Unsupervised Health Indicator Construction via Deep Reinforcement Learning with Terminal-Dominant Reward

Zeqi Wei[1], Zhibin Zhao[1], and Ruqiang Yan[1]

[1]*Xi'an Jiaotong University, Xi'an, 710049, China*

weizeqi@stu.xjtu.edu.cn
zhaozhibin@xjtu.edu.cn
yanruqiang@xjtu.edu.cn

ABSTRACT

In industrial intelligent maintenance, the construction of a reliable health indicator (HI) is crucial for accurate degradation assessment and fault prediction. However, existing methods face two major limitations: fusion-based approaches often suffer from low-quality or irrelevant features that degrade the discriminative capability of the HI, while reconstruction-based approaches rely heavily on high-quality healthy data, which is difficult to obtain in real-world scenarios. To overcome these challenges, this paper proposes an **U**nsupervised **T**erminal-**D**ominant framework for **HI** construction (UTD-HI). The method does not rely on remaining useful life (RUL) labels or pre-defined thresholds. Within a deep reinforcement learning (DRL) paradigm, UTD-HI learns an adaptive feature-weighting policy that suppresses irrelevant features and enhances informative ones. A reward mechanism integrating monotonicity, smoothness, and a sparse terminal constraint is designed, while hindsight experience replay (HER) is introduced to address reward sparsity. Furthermore, by employing different reward strategies in normal and abnormal stages, the framework can automatically and accurately distinguish between healthy and degraded operating conditions. Experimental results on the XJTU-SY bearing dataset demonstrate that the proposed method constructs HIs with superior trendability, monotonicity, and robustness across different operating conditions, thereby offering a practical solution for HI construction in real-world environments.

1. ELECTRONIC SUBMISSION

Prognostic Health Management (PHM) systems play a critical role in ensuring the reliability and safety of machinery by monitoring equipment health and predicting faults. Constructing a health indicator (HI) is necessary for accurately assessing machine status (Lei, Li, Guo, Li, Yan & Lin, 2018). Moreover, accurate HIs facilitate early anomaly detection and support condition-based maintenance strategies, reducing unexpected downtime and maintenance costs. Consequently, HI construction has attracted considerable attention and remains a research focus in recent years (Wang, Tsui & Miao, 2017). However, it is still challenging to create HIs that are both sensitive to machine degradation and robust across different equipment.

In general, HIs can be divided into two categories: physical HIs (PHIs) and virtual HIs (VHIs) (Djeziri, Benmoussa & Zio, 2020). PHIs are derived from monitoring raw signals through signal processing or statistical methods. In signal processing-based methods, PHIs are typically constructed according to the underlying physical failure mechanisms (Yan, Wang, Kong, Xia, Peng & Li, 2021). In contrast, statistical methods are more popular due to their simplicity. Various statistical features such as root mean square (RMS) (Meng, Yan, Chen, Liu & Wu, 2021), kurtosis (Zhong, Wang & Li, 2021), entropy (Yan, Wang, Xia, Zheng, Peng & Xi, 2023), among others, are widely used. PHIs provide clear physical interpretability, but they are often designed for specific tasks, which leads to poor generalization. In addition, these methods require strong expert knowledge, which limits their broader applicability.

VHIs are also categorized into two classes: fusion-based methods and deep learning-based methods. In fusion-based approaches, VHIs are constructed by combining several PHIs into a single indicator that represents overall degradation information (Djeziri, Benmoussa & Zio, 2020). Various techniques have been proposed for this purpose, among which principal component analysis (PCA) is the most widely used. For example, Guo, Wang, Li, Yang, Huang, Yazdi, and Hooi (2024) applied PCA to construct a nonlinear HI for degradation modeling and remaining useful life (RUL) prediction. Similarly, Buchaiah and Shakya (2022) employed 14 dimensionality reduction methods to fuse selected features into a VHI. Fusion-based methods are simple to apply because they only require extracting basic features for fusion.

However, the quality of the constructed VHIs largely depends on the selected features. Inappropriate features may introduce side effects, resulting in insufficient representation of degradation information.

Deep learning-based methods do not fuse PHIs directly. Instead, they extract implicit degradation information from high-dimensional data spaces. Among them, the autoencoder (AE) is the most commonly used approach. These methods aim to reconstruct healthy data and use reconstruction errors as a VHI. For example, González-Muñiz, Diaz, Cuadrado, and Garcia-Perez (2022) leveraged disentangled representations in the latent space of an AE and used the latent reconstruction error as a VHI. Ye and Yu (2021) proposed a long short-term memory convolutional AE that generates HIs based-on reconstruction errors. Such methods construct HIs directly from raw signals, but their performance largely depends on the availability and quality of healthy data. Moreover, their generalization ability is closely tied to the network architecture and training process of the AE.

