



A health indicator enabling both first predicting time detection and remaining useful life prediction: Application to rotating machinery

Yun-Sheng Zhao ^a, Pengfei Li ^{a,b}, Yu Kang ^{a,b,c}, Yun-Bo Zhao ^{a,b,c,*}

^a Department of Automation, University of Science and Technology of China, 96 Jinzhai Road, Hefei, 230027, Anhui, PR China

^b Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, 5089 Wangjiang W Road, Hefei, 230026, Anhui, PR China

^c Institute of Advanced Technology, University of Science and Technology of China, 5089 Wangjiang W Road, Hefei, 230000, Anhui, PR China

ARTICLE INFO

Keywords:

Health indicator
First predicting time
Remaining useful life
Multi-objective optimization problem

ABSTRACT

Remaining Useful Life (RUL) prediction is vital for system functionality. Non-end-to-end approaches is an important type of RUL prediction approaches for their important application in industrial scenarios. In non-end-to-end approaches, Health Indicator (HI) construction is a critical aspect. However, existing HI construction approaches ignore First Predicting Time (FPT) detection, leading to increased domain knowledge demand and system health comprehension difficulty. To address this issue, this paper proposes a multi-objective-optimization-based HI construction approach enabling both FPT detection and RUL prediction. A novel metric called the monotonicity strength index is proposed to address the limitation of the conventional monotonicity. The constructed HI possesses the ability to indicate FPT by taking the detectability metric as an optimization objective. The optimization problem is solved by the combination of the multi-objective ant lion optimizer and the entropy weight method. The superiority of this HI is demonstrated through experiments on the widely used IMS bearing dataset and a gearbox dataset.

1. Introduction

Remaining Useful Life (RUL) of a component or a system, *i.e.*, the length from the current time to the end of the useful life, plays a crucial role in maintaining normal functioning [1]. Accurate RUL prediction enables engineers to proactively maintain systems, preventing severe consequences due to system failure [2]. Numerous studies have been conducted on the prediction of RUL, which can be divided into end-to-end [3–6] and non-end-to-end approaches [7–9].

For industrial scenarios with limited data and high requirements for interpretability, non-end-to-end approaches, rather than end-to-end ones, are often desired for RUL prediction. Such approaches typically involve two components: Health Indicator (HI) construction and degradation modeling, in which, HI is a measure of system health, and the degradation model is used to predict RUL.

An appropriate HI plays a fundamental role in non-end-to-end approaches since it is responsible for accurately reflecting the health status of systems [10,11]. For instance, an HI is expected to demonstrate a monotonic behavior, aligning with the monotonic degradation of a deteriorating system (except for some self-healing systems like lithium batteries [12]). If the HI shows violent fluctuations, it becomes challenging for individuals to understand the current health status of

a system, also leading to a significant increase in the difficulty of predicting RUL by the degradation model. Consequently, an appropriate HI aids in understanding the health status and facilitates precise RUL prediction.

The construction of an appropriate HI entails two key issues that have attracted significant research interest. These issues, as identified in [13], are: (1) How to extract meaningful information from monitoring data that can indicate degradation? (2) How to assess whether the constructed HI is beneficial for RUL prediction? The first issue always involves complex systems as certain raw monitoring data, like flank wear on a cutting tool [7], can be directly used as HIs in certain simple systems. However, for complex systems such as bearings, machine tools, and turbofan engines, cleverer features must be employed to effectively capture health information due to the inherent complexity of their underlying physics. For complex systems, many statistical-based feature extraction approaches have been proposed, including time-domain-based [14], frequency-domain-based [15], time-frequency-domain-based [16], Empirical Modal Decomposition (EMD)-based [17], and other advanced approaches, such as [18–20]. In addition, many machine learning approaches, *e.g.*, self-organizing map [8], the random forest method [21], and deep neural network [7,22],

* Corresponding author at: Department of Automation, University of Science and Technology of China, 96 Jinzhai Road, Hefei, 230027, Anhui, PR China.
E-mail addresses: zys1030@mail.ustc.edu.cn (Y.-S. Zhao), puffylee@ustc.edu.cn (P. Li), kangduyu@ustc.edu.cn (Y. Kang), ybzhao@ustc.edu.cn (Y.-B. Zhao).

have been used for degradation feature extraction. As for the second issue, many metrics have been proposed to evaluate whether an HI is suitable for RUL prediction, such as monotonicity, trendability, prognosability [23], robustness [24], etc.

Based on these feature extraction approaches and HI evaluation metrics, numerous HI construction approaches have been proposed. The task of constructing an HI is converted into an optimization problem, aiming to combine features into an HI. To make the HI perform better, the metrics are utilized as objectives to determine the optimal fusion parameters. These approaches can be categorized into two groups: one group transforms the HI construction problem into Multi-objective Optimization Problems (MOPs) [9,25,26], while the other converts it into single-objective optimization problems [27,28].

However, existing HI construction approaches lack the consideration of the First Predicting Time (FPT) of systems, which may lead to a higher requirement of domain knowledge and greater difficulty in understanding the health status. Typically, systems do not begin to degrade immediately upon operation but undergo a period of healthy operation, whose endpoint is known as the FPT or initial degradation time [29]. In other words, FPT is identified as the moment at which the system displays evident degradation. Upon detection of the FPT, it becomes feasible to initiate the prediction of the RUL of the system. Accurate detection of the FPT is critical, as it directly impacts subsequent RUL prediction and maintenance decisions. Early estimates of the FPT may incorporate irrelevant data into the RUL prediction process, while delayed estimates may exclude pertinent data, both resulting in inaccurate RUL predictions. Additionally, this may lead to resource wastage and delayed maintenance decisions. Although a considerable amount of effort has been devoted to studying FPT detection, these studies often focus on specific systems such as bearings and gears, and utilize specific features for FPT detection [30,31]. However, these approaches need additional features from the HIs that are exclusively constructed for degradation models. This requirement for supplementary features not only increases the demand for domain knowledge but also limits their applicability across different systems, thus presenting challenges for practical implementation. Additionally, engineers cannot intuitively comprehend the health status of the system, as the degradation level and the indication of system degradation are represented by distinct measures. If an HI construction approach can facilitate both FPT detection and RUL prediction, it would substantially reduce the necessity for domain knowledge and enhance the comprehension of the health status of the system for engineers.

With the aforementioned motivation, a general HI that enables simultaneous FPT detection and RUL prediction is constructed by taking the ability to detect FPT, *i.e.* detectability, as an objective in the optimization problem. In the process of HI construction, the construction is converted into an MOP, with the FPT detection being considered by taking the detectability metric as one of the objectives. Three other metrics are also taken into account to ensure accurate RUL prediction. To solve this MOP, a combination of the Multi-Objective Ant Lion Optimizer (MOALO) [32] and the Entropy Weight Method (EWM) [33] is utilized. Based on the constructed HI, FPT detection and RUL prediction can be performed. Two case studies on two rotating machinery datasets are provided to validate the proposed approach at last.

The main contributions of this paper are as follows:

- (1) By integrating detectability into the construction process, a holistic framework is proposed to construct an HI, which is capable of representing the health status of a degradation system and facilitating FPT detection and RUL prediction. This framework could reduce the necessity for domain knowledge in the RUL prediction process and enhance the comprehension of the health status of systems.
- (2) An approach based on the combination of MOALO and the EWM is proposed to obtain the optimal solution from the established MOP. The issue of subjective determination of metric weights

that has been observed in previous studies [9,23,26] can be effectively addressed with the application of EWM, which applies information entropy to evaluate the importance of each metric.

- (3) A new HI evaluation metric called the Monotonicity Strength Index (MSI) is proposed to address the limitation of the commonly used metric, monotonicity, which fails to capture the overall trend of an HI. This is achieved by adding a penalty item and a shape parameter.

The rest of this paper is structured as follows: In Section 2, the HI construction task is decomposed into three sub-tasks and is expressed mathematically. Section 3 presents the methodology of HI construction. Section 4 provides two illustrative examples on a bearing dataset and a gearbox dataset to demonstrate the validity of the approach. Finally, Section 5 concludes this paper.

2. Problem statement of the HI construction task

This section presents the mathematical formulation of the task involving the construction of an HI to enable both FPT detection and RUL prediction. Several studies have formulated the HI construction task as an optimization problem [9,25,26]. However, this formulation is not expressed clearly in these papers. Therefore, we rephrase it using mathematical form for clarity. The HI construction task can be decomposed into the following three sub-tasks:

- (1) **Determine appropriate metrics.** Evaluation metrics for HI are mathematical functions that assess specific properties that HI should adhere to, such as monotonicity and prognosability. By optimizing the performance of HI on these metrics, appropriate HIs can be constructed. Therefore, the key to the performance of the constructed HI primarily depends on the suitability of the selected metrics. The selected metrics aim for accurate FPT detection and RUL prediction are supposed to be:

$$\mathbf{M} = \{M_1(\text{HI}), M_2(\text{HI}), \dots, M_{n_m}(\text{HI})\} \quad (1)$$

where n_m denotes the number of metrics, and a higher value of $M_i(\text{HI})$, where $i = 1, 2, \dots, n_m$, indicates a better performance of the HI on metric $M_i(\text{HI})$.

- (2) **Extract and select features.** The raw monitoring data of an industrial system, which contains various noises and redundant information, cannot directly reflect the health status of the system. Therefore, it is necessary to extract and select features that contain information about the health status of the system from the raw data. Various approaches can be employed to extract features, and the selection of these features can be based on their performance on metrics \mathbf{M} . Let the selected features be denoted as:

$$\mathbf{F} = [F_1, F_2, \dots, F_{n_f}]^T. \quad (2)$$

where n_f denotes the number of selected features. After determining these features, an HI can be constructed using linear fusion, which is a widely used feature fusion approach [9,27,34]:

$$\text{HI} = w^T \mathbf{F} \quad (3)$$

where $w = [w_1, w_2, \dots, w_{n_f}]^T$ is a set of weights. Therefore, the task of constructing the HI can be defined as identifying the optimal solution for the following MOP:

$$\begin{aligned} \max_{w \in \mathbb{R}^{n_f}} g(w) &= [M_1(\text{HI}), M_2(\text{HI}), \dots, M_{n_m}(\text{HI})] \\ &= [M_1(w^T \mathbf{F}), M_2(w^T \mathbf{F}), \dots, M_{n_m}(w^T \mathbf{F})] \end{aligned} \quad (4)$$

- (3) **Solve the established MOP.** After the determination of \mathbf{M} and \mathbf{F} , the only remaining task is to obtain an optimal solution from the optimization problem, *i.e.*, Eq. (4).

