzy entropy and multi-cluster feature selection based intelligent fault diagnosis of rolling bearing. *ISA Trans* 123:136–151. <https://doi.org/10.1016/j.isatra.2021.05.042>
7. Zhu X, Zhang Y, Zhu Y (2012) Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features. *J Mech Sci Technol* 26:2649–2657. <https://doi.org/10.1007/s12206-012-0716-9>
8. Guan X, Chen G (2019) Sharing pattern feature selection using multiple improved genetic algorithms and its application in bearing fault diagnosis. *J Mech Sci Technol* 33:129–138. <https://doi.org/10.1007/s12206-018-1213-6>
9. Kateris D, Moshou D, Pantazi X-E, Gravalos I, Sawalhi N, Loutridis S (2014) A machine learning approach for the condition monitoring of rotating machinery. *J Mech Sci Technol* 28:61–71. <https://doi.org/10.1007/s12206-013-1102-y>
10. Gu Y-K, Zhou X-Q, Yu D-P, Shen Y-J (2018) Fault diagnosis method of rolling bearing using principal component analysis and support vector machine. *J Mech Sci Technol* 32:5079–5088. <https://doi.org/10.1007/s12206-018-1004-0>
11. Li Y, Zhang W, Xiong Q, Luo D, Mei G, Zhang T (2017) A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM. *J Mech Sci Technol* 31:2711–2722. <https://doi.org/10.1007/s12206-017-0514-5>
12. Zhou Y, Liu C, Yu X, Liu B, Quan Y (2022) Tool wear mechanism, monitoring and remaining useful life (RUL) technology based on big data: a review. *SN Appl Sci* 4:232. <https://doi.org/10.1007/s42452-022-05114-9>
13. Caesarendra W, Tjahjowidodo T (2017) A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines* 5:21. <https://doi.org/10.3390/machines5040021>
14. Kumar PS, Kumaraswamidhas LA, Laha SK (2021) Selection of efficient degradation features for rolling element bearing prognosis using Gaussian Process Regression method. *ISA Trans* 112:386–401. <https://doi.org/10.1016/j.isatra.2020.12.020>
15. Liao L (2013) Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Trans Ind Electron* 61:2464–2472. <https://doi.org/10.1109/TIE.2013.2270212>



16. Coble J, Hines JW (2009) Identifying optimal prognostic parameters from data: a genetic algorithms approach. In: Annual Conference of the PHM Society
17. Ben AJ, Chebel-Morello B, Saidi L, Malinowski S, Fnaiech F (2015) Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mech Syst Signal Process* 56:150–172. <https://doi.org/10.1016/j.ymssp.2014.10.014>
18. Javed K, Gouriveau R, Zerhouni N, Nectoux P (2014) Enabling health monitoring approach based on vibration data for accurate prognostics. *IEEE Trans Ind Electron* 62:647–656. <https://doi.org/10.1109/TIE.2014.2327917>
19. Liu K, Gebraeel NZ, Shi J (2013) A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. *IEEE Trans Autom Sci Eng* 10:652–664. <https://doi.org/10.1109/TASE.2013.2250282>
20. Huang F, Sava A, Adjallah KH, Wang Z (2021) Fuzzy model identification based on mixture distribution analysis for bearings remaining useful life estimation using small training data set. *Mech Syst Signal Process* 148:107173. <https://doi.org/10.1016/j.ymssp.2020.107173>
21. Orfanidis SJ (1995) Introduction to signal processing. Prentice-Hall, Inc.
22. Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc* 83:596–610. <https://doi.org/10.1080/01621459.1988.10478639>
23. Craven P, Wahba G (1978) Smoothing noisy data with spline functions. *Numer Math* 31:377–403. <https://doi.org/10.1007/BF01404567>
24. Wolfe JH (1970) Pattern clustering by multivariate mixture analysis. *Multivariate Behav Res* 5:329–350. [https://doi.org/10.1207/s15327906mbr0503\\_6](https://doi.org/10.1207/s15327906mbr0503_6)
25. Chiu SL (1994) Fuzzy model identification based on cluster estimation. *J Intell fuzzy Syst* 2:267–278. <https://doi.org/10.3233/IFS-1994-2306>
26. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-Morello B, Zerhouni N, Varnier C (2012) PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In: IEEE International Conference on Prognostics and Health Management, PHM'12. IEEE Catalog Number: CPF12PHM-CDR, pp 1–8
27. Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639. <https://doi.org/10.1021/ac60319a045>
28. Paulter NG, Larson DR, Blair JJ (2004) The IEEE standard on transitions, pulses, and related waveforms, Std-181-2003. *IEEE Trans Instrum Meas* 53:1209–1217. <https://doi.org/10.1109/TIM.2004.831470>
29. Pan YN, Chen J, Li XL (2009) Spectral entropy: a complementary index for rolling element bearing performance degradation assessment. *Proc Inst Mech Eng Part C J Mech Eng Sci* 223:1223–1231. <https://doi.org/10.1243/09544062JMES1224>
30. Yan R, Gao RX (2007) Approximate entropy as a diagnostic tool for machine health monitoring. *Mech Syst Signal Process* 21:824–839. <https://doi.org/10.1016/j.ymssp.2006.02.009>
31. Caesarendra W, Kosasih B, Tieu AK, Moodie CAS (2015) Application of the largest Lyapunov exponent algorithm for feature extraction in low speed slew bearing condition monitoring. *Mech Syst Signal Process* 50:116–138. <https://doi.org/10.1016/j.ymssp.2014.05.021>
32. Logan D, Mathew J (1996) Using the correlation dimension for vibration fault diagnosis of rolling element bearings—I. Basic concepts *Mech Syst Signal Process* 10:241–250. <https://doi.org/10.1006/mssp.1996.0018>
33. Hellendoorn H, Driankov D (2012) Fuzzy model identification: selected approaches. Springer Science & Business Media
34. Jang J-S (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685. <https://doi.org/10.1109/21.256541>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.