It is evident that these methods have achieved promising results in HI construction. Beyond the aforementioned limitations, applying constructed HIs to RUL prediction or maintenance strategy optimization typically requires a failure threshold. However, this threshold often differs across devices, making it difficult to generalize a fixed, manually defined one. To address this issue, some studies have attempted to normalize HIs using their maximum and minimum values (Ni, Ji & Feng, 2022). Nevertheless, such normalization remains device-dependent and introduces additional limitations.

This paper proposes an unsupervised HI construction method based on a deep reinforcement learning (DRL) paradigm. DRL is used to assign different weights to features, which improves both trendability and monotonicity. Firstly, time-domain and frequency-domain features are extracted from raw signals and modeled as the state space. Secondly, a reward with monotonicity and smoothness constraints is designed. To overcome the limitations of failure thresholds, a dominance reward is introduced at the terminal state, and hindsight experience replay (HER) is adopted to learn from this sparse terminal reward. Finally, DRL outputs a weight for each fused feature, assigning low weights to features with little contribution to degradation. The HI is then obtained as a weighted sum of the features. Moreover, by employing different reward functions for normal and abnormal stages, the proposed method can automatically and accurately distinguish between healthy and degraded conditions.

2. METHODOLOGY

2.1. Framework of UTD-HI

The overall framework of the proposed unsupervised HI construction (UTD-HI) is illustrated in Figure 1, where raw sensor signals are first processed to extract a set of time-domain and frequency-domain features, which together form the state space for the DRL agent. The agent is trained to assign adaptive weights to these features. To guide the training, a reward function is designed with three components: (1) monotonicity constraint, (2) smoothness constraint, and (3) a terminal dominance reward that alleviates the need for a manually defined failure threshold. HER is further incorporated to address the sparsity of the terminal reward and improve training result. During inference, the trained DRL agent outputs weights for each feature, giving lower weights to those less relevant to degradation. Finally, the HI is obtained as the weighted sum of the fused features, which ensures better trendability and monotonicity.

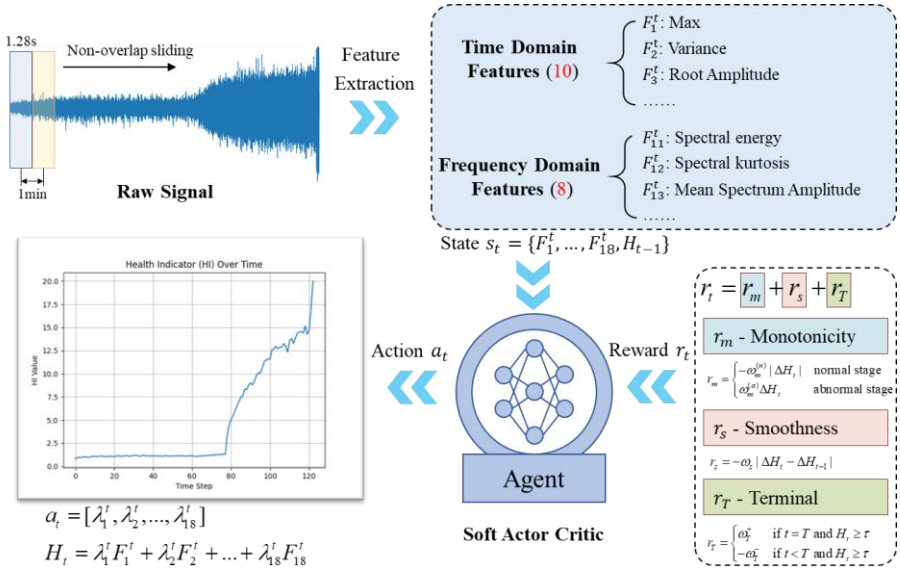

Figure 1. Framework of the UTD-HI

2.2. Basic Features Extraction

To cover the degradation information as comprehensively as possible, 10 time-domain features and 8 frequency-domain features are extracted as shown in Table 1. In the time domain, the raw vibration signal is represented as x_i, where $i=1,\ldots,N$ and N is the total number of samples. The mean value of the signal is denoted by μ. In the frequency domain, the spectrum obtained by the Fast Fourier Transform (FFT) is denoted as X_k, where $k=1,\ldots,K$ and K is the number of frequency bins. Each spectral component corresponds to a frequency f_k which is given by $f_k=\frac{k}{K}\times f_s$, where f_s is the sampling frequency.