These three issues are solved in Section 3.

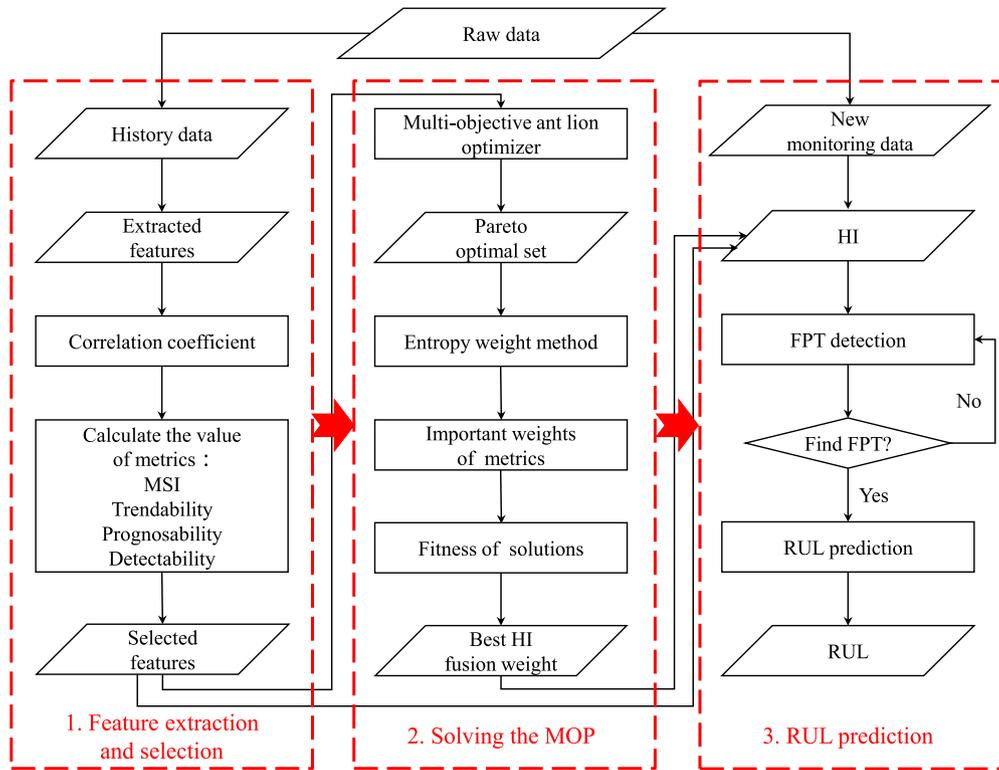


Fig. 1. The flowchart of the entire process of predicting RUL based on the proposed HI construction approach.

Remark 1. The problems related to HI construction can be categorized based on the availability of actual physical quantities that represent the health status. In systems where such quantities are present, they can be directly measured, exhibit monotonic changes, and have predefined thresholds for system failure. However, in practice, monitoring these physical quantities directly online can pose challenges (e.g., some physical quantities may require measurement downtime, thus affecting work efficiency). In such scenarios, a mapping relationship between easily measurable physical quantities and the HI can be established using supervised learning methods, as demonstrated in previous works [7,35]. On the other hand, there are systems where such physical quantities do not exist, and only a virtual HI can be constructed based on specific metrics. Our work falls into this category, which motivated the adoption of a multi-objective optimization approach to address this challenge.

3. Methodology of HI construction

In this section, the methodology of HI construction is proposed. The subsequent subsections address the three aforementioned issues raised in the previous section. Firstly, we introduce the metrics that serve as the objectives of the MOP, including the novel metric MSI. Then, the method of extracting and selecting features as candidate elements of the HI is introduced. At last, the solving of the MOP is achieved by the combination of the MOALO and EWM. The entire process is visually depicted in Fig. 1, providing a clear illustration of the methodology.

3.1. The determination of HI evaluation metrics

This section determines the metrics applied as objectives in Eq. (4). First, a review of commonly used metrics in the literature for assessing HIs is provided. Subsequently, to address the limitation of the traditional monotonicity metric that ignores the overall trend of an HI, a novel metric called MSI is introduced. Lastly, four significant metrics are selected as the objectives of Eq. (4), along with a justification.

3.1.1. Some commonly used HI assessment metrics

Several metrics have been proposed in the literature to assess the suitability of an HI for predicting the RUL [23,24]. To facilitate the introduction of these metrics, certain mathematical assumptions must be made. Assume there are m run-to-failure trajectories obtained from a set of systems with identical specifications. For system j , where $j = 1, 2, \dots, m$, the number of monitoring data is denoted by N_j . For a given feature F_i , where $i = 1, 2, \dots, n_f$, $F_i = \{F_{i,j}\}_{j=1:m}$ represents the sequence of feature i extracted from the monitoring data of system j . Here $F_{i,j} = [f_{i,j,1}, f_{i,j,2}, \dots, f_{i,j,N_j}]^T$. Since an HI is the fusion of features, there are no differences in assessing features and HIs using a metric. Therefore, an HI is treated as a feature in the rest of Section 3.1.

For all the metrics below, a higher value indicates that the HI performs better in that aspect.

Monotonicity: System degradation can be classified into two types: monotonic and non-monotonic. Unlike certain systems, such as lithium batteries [12] and self-healing materials [36], which exhibit non-monotonic degradation due to their self-healing characteristic, the degradation of mechanical systems is typically irreversible due to their physics. For this reason, the degradation process of a system should be monotonic. Monotonicity can be measured using the following equation [23]:

$$\text{Mon}(F_i) = \text{mean} \left(\frac{|\text{No. of } d/df_{i,j} > 0 - \text{No. of } d/df_{i,j} < 0|}{N_j - 1} \right) \quad (5)$$

where $d/df_j = f_{i,j,k+1} - f_{i,j,k}$ represents the difference between two consecutive values in $F_{i,j}$. Additionally, $\text{No. of } d/df_{i,j} > 0$ and $\text{No. of } d/df_{i,j} < 0$ denote the number of positive and negative differences, respectively.

Trendability: To ensure that a feature is capable of indicating the degradation process of systems with the same specifications, it is expected to show similar degradation patterns among them. Hence, trendability is proposed to evaluate this similarity. A trendability index was proposed in [23], but highly sensitive to noise. Afterward, the trendability of different trajectories is measured using the minimum

correlation coefficient, as suggested in [37]:

$$\text{Tre}(F_i) = \min(|\text{corr}(F_{i,j_1}, F_{i,j_2})|) \quad (6)$$

where $\text{corr}(\cdot, \cdot)$ denotes the Pearson correlation coefficient of two sequences, where $j_1, j_2 = 1, 2, \dots, m$. Note that the correlation coefficient needs to be calculated between sequences of equal length, thus Eq. (6) cannot be applied directly. In practice, linear interpolation is first performed for shorter sequences.

Prognosability: In RUL prediction, it is necessary to establish a critical threshold to determine the occurrence of system failure. As a result, an ideal HI should demonstrate consistent values when different systems fail. To this end, the concept of prognosability has been introduced, which is quantified as follows [23]:

$$\text{Pro}(F_i) = \exp\left(\frac{-\text{std}(F_{i,\text{begin}})}{\text{mean}|F_{i,\text{end}} - F_{i,\text{begin}}|}\right) \quad (7)$$

where $F_{i,\text{begin}} = [f_{i,1,1}, f_{i,2,1}, \dots, f_{i,m,1}]^T$ and $F_{i,\text{end}} = [f_{i,1,N_1}, f_{i,2,N_2}, \dots, f_{i,m,N_m}]^T$ denote the vector consisting of initial and failure values of a particular feature in different systems, respectively. In addition, $\text{std}(\cdot)$ and $\text{mean}(\cdot)$ represent the standard deviation and mean, respectively.

Detectability: As mentioned before, the system operation process is usually multi-stage, with each stage displaying distinct degradation patterns. In most cases, the operation process can be divided into the normal stage and the rapid degradation stage according to the FPT. The detectability of the two-stage degradation of systems can be computed using the following formula [27]:

$$\text{Det}(F_i) = \text{mean}\left(\exp\left(-\frac{\alpha\sigma_{i,j}}{\sqrt{|f_{i,j}^c - \mu_{i,j}|}}\right)\right) \quad (8)$$

where α is a scale factor, $f_{i,j}^c$ denotes the monitoring value of the point after the FPT. In addition, $\mu_{i,j}$ and $\sigma_{i,j}$ denote the mean and standard deviation of monitoring values during the normal stage, respectively.

3.1.2. Our proposed monotonicity strength index

A new metric that can measure the monotonicity level of HIs, named Monotonicity Strength Index (MSI), is proposed in this subsection, to address the limitation of the conventional one, i.e. Eq. (5).

We first illustrate the limitation of the conventional monotonicity index (Eq. (5)) through an example. The monotonicity achieved by Eq. (5) in certain situations can be misleading, which could be attributed to its neglect of the magnitude of the changes between consecutive monitoring points in its calculations. Fig. 2 shows two HIs, $y_1 = 0.5t + 0.5 \times (-1)^t$ and $y_2 = 0.5 \times (t \bmod 4)$, where $t = 0, 1, \dots, 20$. To be specific, y_1 has a trending folding line shape, while y_2 is essentially horizontal. According to Eq. (5), the monotonicity values of y_1 and y_2 are $\text{Mon}(y_1) = 0$ and $\text{Mon}(y_2) = 0.5$, respectively. This suggests that y_2 is preferred over y_1 based on Eq. (5), which is counterintuitive. In practice, however, y_1 would be preferred over y_2 as an HI, because y_1 provides information about degradation while y_2 does not. Therefore, it is misleading to evaluate the monotonicity of an HI by Eq. (5) in this situation.

This phenomenon occurs because Eq. (5) is only related to the sign of the difference between consecutive monitoring points, disregarding its magnitude. The magnitudes of these differences affect the overall trend of an HI, with larger differences having a greater impact. However, Eq. (5) treats a large difference and a small one equally, which is equivalent to giving them the same impact weight to the overall monotonicity. In the example above, y_1 has an equal number of positive differences and negative differences, with magnitudes of 1.5 and 0.5 respectively. Eq. (5) disregards their magnitudes, leading to $\text{Mon}(y_1) = 0$, which does not match the overall increasing trend.

Therefore, we design the MSI to evaluate the monotonicity of an HI. MSI assigns different levels of importance to these differences based on

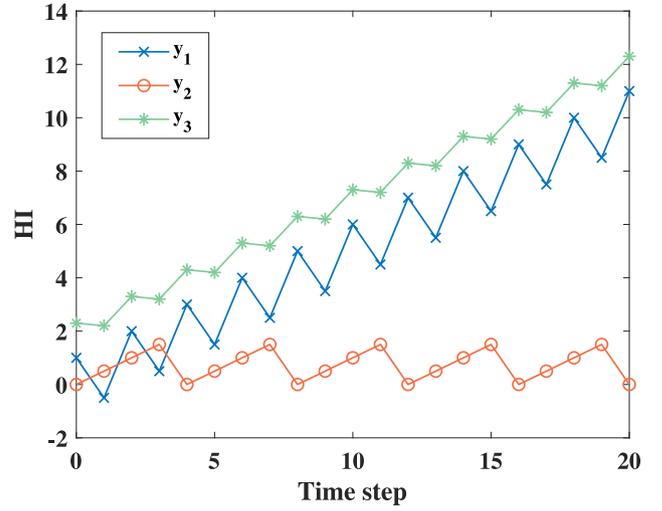


Fig. 2. An illustrative example to show the limitation of the widely used metric monotonicity (Eq. (5)), where $y_1 = 0.5t + 0.5 \times (-1)^t$, $y_2 = 0.5 \times (t \bmod 4)$, and $y_3 = 0.5t + 0.3 \times (-1)^t + 2$.

their respective magnitudes. Furthermore, since an HI is expected to be monotonic, differences with the opposite sign are supposed to be penalized in MSI.

In light of the foregoing, we first define a function $f(\cdot)$ to evaluate the effect of a difference on the HI:

$$f(d/df_{i,j}) = \begin{cases} (d/df_{i,j})^\beta, & d/df_{i,j} > 0 \\ p \cdot |d/df_{i,j}|^\beta, & d/df_{i,j} \leq 0 \end{cases} \quad (9)$$

where $p \leq -1$ is a penalty factor, $\beta \leq 0$ is a shape parameter introduced to control the flexibility and distortion of the function, and $d/df_j = f_{i,j,k+1} - f_{i,j,k}$ has the same meaning as in Eq. (5). Here, the upward trend of HI is used as the baseline trend in this article, noting that a downward trend can be readily obtained by taking the reciprocal of an upward trend.

Compared to Eq. (5), the influence of each difference on the overall monotonicity is evaluated in Eq. (9). The weight of each difference is directly linked to its magnitude. Additionally, penalties are applied to negative differences to ensure that the HI exhibits an upward trend.

Afterward, the MSI is defined as below:

$$\text{MSI}(F_i) = \text{mean}\left(\frac{|\sum_{k=1}^{N_j} f(d/df_{i,j})|}{N_j - 1}\right). \quad (10)$$

It should be noted that our proposed MSI is an extension form of the conventional monotonicity metric. When $p = -1$ and $\beta = 0$, it can be observed that $f(d/df_{i,j}) = 1$ when $d/df_{i,j} > 0$ and $f(d/df_{i,j}) = -1$ when $d/df_{i,j} \leq 0$, so MSI degenerates into Mon (i.e. Eq. (5)). An HI with a large MSI value is preferred, as it indicates better monotonicity and an obvious degradation trend.

Having established the MSI as above, it holds the potential to address the limitation of the conventional monotonicity index illustrated by the examples in Fig. 2. We use the MSI to evaluate the monotonicity of y_1 and y_2 with $p = -1$ and $\beta = 1$. The resulting values are $\text{MSI}(y_1) = 0$ and $\text{MSI}(y_2) = -5$, indicating that y_1 is preferred. Therefore, the MSI can correctly reflect the overall monotonicity of an HI, while Eq. (5) cannot.

In addition, MSI also possesses the ability to measure the level of noise present in an HI to some extent, referred to as “robustness” [38]. System degradation measurements often contain significant uncertainty and randomness due to various sources of noise, which may affect the accuracy of RUL prediction. Therefore, the constructed HI needs to exhibit a stable degradation trend, and the level of stability of an HI is called robustness.

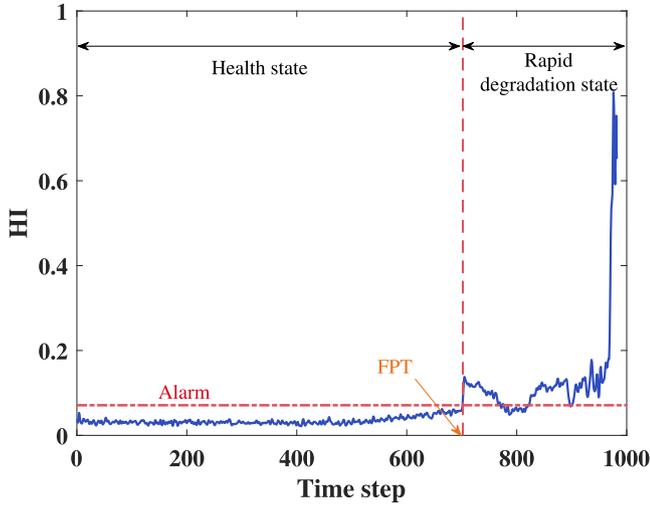


Fig. 3. The two-stage run-to-failure trajectory of a bearing.

The calculation method of MSI gives it the ability to assess the level of noise. Typically, more intense noise leads to larger fluctuations in HI, which also means larger and more negative differences. These negative differences are penalized according to their magnitude, as indicated in Eq. (9). Consequently, larger or more negative differences lead to a lower value of MSI. Therefore, MSI can reflect the noise level in an HI to some extent. The greater the value of p , the less tolerant the MSI is to noise in an HI. In contrast, Eq. (5) does not reflect robustness due to its treating differences with varying magnitudes equally.

Another example is used to illustrate the ability of MSI to reflect the robustness of an HI. Another HI in Fig. 2, $y_3 = 0.5t + 0.3 \times (-1)^t + 2$, where $t = 0, 1, \dots, 20$, exhibits the same overall trend as y_1 , but with a smaller level of noise. In this case, $\text{MSI}(y_3) = 9$ when $p = 2$ and $\beta = 1$, which performs better than y_1 . However, if we use Eq. (5) for calculation, both of these results are 0, demonstrating that MSI has an advantage over the traditional monotonicity in this aspect.

3.1.3. Metric selection for FPT detection and RUL prediction

In order to create an HI capable of detecting FPT and predicting RUL, specific metrics need to be chosen from the aforementioned options. These selected metrics will be used to characterize the desired attributes of the HI.

To ensure the HI possesses the ability to indicate the FPT, the detectability metric (Eq. (8)) must be employed. As illustrated in Fig. 3, a typical run-to-failure trajectory of a bearing demonstrates a clear transition between a healthy state and a rapid degradation state. This transition is characterized by an abrupt jump known as the FPT. To ensure reliable RUL prediction, the HI must possess the capability to identify the FPT. Consequently, the detectability metric is selected to provide the constructed HI for the ability to indicate the FPT.

Besides detectability, the constructed HI needs to perform well on RUL prediction. Therefore, additional metrics, namely monotonicity, trendability, and prognosability, are taken into account. However, due to the limitations identified with monotonicity (as discussed in Section 3.1.2), it has been substituted with the MSI. As a result, the construction of the HI will employ MSI, trendability, prognosability, and detectability as objectives.

Then Eq. (4) is transformed as follows:

$$\begin{aligned} \max_{w \in \mathbb{R}^M} g(w) &= [\text{MSI}(\text{HI}), \text{Tre}(\text{HI}), \text{Pro}(\text{HI}), \text{Det}(\text{HI})] \\ &= [\text{MSI}(w^T \mathbf{F}), \text{Tre}(w^T \mathbf{F}), \dots, \text{Pro}(w^T \mathbf{F}), \text{Det}(w^T \mathbf{F})]. \end{aligned} \quad (11)$$

Table 1

The extracted time-domain and frequency-domain features.

No.	Feature	Definition
1	Mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
2	Standard Deviation	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
3	Root Mean Square	$x_{rms} = \sqrt{x_{en}}$
4	Peak	$x_p = \max(x_i)$
5	Peak to Peak	$x_{p2p} = \max(x_i) - \min(x_i)$
6	Mean Absolute	$x_{ma} = \frac{1}{n} \sum_{i=1}^n x_i $
7	Energy	$x_{en} = \frac{1}{n} \sum_{i=1}^n x_i^2$
8	Clearance Factor	$x_{clf} = \frac{n^2 x_p}{(\sum_{i=1}^n \sqrt{ x_i })^2}$
9	Crest factor	$x_{crf} = x_p / x_{rms}$
10	Impulse factor	$x_{if} = x_p / \bar{x} $
11	Kurtosis	$x_{ku} = \frac{1}{n} \sum_{i=1}^n x_i^4$
12	SINAD	$x_{\text{SINAD}} = 10 \log(P_{\text{signal}} / P_{\text{noise}})$
13	SNR	$x_{\text{SNR}} = 10 \log \frac{P_{\text{signal}} + P_{\text{noise}} + P_{\text{distortion}}}{P_{\text{noise}} + P_{\text{distortion}}}$
14	Shape Factor	$x_{sf} = x_{rms} / \bar{x} $
15	Skewness	$x_{sk} = \frac{1}{n} \sum_{i=1}^n x_i^3$
16	Mean Frequency	$x_{mf} = \frac{1}{N \sum_{i=1}^N X_i}$
17	Frequency Center	$x_{fc} = \frac{\sum_{i=1}^N f_i X_i}{\sum_{i=1}^N X_i}$
18	Standard Deviation Frequency	$x_{stdf} = \frac{\sum_{i=1}^N (f_i - x_{fc})^2 X_i}{\sum_{i=1}^N X_i}$
19	Mean Square of Frequency	$x_{msf} = \frac{\sum_{i=1}^N (f_i)^2 X_i}{\sum_{i=1}^N X_i}$

3.2. Feature extraction and selection

Due to the low information density and redundancy in raw monitoring data, it is necessary to extract and select features to be considered as candidate elements for constructing the HI. Different approaches for extracting statistical features can be considered, such as time-domain analysis, frequency-domain analysis, time-frequency domain analysis, and so on. In this study, 31 features are extracted from the raw data, consisting of 13 time-domain features, 6 frequency-domain features, 8 wavelet packet energy features (3-depth decomposition of wavelet packet [39]), and 4 EMD energy features [17]. The specific names and formulas of the 13 time-domain features and 6 frequency-domain features are presented in Table 1. All 31 features are labeled from No. 1 to No. 31.