Table 1. Manually extracted features

Time-domain features			
F_1: Max	$\max(x_i)$	F_2: Variance	$\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$
F_3: Root Amplitude	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$	F_4: Absolute Mean	$\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$
F_5: Root Mean Square	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$	F_6: Kurtosis	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\sigma^4}$
F_7: Crest Factor	$\frac{\max(x_i)}{F_5}$	F_8: Skewness	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3}$
F_9: Waveform Factor	$\frac{F_5}{F_4}$	F_{10}: Margin Factor	$\frac{F_1}{(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert})^2}$
Frequency-domain features			
F_{11}: Spectral Energy	$\sum_{k=1}^{K}\lvert X_k\rvert^2$	F_{12}: Spectral Kurtosis	$\frac{\frac{1}{K}\sum_{k=1}^{K}(\lvert X_k\rvert-\mu_f)^4}{\sigma_f^4}$
F_{13}: Mean Spectrum Amplitude	$\frac{1}{K}\sum_{k=1}^{K}\lvert X_k\rvert$	F_{14}: Frequency Center	$F_{14}=\frac{\sum_{k=1}^{K}f_k\lvert X_k\rvert}{\sum_{k=1}^{K}\lvert X_k\rvert}$
F_{15}: Root Mean Square Frequency	$\sqrt{\frac{\sum_{k=1}^{K}f_k^2\lvert X_k\rvert}{\sum_{k=1}^{K}\lvert X_k\rvert}}$	F_{16}: Variance Frequency	$\frac{\sum_{k=1}^{K}(f_k-F_{14})^2\lvert X_k\rvert}{\sum_{k=1}^{K}\lvert X_k\rvert}$
F_{17}: Root Variance Frequency	$\sqrt{F_{16}}$	F_{18}: Spectral Center of Gravity	$\frac{\sum_{k=1}^{K}f_k\lvert X_k\rvert^2}{\sum_{k=1}^{K}\lvert X_k\rvert^2}$

To reduce the influence of magnitude differences among features and facilitate subsequent feature fusion, each feature is normalized. Specifically, the mean of the first ten samples of each feature is calculated as a reference value. Then, all values of this feature are divided by the reference mean to obtain the normalized feature. This method ensures that each feature has a comparable scale while preserving the relative variation trends, which is important for degradation analysis and health index construction.

2.3. DRL Agent

2.3.1. Soft Actor-Critic

Soft Actor-Critic (SAC) is adopted in this study as the reinforcement learning agent (Haarnoja, Zhou, Hartikainen, Tucker, Ha, Tan, Kumar, Zhu, Gupta, Abbeel, and Levine, 2018). SAC is well-known for its superior performance in continuous action spaces, owing to the introduction of an entropy term in its objective function. Specifically, SAC aims to maximize not only the expected cumulative reward but also the policy entropy, which encourages exploration and improves robustness against perturbations. The optimal policy function of SAC is formulated as:

$$\pi^*=\arg\max_{\pi}E_{(s_t,a_t)\sim\rho_\pi}\left[\sum_t r(s_t,a_t)+\alpha H\left(\pi(\cdot\mid s_t)\right)\right] \quad (1)$$

where ρ_π denotes the state–action distribution induced by policy π, $r(s_t,a_t)$ is the reward function, $H(\cdot)$ is the entropy, and α is the temperature coefficient that balances reward maximization and entropy maximization. The information entropy is defined as:

$$H(P)=\underset{x\sim P}{E}\left[-\log P(x)\right] \quad (2)$$

where x follow the probability distribution P. The soft state-action value function in SAC is updated according to the soft Bellman iteration:

$$\begin{aligned}&Q(s_t,a_t)\\&=r(s_t,a_t)+\gamma E_{s_{t+1},a_{t+1}}\left[Q(s_{t+1},a_{t+1})-\alpha\log\left(\pi(a_{t+1}|s_{t+1})\right)\right]\end{aligned} \quad (3)$$

where γ is discount factor and α is the temperature coefficient, which is adaptively tuned during training. By maximizing both reward and entropy, SAC prevents premature convergence to deterministic policies. In addition, SAC employs two *Q*-functions to reduce overestimation bias and applies the reparameterization trick to enable efficient gradient-based optimization.