In order to ensure that only the most relevant features are utilized for HI construction, it is necessary to evaluate the extracted features using appropriate metrics. To maintain consistency with the objectives of HI construction, the metrics determined in Section 3.1 are also employed to assess the performance of the extracted features. Features that perform poorly on these metrics are considered unsuitable and are consequently eliminated from consideration for inclusion in the HI.

Furthermore, it is important to consider the possibility of redundancy among the features, as some of them may be highly similar [13]. To address this issue, a correlation matrix is established before HI construction. If one feature is too similar to another one, only one of them will be kept. In this way, redundancy and unnecessary waste of computing resources can be avoided.

Remark 2. The feature extraction approach discussed in this section is presented as an illustrative example and can be interchanged. When implementing the framework proposed in this study, researchers have the flexibility to select the most suitable features based on the specific application requirements. Furthermore, the integration of novel and advanced feature extraction methods has the potential to enhance the overall performance of the proposed approach.

3.3. Method of solving the MOP

This subsection presents a method for obtaining the optimal solution of the MOP converted from the HI construction task (i.e., Eq. (11)). For the convenience of presentation, the following definitions regarding multi-objective optimization are stated [32]:

Definition 1 (Pareto dominance). For $\forall w, w' \in \mathbb{R}^{n_f}$, w is said to dominate w' (denote as $w > w'$) if and only if:

$$\begin{aligned} \forall i \in \{1, 2, \dots, n_m\} : M_i(w^T F) &\geq M_i(w'^T F) \\ \wedge \exists i \in \{1, 2, \dots, n_m\} : M_i(w^T F) &> M_i(w'^T F) \end{aligned}$$

Definition 2 (Pareto optimality). Suppose $w \in \mathbb{R}^{n_f}$, w is called Pareto-optimal and is a solution of Eq. (11) if and only if:

$$\{\nexists w' \in \mathbb{R}^{n_f} | w' > w\} \quad (12)$$

Definition 3 (Pareto optimal set). The set of all Pareto-optimal solutions is called Pareto optimal set (also called Pareto set for short) as follows:

$$W^* := \{w, w' \in \mathbb{R}^{n_f} | \nexists w' > w\} \quad (13)$$

Definition 4 (Pareto optimal front). A set containing the value of objective functions for Pareto solutions set:

$$G := \{g(w) | w \in W^*\} \quad (14)$$

Some studies weight the metrics to form a single objective function in order to obtain the optimal solution [9,23,26]. However, these weights are subjectively determined and need to be recalculated for different scenarios, resulting in inefficiency. Moreover, the subjective nature of this determination lacks consistent standards, leading to the instability of the effectiveness of the method across different tasks.

To address the aforementioned issue, we employ a combination of MOALO and EWM to solve Eq. (11) in our study. Initially, MOALO is employed to obtain the Pareto solution set. Subsequently, EWM, which incorporates the consideration of information entropy, is utilized to derive the optimal solution from the Pareto solution set. It is important to note that subjective determination is not involved in any part of the process.

We initially solve the MOP defined by Eq. (11) to obtain the Pareto optimal set W^* , which can be achieved by using multi-objective optimization algorithms. Many optimization algorithms are utilized to solve MOPs, with non-dominated sorting genetic algorithm-II (NSGA-II) [40] and multi-objective particle swarm optimization (MOPSO) [41] being the most popular in the literature. In this work, we apply MOALO as the optimization algorithm, which is a population-based optimization algorithm inspired by the predatory behavior of antlions. MOALO is selected due to its proven superiority in terms of accuracy and robustness when compared to NSGA-II, as well as its demonstrated performance for optimizing separated regions compared to MOPSO [32]. With the utilization of MOALO on Eq. (11), a Pareto optimal set of groups of fusion weights, W^* , can be obtained.

Since the Pareto optimal set is insufficient to achieve the target of finding the optimal set of fusion weights due to its inclusion of multiple solutions, the Pareto set W^* needs to be transformed into the optimal solution w^* . The performance of Pareto optimal solutions varies across different metrics, making it challenging to directly choose the best solution. In fact, this constitutes a multi-criteria decision-making problem [42].

EWM is one of the most popular ways to tackle multi-criteria decision-making problems, which can be utilized to obtain w^* from W^* . The fundamental concept underlying this method is to assign greater importance to a metric that contains more information, which can be achieved by quantifying the amount of information using information

entropy. Using this method, the Pareto optimal solutions can be ranked to obtain the optimal solution. The algorithm for obtaining the optimal group of HI fusion weights using EWM is presented below as Algorithm 1. The Pareto front, which consists of metric values of HIs that are fused by weighting the extracted features using fusion weights, is denoted as $G = [G_1, G_2, \dots, G_{n_s}]^T$, where $G_i = [g_{i,1}, g_{i,2}, g_{i,3}, g_{i,4}]^T$, $i = 1, 2, \dots, n_s$. Here, G_i is a vector of metric values, and $W^* = \{w_1, w_2, \dots, w_{n_s}\}$ denotes the Pareto optimal set. The optimal solution w^* can be obtained utilizing Algorithm 1.

Algorithm 1: The execution process of EWM.

Input: Pareto front G , Pareto optimal set W^*
Output: The best HI fusion weight vector w^*

```

1 for  $i \leftarrow 1$  to  $n_s$  do
2   for  $j \leftarrow 1$  to 4 do
3      $z_{i,j} \leftarrow \frac{g_{i,j} - \min(g_{1,j}, g_{2,j}, \dots, g_{n_s,j})}{\max(g_{1,j}, g_{2,j}, \dots, g_{n_s,j}) - \min(g_{1,j}, g_{2,j}, \dots, g_{n_s,j})}$ 
4      $p_{i,j} \leftarrow z_{i,j} / \sum_{i=1}^{n_s} z_{i,j}$ 
5   end
6 end
7 for  $j \leftarrow 1$  to 4 do
8    $E_j \leftarrow -\ln(s)^{-1} \sum_{i=1}^{n_s} p_{i,j} \ln p_{i,j}$ 
9 end
10 for  $j \leftarrow 1$  to 4 do
11    $\theta_j \leftarrow (1 - E_j) / (4 - \sum_{j=1}^4 E_j)$ 
12 end
13  $\Theta \leftarrow [\theta_1, \theta_2, \theta_3, \theta_4]^T$  for  $i \leftarrow 1$  to  $n_s$  do
14    $\text{Fit}_i \leftarrow \Theta^T G_i$ 
15 end
16  $i^* = \arg \max_i (\text{Fit}_i)$ 
17  $w^* = w_{i^*}$ 

```

The HI construction algorithm by MOALO and EWM is summarized as Algorithm 2.

Algorithm 2: The algorithm of HI construction.

Input: Raw run-to-failure data.
Output: The optimal HI.

Phrase 1: Calculate the optimal fusion weight off-line

- Construct a feature space consisting of the 31 features mentioned in Section 3.2, derived from raw run-to-failure training data.
- Compute the correlation among these 31 features, and delete redundant features.
- Compute the four metrics by Eqs. (6)–(10).
- Add the metric value of these features, and select the features that perform well on these metrics.
- Solve the MOP constructed by Eq. (11), obtain the Pareto optimal set and Pareto front of the MOP.
- Input the Pareto optimal set and the Pareto front into Algorithm 1, obtain the optimal fusion weights.

Phrase 2: Construct HI on-line

- while** The monitor system is running **do**
- Compute the features of the running system.
- Compute the HI by multiplying the features with the optimal fusion weights.
- end**

Remark 3. In fact, the approach proposed in this article is a framework for HI construction, and both the feature extraction part and the optimization objective can be replaced according to actual applications. Therefore, we believe that this approach has good scalability.

Table 2
Structure of the IMS dataset.

Dataset	Fault bearing	Accelerometers location	Monitoring points	Monitoring points after pre-processing
1	Bearing 3 & 4	x & y axes	2156	2000
2	Bearing 5	x-axis	984	982
3	Bearing 11	x-axis	6324	6323

Remark 4. The computational load of our approach is primarily centered on the offline phase, where the optimal fusion weights are obtained using the training set through MOALO and EWM; while in the online stage, only the required features need to be extracted from the monitoring data, and then the monitoring data is fused using the fusion weights. Consequently, this approach does not demand high computational efficiency.

4. Case studies on rotating machinery datasets

This section presents two case studies on two rotating machinery datasets to illustrate the superiority of the proposed framework. All experiments are performed in MATLAB R2021a.

4.1. The case study on the IMS bearing dataset

Generated by the NSF I/UCR for Intelligent Maintenance Systems (IMS), the IMS bearing dataset [43] has been widely adopted by many studies to demonstrate the validity of their approaches of diagnostics and prognostics [13,44,45].

4.1.1. Experiment setup

The test rig setup is shown in Fig. 4, which comprised four ZA-2115 double row bearings installed on a shaft, with a constant rotation speed of 2000 RPM and a radial load of 6000 lbs. High-sensitivity quartz ICP accelerometers (PCB 353B33) were attached to the bearing housing to collect acceleration data at a sampling frequency of 20480 Hz. Data were collected every 10 min for 1 s, and the test would be terminated if any bearing failed. Three tests were conducted, resulting in the collection of data from a total of 12 bearings. As illustrated in Table 2, 4 bearings failed during the tests while the other 8 bearings held up until the end of the tests. In this case study, only run-to-fail bearing data are used because the others do not contain the information of failure.

The IMS dataset contains outliers and inconsistent timestamps, necessitating pre-processing prior to utilization. Note that the interval between the 156th monitoring point (October 23, 2003, 10:14:13) and the 157th monitoring point (October 29, 2003, 14:39:46) in dataset 1 is six days instead of the stated 10 min, and data show inconsistent characteristics before and after. To eliminate the interference of this inconsistency, data collected at the first 156 monitoring points are removed. Additionally, the final two monitoring points of dataset 2 and the last one of dataset 3 are also removed due to issues with measurements most likely brought on by bearing failures. Their monitoring values only consist of 0.2 and 0, which are abnormal. The results of the aforementioned pre-processing are also shown in Table 2.

Data collected from bearing 3 and 5 are selected as the training data, while data from bearing 11 is chosen as the test data. Bearing 4 is excluded from the training and test data as it exhibits a different degradation trajectory, as depicted in Fig. 9. One possible explanation for this inconsistency relates to the data set description, which reported that bearings 3 and 4 failed simultaneously during test 1. However, considering the low probability of two bearings failing exactly at the same time, it is more plausible to attribute the cessation of test 1 primarily to the failure of bearing 3, indicating that bearing 4 was not completely damaged and thus exhibits a different degradation trajectory.

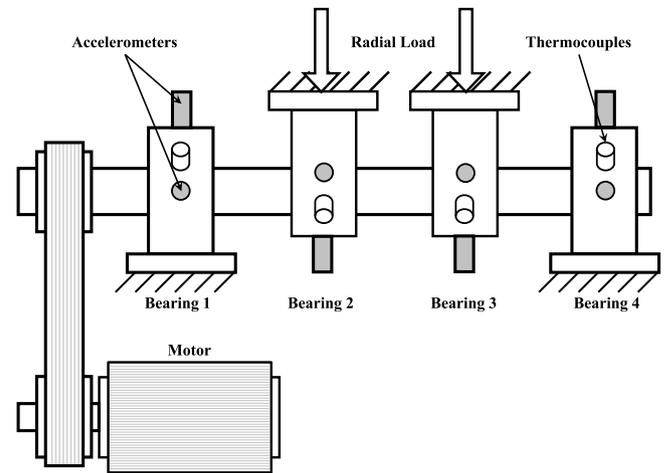


Fig. 4. Bearing test rig and sensor placement illustration.

4.1.2. Applying our approach on the IMS dataset

A comprehensive set of 31 features is extracted from raw data, which includes 13 time-domain features, 6 frequency-domain features, 8 wavelet packet energy features, and 4 EMD features, as described in Section 3.2. To ensure comparability, max–min normalization is applied to each feature. The features of bearing 3 after normalization are as Fig. 5. Most features are observed to behave differently during the health phase than they do during degradation, suggesting their potential to detect a failure and that their fusion can be used to construct an HI.

The appropriate features should be selected from the 31 features as candidate features for the construction of HI. The correlation coefficient matrix of the 31 features is calculated to reduce redundancy, and the result is shown in Fig. 6, as a heatmap. If the correlation coefficient between two features is above 0.9, the one with the worse result on the sum of MSI, trendability, prognosability, and detectability will be eliminated. Fourteen features remain after doing this.

To select the features that are suitable for HI construction, the values of the four metrics of the remaining 14 features are calculated by Eq. (6)–(10), with $\alpha = 5$ in Eq. (8), $p = 2$ and $\beta = 2$ in Eq. (9). The values of all metrics are normalized into [0, 1], and the results are shown in Fig. 7. The characteristics of features exhibited in Fig. 5 are corroborated by Fig. 7. For instance, feature No. 8 exhibits a significant jump in the near-failure phase, resulting in a high detectability score, with a poor MSI value due to the huge fluctuation. The values of all four evaluation metrics for each feature are summed together, and a threshold of 2.5 is chosen to select the most promising features for HI construction. The results indicate that features No. 5, 6, 14, 16, and 20 demonstrated superior performance across all four metrics when compared to other features, thereby they are selected for fusion.

To obtain the Pareto solutions, MOALO is executed. The number of solutions is set to 100, and the number of iterations is set to 200. The bound of weights is $[-5, 5]$, and $\sum_{i=1}^5 w_i = 1$. Fig. 8 illustrates the optimization results obtained using MOALO, where each data point represents the HI corresponding to the optimized set of weights. The 3D coordinates and color of each point in the plot represent the metric values for the corresponding HI.

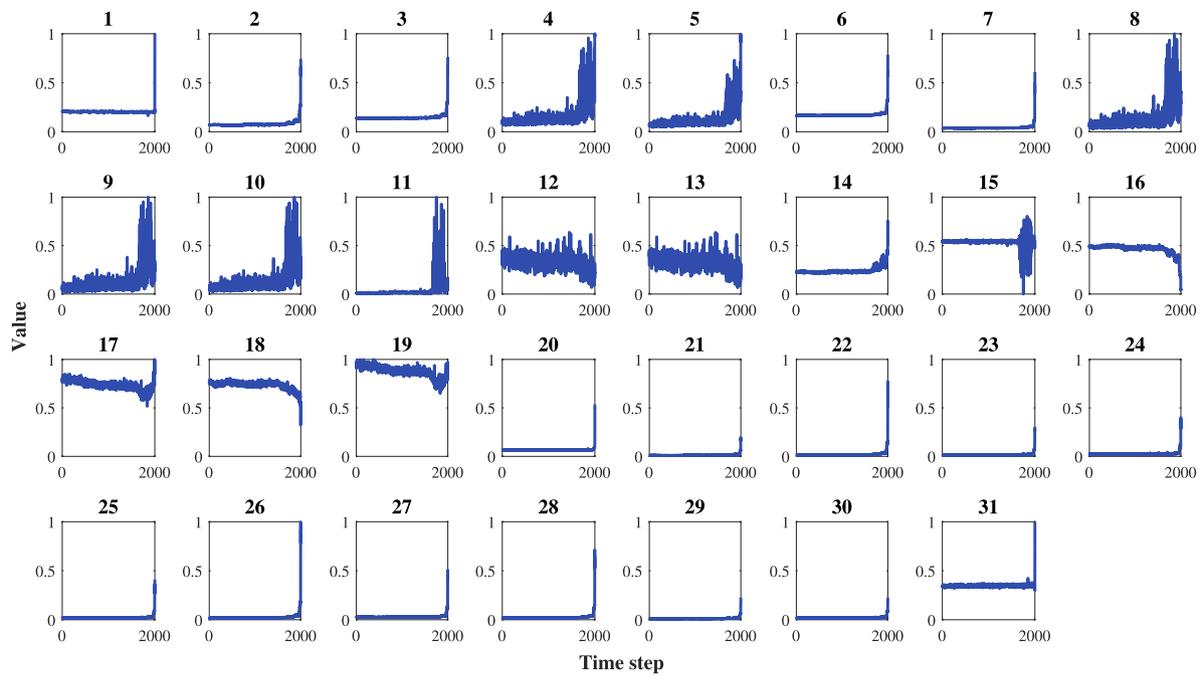


Fig. 5. All normalized features of bearing 3.

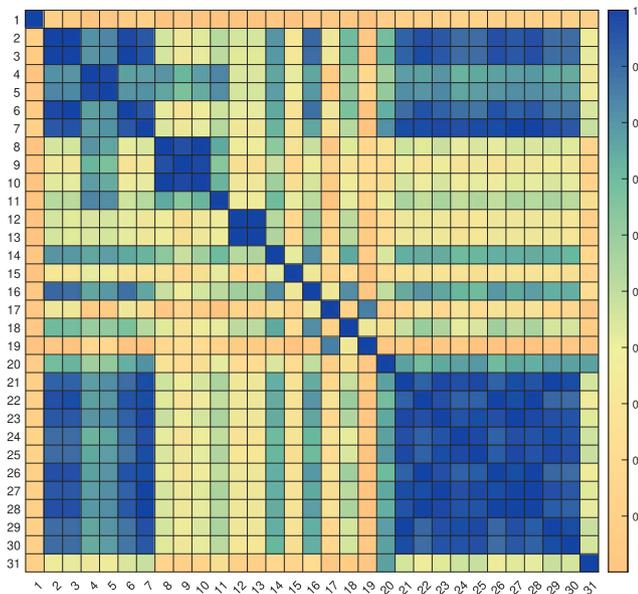


Fig. 6. Correlation heatmap of all extracted bearing features.

As all these HIs are not dominated by each other, EWM is applied to rank them. Following the procedure in Algorithm 1, the importance weight of each metric is obtained and shown in Table 3. In this way, an optimal HI fusion weights vector with the highest fitness is obtained, as shown in Table 4.

4.1.3. Results and discussion

Based on the HI fusion weights vector in Table 4, HIs of bearing 3, 4, 5, and 11 are constructed, as shown in Fig. 9. All bearings exhibited similar degradation trends, except for bearing 4. This also to some extent confirms our speculation in 4.1.1.

To show the superiority of the constructed HI, the selected features and another HI constructed in [9] are compared to our HI. The HI constructed in [9] applies Eq. (5)–(7) and robustness in [24] being

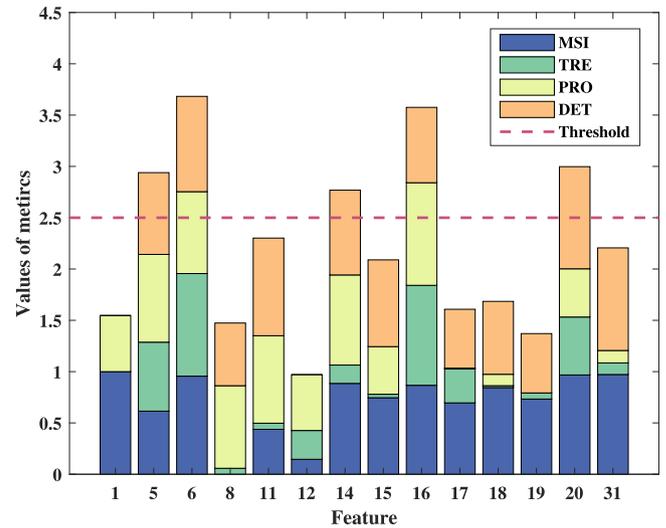


Fig. 7. The metric values of the bearing features.

Table 3

The importance weight of the four metrics of the bearing experiment.

Metric	MSI	Tre	Pro	Det
Importance weight	0.0298	0.2989	0.3877	0.2837

Table 4

The optimal HI fusion weights of bearing features.

Feature	No. 5	No. 6	No. 14	No. 16	No. 20
Fusion weight	-0.0332	2.7027	0.0884	-0.2499	-1.5080

objectives, and the weights of these objectives were determined subjectively by humans. As the HI is constructed based on a genetic algorithm, it is named GA-HI in this paper.

The four metrics which are used as the objectives of the MOP, i.e., MSI, trendability, prognosability, and detectability are utilized to

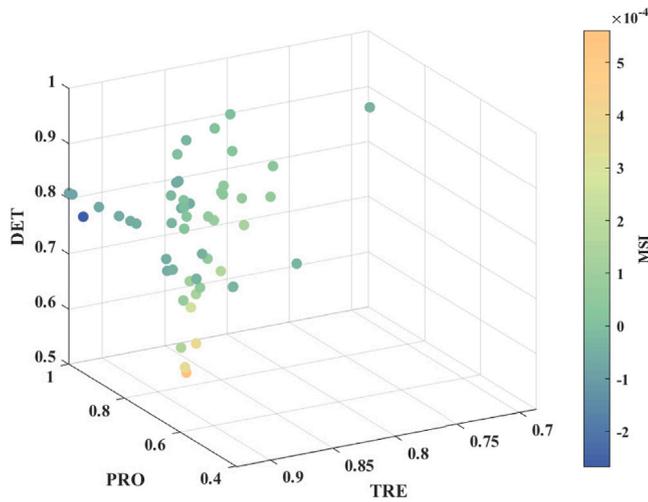


Fig. 8. The Pareto front obtained by MOALO.

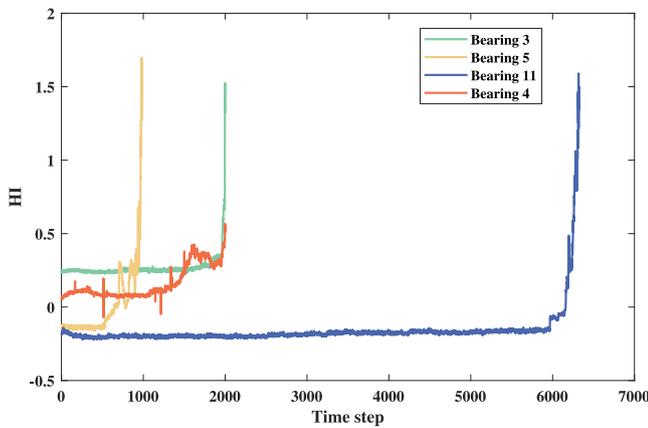


Fig. 9. The constructed HIs of bearing 3, 4, 5, 11.

compare the performance of these HIs. Besides, the fitness synthesized by each metric based on the weights in Table 3 is also calculated. The results of the evaluation are shown in Table 5, as well as the evaluation results on the 5 selected features. The results show that the constructed HI does not perform best under every metric, but achieves the best performance in terms of the combined fitness than the other features, displaying the ability of the proposed HI to indicate the degradation of bearings.

To demonstrate the ability of MSI to indicate the robustness level of HIs, another metric to measure robustness is introduced from [24], which we refer to as “Rob”. Even if Rob is not included as an objective in the MOP, the proposed HI still performs well and is comparable to the results of features No. 14 and No. 16. In addition, our HI outperforms GA-HI in most metrics except MSI, even when Rob is taken as an objective of the GA-HI. These results demonstrate that MSI has the ability to indicate the degree of robustness in HIs to some extent. However, our constructed HI is not as good as GA-HI in MSI, which probably happens due to the trade-off between detectability.

To demonstrate the effectiveness of the constructed HI in detecting FPT detection, the FPT detection method described in [31] was employed on the data of bearing 11. The results of FPT detection on the selected features and the proposed HI in this work are presented in Fig. 10 and Table 6. The results reveal that the FPT value obtained from the proposed HI is 5967, which corresponds to the results on feature No. 6 and feature No. 16 and occurs earlier than in the case of the other 3 features. This finding indicates that the proposed HI can

provide accurate FPT detection. Moreover, Table 5 suggests that the constructed HI exhibits the highest detectability compared to all the other features. These results collectively indicate that the proposed HI construction approach possesses the ability to precisely identify the FPT of a system, thereby enabling a more distinct differentiation between the health and failure phases.

Experiments are also set to demonstrate the effective help of the constructed HI on RUL prediction. The Wiener-process-based RUL prediction method proposed in [46] is applied for RUL prediction. Specifically, we employ an exponential model, *i.e.*, $\mu(t; \theta) = \theta e^{\theta t}$.

In order to demonstrate the superiority of the proposed method in RUL prediction, three groups of comparative experiments were conducted. The first group is designed to illustrate the importance of FPT detection for RUL prediction, thus comparing the results of RUL prediction without considering FPT detection and with considering FPT detection. The second and third comparative objects are respectively the optimal performing feature 6 selected from the chosen features, and the previously mentioned HI constructed in [9], both of which were predicted under the consideration of FPT.

The results of RUL prediction are displayed in Fig. 11. Because of the significant deviation in the initial stage of RUL prediction associated with the specific RUL prediction method, we commence displaying the RUL prediction results from 6125 rather than from FPT (5967), ensuring a clearer presentation in Fig. 11 of the RUL prediction outcomes. The significant deviation of the RUL prediction result obtained without considering FPT detection from the actual RUL highlights the necessity of FPT detection in RUL prediction. Furthermore, it can be observed that the result on the proposed HI with the consideration of FPT detecting is closer to the actual RUL than feature No. 6 and GA-HI. Furthermore, during the end period of the life-cycle, the RUL prediction result derived from our HI exhibits a high degree of proximity to the true RUL, showing the validity of constructed HI on RUL prediction.

4.2. The case study on a gearbox dataset

The gear life test dataset is from [47], which was performed by a gear contact fatigue test rig provided by the University of Newcastle in the UK. Two gears made of 20Cr2Ni4 A, with tooth numbers of 27 and 25 respectively, were continuously monitored by accelerometers mounted on the gearbox cover. The sampling frequency, sampling duration, and sampling period were set to 25 kHz, 0.2 s, and 1 min, respectively. The experiments were terminated when the amplitude of the vibration signal exceeded a predetermined threshold, indicating gear failure. A total of four run-to-failure tests are accessible, outlined in Table 7. Although the testing conditions for each assessment may vary slightly, given the lack of significant differences, it can be considered that their features provide a similar indication of the health status of gears. Consequently, tests 1, 2, and 3 are chosen for training, while dataset 4 is selected for testing.

Following the feature extraction and selection approach outlined in Section 4.1, six key features, namely Energy, Kurtosis, SINAD, Shape Factor, Wavelet Packet Energy 30, and Intrinsic Mode Function Energy 3, are identified for fusion. After the features are normalized and smoothed, they are input into Eq. (11), which is then solved by MOALO and EWM. The resultant optimal fusion weights are shown in Table 8, alongside the fusion weights from GA-HI for comparative analysis.

The constructed HI is presented in Fig. 12 and compared with other HIs. Comparison results among various HIs are delineated in Table 9. Our constructed HI outperforms others on three key metrics, and achieves the optimal fitness. It is worth noting that our HI performs best in Rob, despite it not being a specified objective in the MOP, which further demonstrates the ability of MSI to indirectly indicate robustness. Furthermore, Our proposed HI promptly detects FPT, aiding in the monitoring of system health. The GA-HI exhibits almost the same characteristics as Energy due to its emphasis on monotonicity (given a weighted importance of 0.9 in [9]). Such targeted emphasis demands

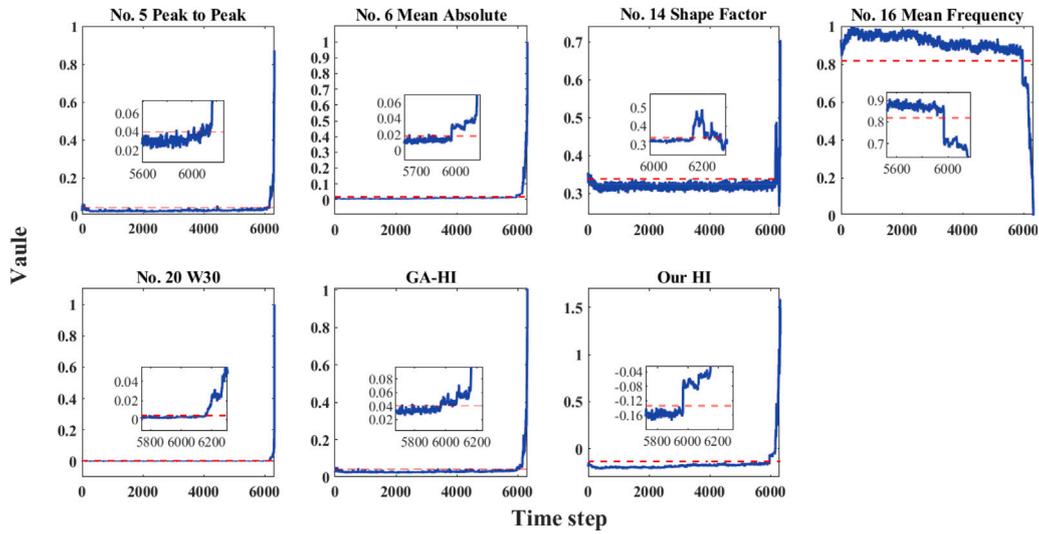


Fig. 10. FPT detection results of bearing 11 by different HIs.

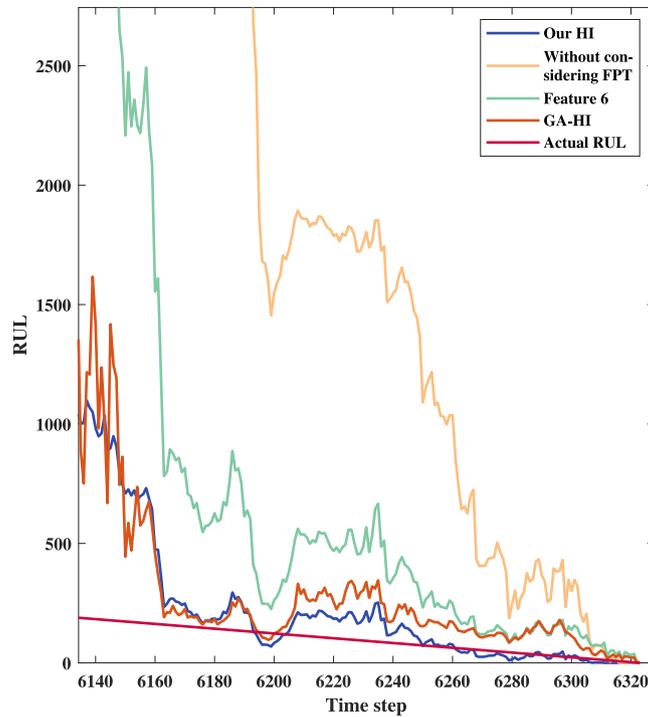


Fig. 11. RUL prediction results of different bearing HIs.

Table 5
Comparison results among different features and HIs for bearing dataset.

Feature	MSI	Tre	Pro	Det	Rob	Fitness
No. 5	-7.2601×10^{-4}	0.5458	0.8109	0.7739	0.9244	0.6970
No. 6	1.0352×10^{-5}	0.8153	0.7670	0.8907	0.9368	0.7937
No. 14	-1.4379×10^{-4}	0.1573	0.8437	0.7922	0.9857	0.5988
No. 16	-1.8430×10^{-4}	0.7946	0.9602	0.7044	0.9893	0.8095
No. 20	3.6312×10^{-5}	0.4665	0.4533	0.9545	0.9142	0.5860
GA-HI	7.2362×10^{-6}	0.6970	0.8253	0.4722	0.9465	0.6622
Our HI	-4.8171×10^{-5}	0.8180	0.9208	0.7752	0.9798	0.8214

Table 6
The results of FPT detection on different bearing HIs.

Feature	No. 5	No. 6	No. 14	No. 16	No. 20	GA-HI	Our HI
FPT	6115	5967	6159	5967	6161	5967	5967

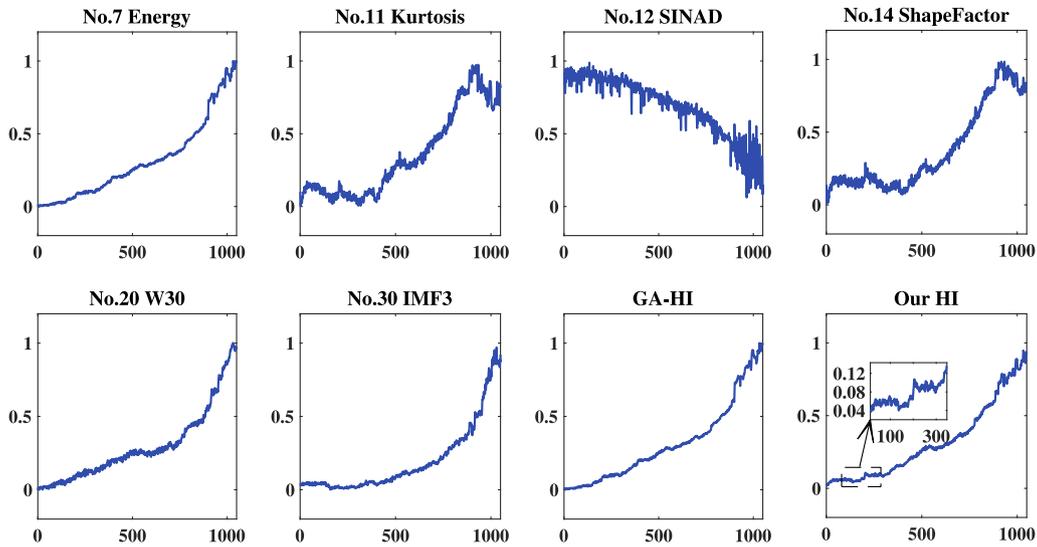


Fig. 12. HIs of gear 4.

Table 7

The detailed information of the gear dataset from [47].

No. of test	1	2	3	4
Load (N m)	4400	4070	4070	4400
Rotational speed (rpm)	1100	800	800	1100
The number of samples	1050	2820	2820	1050
Role		Training data		Test data

Table 8

The optimal HI fusion weights of gear features.

Feature	No. 7	No. 11	No. 12	No. 14	No. 20	No. 30
Our HI	0.3700	-0.0098	0.0046	0.1596	0.2163	0.2592
GA-HI	0.9992	5.162×10^{-4}	1.695×10^{-4}	8.186×10^{-4}	-3.476×10^{-4}	-3.089×10^{-4}

Table 9

Comparison results among different features and HIs for the gear dataset.

Feature	MSI	Tre	Pro	Det	Rob	FPT	Fitness
No. 7	7.8780×10^{-6}	0.9033	0.9950	0.6748	0.9921	201	0.8124
No. 11	-1.2733×10^{-4}	0.0128	0.3461	0.5978	0.9768	445	0.2534
No. 12	-1.335×10^{-3}	0.3325	0.7729	0.5824	0.9777	404	0.5708
No. 14	-1.4537×10^{-4}	0.1094	0.3985	0.4817	0.9840	646	0.2951
No. 20	-1.5020×10^{-5}	0.9114	0.9866	0.3373	0.9866	916	0.7873
No. 30	-1.2707×10^{-5}	0.8167	0.8417	0.6932	0.9784	502	0.7040
GA-HI	7.893×10^{-6}	0.9035	0.9949	0.6751	0.9922	201	0.8124
Our HI	2.0830×10^{-6}	0.9670	0.9900	0.7206	0.9949	201	0.8232

domain expertise that can impact utility in practical applications. Conversely, our approach reduces the demand for specialized knowledge, presenting direct applicability.

Comparisons of RUL predictions using various HIs are conducted, similar to Section 4.1. The RUL prediction without considering FPT detection is assessed, with a specific focus on the RUL prediction results of GA-HI. The predicted RUL derived from the proposed HI closely aligns with the actual RUL, as demonstrated in Fig. 13. This alignment is improved compared to the GA-HI approach, highlighting the efficacy of our HI in taking into account various metrics comprehensively. However, the difference in our prediction results is small whether or not we consider FPT, which may be due to some stable running stage data being removed from the dataset (this may also be the reason why the dataset we obtained is different from what was annotated in [47]). Nevertheless, these experimental results are sufficient to demonstrate

that the suggested framework can generate a health indicator (HI) that aids in fault propagation time (FPT) detection and remaining useful life (RUL) prediction to a certain extent.

5. Conclusion

In this work, a novel HI that enables FPT detection, and subsequent RUL prediction is constructed. To construct this HI, four HI assessment metrics, namely MSI, trendability, prognosability, and detectability, are utilized as optimization objectives. The proposed approach takes FPT detection into account by taking detectability as an objective of the MOP. The introduction of MSI overcomes the limitation of the traditional monotonicity metric in capturing the overall trend of an HI. To solve the established MOP effectively, we utilize the combination of MOALO and EWM, which eliminates subjectivity in manually determining the importance of assessment metrics. The validity of the

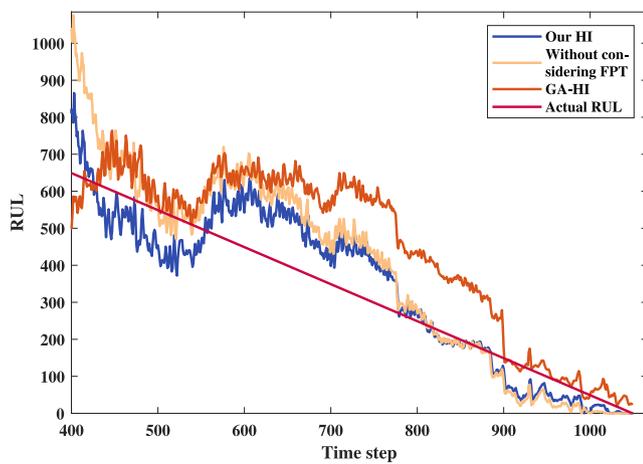


Fig. 13. RUL prediction results of different gear HIs.

constructed HI is demonstrated on two rotating machinery datasets. This approach has the potential to be employed for assessing the health of diverse systems such as turbfan engines, machine tools, gearboxes, and vehicle exhaust monitoring equipment.

However, there are areas for further improvement in this approach. Firstly, the current fusion of features is accomplished through linear weighting, potentially neglecting the nonlinear and intricate relationships between features. Secondly, the constructed HI remains fixed for a particular type of system, overlooking the heterogeneity among individuals. Thirdly, this approach has only been validated on vibration signals of rotating machinery, and its effectiveness on other systems such as lithium batteries and machine tools needs further investigation. To enhance the representation ability of the health status of the HI, future work will concentrate on addressing these challenges.

CRediT authorship contribution statement

Yun-Sheng Zhao: Data curation, Formal analysis, Investigation, Methodology, Software, Validation. **Pengfei Li:** Formal analysis, Investigation, Methodology, Writing – original draft. **Yu Kang:** Conceptualization, Project administration, Resources, Supervision. **Yun-Bo Zhao:** Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 62173317, 62033012), and the Key Research and Development Program of Anhui (No. 202104a05020064).

References

- [1] S. Zheng, K. Ristovski, A. Farahat, C. Gupta, Long short-term memory network for remaining useful life estimation, in: 2017 IEEE International Conference on Prognostics and Health Management, ICPHM, Dallas, TX, USA, 2017, pp. 88–95, <http://dx.doi.org/10.1109/ICPHM.2017.7998311>.
- [2] H. Zhang, M. Chen, J. Shang, C. Yang, Y. Sun, Stochastic process-based degradation modeling and RUL prediction: From Brownian motion to fractional Brownian motion, *Sci. China Inf. Sci.* 64 (7) (2021) 171201, <http://dx.doi.org/10.1007/s11432-020-3134-8>.
- [3] P. Li, Z. Zhang, R. Grosu, Z. Deng, J. Hou, Y. Rong, R. Wu, An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries, *Renew. Sustain. Energy Rev.* 156 (2022) 111843, <http://dx.doi.org/10.1016/j.rser.2021.111843>.
- [4] R. Jin, Z. Chen, K. Wu, M. Wu, X. Li, R. Yan, Bi-LSTM-based two-stream network for machine remaining useful life prediction, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–10, <http://dx.doi.org/10.1109/TIM.2022.3167778>.
- [5] J. Zhang, C. Huang, M.-Y. Chow, X. Li, J. Tian, H. Luo, S. Yin, A data-model interactive remaining useful life prediction approach of lithium-ion batteries based on PF-BiGRU-TSAM, *IEEE Trans. Ind. Inform.* (2023) <http://dx.doi.org/10.1109/TII.2023.3266403>.
- [6] J. Zhang, Y. Jiang, X. Li, H. Luo, S. Yin, O. Kaynak, Remaining useful life prediction of lithium-ion battery with adaptive noise estimation and capacity regeneration detection, *IEEE/ASME Trans. Mechatronics* 28 (2) (2022) 632–643, <http://dx.doi.org/10.1109/TMECH.2022.3202642>.
- [7] H. Sun, D. Cao, Z. Zhao, X. Kang, A hybrid approach to cutting tool remaining useful life prediction based on the Wiener process, *IEEE Trans. Reliab.* 67 (3) (2018) 1294–1303, <http://dx.doi.org/10.1109/TR.2018.2831256>.
- [8] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A model-based method for remaining useful life prediction of machinery, *IEEE Trans. Reliab.* 65 (3) (2016) 1314–1326, <http://dx.doi.org/10.1109/TR.2016.2570568>.
- [9] G. Qiu, Y. Gu, J. Chen, Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model, *Measurement* 150 (2020) 107097, <http://dx.doi.org/10.1016/j.measurement.2019.107097>.
- [10] Q. Ni, J.C. Ji, K. Feng, Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network, *IEEE Trans. Ind. Inform.* 19 (2) (2023) 1301–1311, <http://dx.doi.org/10.1109/TII.2022.3169465>.
- [11] L. Guo, Y. Yu, A. Duan, H. Gao, J. Zhang, An unsupervised feature learning based health indicator construction method for performance assessment of machines, *Mech. Syst. Signal Process.* 167 (2022) 108573, <http://dx.doi.org/10.1016/j.ymssp.2021.108573>.
- [12] S. Zhang, Q. Zhai, X. Shi, X. Liu, A Wiener process model with dynamic covariate for degradation modeling and remaining useful life prediction, *IEEE Trans. Reliab.* 72 (1) (2023) 214–223, <http://dx.doi.org/10.1109/TR.2022.3159273>.
- [13] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to RUL prediction, *Mech. Syst. Signal Process.* 104 (2018) 799–834, <http://dx.doi.org/10.1016/j.ymssp.2017.11.016>.
- [14] A. Aasi, R. Tabatabaei, E. Aasi, S.M. Jafari, Experimental investigation on time-domain features in the diagnosis of rolling element bearings by acoustic emission, *J. Vib. Control* 28 (19–20) (2022) 2585–2595, <http://dx.doi.org/10.1177/10775463211016130>.
- [15] T. Yan, D. Wang, B. Hou, Z. Peng, Generic framework for integration of first prediction time detection with machine degradation modelling from frequency domain, *IEEE Trans. Reliab.* 71 (4) (2021) 1464–1476, <http://dx.doi.org/10.1109/TR.2021.3087698>.
- [16] Y. Duan, X. Cao, J. Zhao, X. Xu, Health indicator construction and status assessment of rotating machinery by spatio-temporal fusion of multi-domain mixed features, *Measurement* 205 (2022) 112170, <http://dx.doi.org/10.1016/j.measurement.2022.112170>.
- [17] Z. Gao, Y. Liu, Q. Wang, J. Wang, Y. Luo, Ensemble empirical mode decomposition energy moment entropy and enhanced long short-term memory for early fault prediction of bearing, *Measurement* 188 (2022) 110417, <http://dx.doi.org/10.1016/j.measurement.2021.110417>.
- [18] Q. Ni, J.C. Ji, K. Feng, B. Halkon, A Fault Information-guided Variational Mode Decomposition (FIVMD) method for rolling element bearings diagnosis, *Mech. Syst. Signal Process.* 164 (2022) 108216, <http://dx.doi.org/10.1016/j.ymssp.2021.108216>.
- [19] X. Yan, M. Jia, Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings, *Mech. Syst. Signal Process.* 122 (2019) 56–86, <http://dx.doi.org/10.1016/j.ymssp.2018.12.022>.
- [20] J. Zheng, S. Cao, H. Pan, Q. Ni, Spectral envelope-based adaptive empirical Fourier decomposition method and its application to rolling bearing fault diagnosis, *ISA Trans.* 129 (2022) 476–492, <http://dx.doi.org/10.1016/j.isatra.2022.02.049>.
- [21] S. Buchaiah, P. Shakyia, Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection, *Measurement* 188 (2022) 110506, <http://dx.doi.org/10.1016/j.measurement.2021.110506>.

- [22] A. Hu, J. Sun, L. Xiang, Y. Xu, Rotating machinery fault diagnosis based on impact feature extraction deep neural network, *Meas. Sci. Technol.* 33 (11) (2022) 114004, <http://dx.doi.org/10.1088/1361-6501/ac7eb1>.
- [23] J. Coble, J.W. Hines, Identifying optimal prognostic parameters from data: a genetic algorithms approach, in: *Annual Conference of the PHM Society*, vol. 1, San Diego, CA, USA, 2009, URL: <https://www.papers.phmsociety.org/index.php/phmconf/article/view/1404>.
- [24] B. Zhang, L. Zhang, J. Xu, Degradation feature selection for remaining useful life prediction of rolling element bearings, *Qual. Reliab. Eng. Int.* 32 (2) (2016) 547–554, <http://dx.doi.org/10.1002/qre.1771>.
- [25] P. Baraldi, G. Bonfanti, E. Zio, Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics, *Mech. Syst. Signal Process.* 102 (2018) 382–400, <http://dx.doi.org/10.1016/j.ymssp.2017.09.013>.
- [26] Z. Chen, D. Zhou, E. Zio, T. Xia, E. Pan, A deep learning feature fusion based health index construction method for prognostics using multiobjective optimization, *IEEE Trans. Reliab.* (2022) 1–15, <http://dx.doi.org/10.1109/TR.2022.3215757>.
- [27] T. Yan, D. Wang, T. Xia, E. Pan, Z. Peng, L. Xi, New shapeness property and its convex optimization model for interpretable machine degradation modeling, *IEEE Trans. Reliab.* (2022) 1–13, <http://dx.doi.org/10.1109/TR.2022.3190514>.
- [28] K. Liu, A. Chehade, C. Song, Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis, *IEEE Trans. Autom. Sci. Eng.* 14 (3) (2017) 1504–1514, <http://dx.doi.org/10.1109/TASE.2015.2446752>.
- [29] H. Cheng, X. Kong, Q. Wang, H. Ma, S. Yang, The two-stage RUL prediction across operation conditions using deep transfer learning and insufficient degradation data, *Reliab. Eng. Syst. Saf.* 225 (2022) 108581, <http://dx.doi.org/10.1016/j.res.2022.108581>.
- [30] D. Wang, Q. Miao, Q. Zhou, G. Zhou, An intelligent prognostic system for gear performance degradation assessment and remaining useful life estimation, *J. Vib. Acoust.* 137 (2) (2015) <http://dx.doi.org/10.1115/1.4028833>.
- [31] N. Li, Y. Lei, J. Lin, S.X. Ding, An improved exponential model for predicting remaining useful life of rolling element bearings, *IEEE Trans. Ind. Electron.* 62 (12) (2015) 7762–7773, <http://dx.doi.org/10.1109/TIE.2015.2455055>.
- [32] S. Mirjalili, P. Jangir, S. Saremi, Multi-objective ant lion optimizer: A multi-objective optimization algorithm for solving engineering problems, *Appl. Intell.* 46 (2017) 79–95, <http://dx.doi.org/10.1007/s10489-016-0825-8>.
- [33] Y. Zhu, D. Tian, F. Yan, Effectiveness of entropy weight method in decision-making, *Math. Probl. Eng.* 2020 (2020) 1–5, <http://dx.doi.org/10.1155/2020/3564835>.
- [34] N.M. Thoppil, V. Vasu, C.S.P. Rao, Health indicator construction and remaining useful life estimation for mechanical systems using vibration signal prognostics, *Int. J. Syst. Assur. Eng. Manag.* 12 (5) (2021–10) 1001–1010, <http://dx.doi.org/10.1007/s13198-021-01190-z>.
- [35] J. Zhang, J. Tian, A.M. Alcaide, J.I. Leon, S. Vazquez, L.G. Franquelo, H. Luo, S. Yin, Lifetime extension approach based on the Levenberg–Marquardt neural network and power routing of DC–DC converters, *IEEE Trans. Power Electron.* 38 (8) (2023) 10280–10291, <http://dx.doi.org/10.1109/TPEL.2023.3275791>.
- [36] Y. Yang, X. Ding, M.W. Urban, Chemical and physical aspects of self-healing materials, *Prog. Polym. Sci.* 49 (2015) 34–59, <http://dx.doi.org/10.1016/j.progpolymsci.2015.06.001>.
- [37] M.L. Baptista, K. Goebel, E.M.P. Henriques, Relation between prognostics predictor evaluation metrics and local interpretability SHAP values, *Artificial Intelligence* 306 (2022) 103667, <http://dx.doi.org/10.1016/j.artint.2022.103667>.
- [38] J. Li, Y. Zi, Y. Wang, Y. Yang, Health indicator construction method of bearings based on Wasserstein dual-domain adversarial networks under normal data only, *IEEE Trans. Ind. Electron.* 69 (10) (2022) 10615–10624, <http://dx.doi.org/10.1109/TIE.2022.3156148>.
- [39] X. Yu, Z. Liang, Y. Wang, H. Yin, X. Liu, W. Yu, Y. Huang, A wavelet packet transform-based deep feature transfer learning method for bearing fault diagnosis under different working conditions, *Measurement* 201 (2022) 111597, <http://dx.doi.org/10.1016/j.measurement.2022.111597>.
- [40] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197, <http://dx.doi.org/10.1109/4235.996017>.
- [41] C.C. Coello, M.S. Lechuga, MOPSO: A proposal for multiple objective particle swarm optimization, in: *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, vol. 2, IEEE, Honolulu, HI, USA, 2002, pp. 1051–1056, <http://dx.doi.org/10.1109/CEC.2002.1004388>.
- [42] A. Jahan, K.L. Edwards, Chapter 3 - Multi-criteria decision-making for materials selection, in: A. Jahan, K.L. Edwards (Eds.), *Multi-Criteria Decision Analysis for Supporting the Selection of Engineering Materials in Product Design*, Butterworth-Heinemann, Boston, 2013, pp. 31–41, <http://dx.doi.org/10.1016/B978-0-08-099386-7.00003-9>.
- [43] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, *J. Sound Vib.* 289 (4–5) (2006) 1066–1090, <http://dx.doi.org/10.1016/j.jsv.2005.03.007>.
- [44] L. Cui, X. Wang, H. Wang, J. Ma, Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter, *IEEE Trans. Instrum. Meas.* 69 (6) (2019) 2858–2867, <http://dx.doi.org/10.1109/TIM.2019.2924509>.
- [45] Z. Zhang, W. Huang, Y. Liao, Z. Song, J. Shi, X. Jiang, C. Shen, Z. Zhu, Bearing fault diagnosis via generalized logarithm sparse regularization, *Mech. Syst. Signal Process.* 167 (2022) 108576, <http://dx.doi.org/10.1016/j.ymssp.2021.108576>.
- [46] X.-S. Si, An adaptive prognostic approach via nonlinear degradation modeling: Application to battery data, *IEEE Trans. Ind. Electron.* 62 (8) (2015) 5082–5096, <http://dx.doi.org/10.1109/TIE.2015.2393840>.
- [47] Y. Qin, J. Yang, J. Zhou, H. Pu, X. Zhang, Y. Mao, Dynamic weighted federated remaining useful life prediction approach for rotating machinery, *Mech. Syst. Signal Process.* 202 (2023) 110688, <http://dx.doi.org/10.1016/j.ymssp.2023.110688>.