2.3.2. State

The state s_t represents the input to the DRL agent at step t. It is composed of two parts: (i) the candidate fusion features $\{F_1,F_2\ldots,F_{18}\}$, which contain time-domain and frequency-

domain degradation information, and (ii) the previously constructed health indicator H_{t-1}, which reflects the degradation trend up to step $t-1$. By including both the raw features and the past HI, the state provides sufficient information for the agent to learn feature weighting strategies that are consistent with the degradation process. Formally, the state is defined as:

$$s_t = \left\{F_1^t, \ldots, F_{18}^t, H_{t-1}\right\} \tag{4}$$

2.3.3. Action

Given a state s_t, the agent outputs an action vector:

$$a_t = \left\{\lambda_1^t, \lambda_2^t, \ldots, \lambda_{18}^t\right\}, \lambda_i^t \in [0,1] \tag{5}$$

where each element λ_i^t corresponds to the weight assigned to the i-th feature at step t. These weights determine the contribution of each feature to the fused HI. By adjusting the values of λ_i^t, the agent is capable of emphasizing degradation-sensitive features while suppressing noise or less relevant ones. The bounded range [0,1] ensures numerical stability and prevents the agent from assigning excessively large or negative weights during HI construction.

2.3.4. Reward

Reward is the most critical component in reinforcement learning. In this study, the reward function is divided into three parts: monotonicity, smoothness, and a terminal threshold constraint. The overall reward is defined as:

$$r_t = r_m + r_s + r_T \tag{6}$$

where r_m is the monotonicity reward, r_s is the smoothness reward, and r_T is the terminal reward.

- Monotonicity reward (r_m)

The form of r_m depends on the degradation stage of the machine, and is given as:

$$r_m = \begin{cases} -\omega_m^{(n)} \Delta H_t & \text{normal stage} \\ \omega_m^{(a)} \Delta H_t & \text{abnormal stage} \end{cases} \tag{7}$$

where $\omega_m^{(n)} > 0$ and $\omega_m^{(a)} > 0$ are coefficients, and ΔH_t denotes the change in HI at step t. This design ensures that the HI remains as stable as possible during the normal stage and increases steadily during degradation, which matches the real-world degradation process. The degradation stage is determined by the slope of HI: if the slope exceeds 0.15, the machine is considered in the abnormal stage; otherwise, it is regarded as normal.

- Smoothness reward (r_s)

The smoothness reward constrains the HI to grow smoothly by minimizing fluctuations. It is defined as:

$$r_s = -\omega_s \left|\Delta H_t - \Delta H_{t-1}\right| \tag{8}$$

where $\omega_s > 0$ is a coefficient. This formulation penalizes large second-order differences, ensuring a smoother HI trajectory.

- Terminal reward (r_T)

The terminal reward ensures that the HI exceeds the failure threshold τ at the final step T, while remaining within $[0, \tau)$ before failure. It is defined as:

$$r_T = \begin{cases} \omega_T^+ & t = T \text{ and } H_t \geq \tau \\ -\omega_T^- & t < T \text{ and } H_t \geq \tau \end{cases} \tag{9}$$

where $\omega_T^+ > 0$ and $\omega_T^- > 0$ are coefficients. This encourages the HI to cross the failure threshold only at the terminal step, avoiding premature exceedance. Since the terminal reward r_T is very sparse, it may hinder the efficiency of policy learning. To address this issue, this study introduces hindsight experience replay (HER) (Andrychowicz, Wolski, Ray, Schneider, Fong, Welinder, McGrew, Tobin, Abbeel, and Zaremba, 2017). HER relabels failed trajectories with alternative goals, allowing the agent to extract useful training signals even when the original goal is not achieved. By augmenting the replay buffer with such relabeled experiences, HER effectively improves sample efficiency and accelerates convergence in sparse-reward settings.

3. EXPERIMENT AND DISCUSSION

The effectiveness of the proposed method is verified using the XJTU-SY dataset. Three evaluation metrics are employed to compare with baseline methods: monotonicity, trendability, and the hybrid metric.

3.1. Dataset Description and Preprocessing

The XJTU-SY dataset contains run-to-failure vibration signals collected from 15 bearings under accelerated degradation tests (Wang, Lei, Li, and Li, 2018). Three operating conditions were set in the experiments, with five bearings being tested under each condition, as summarized in Table 2. The vibration signals were collected every 60 seconds with a sampling frequency of 25.6 kHz and a sampling duration of 1.28 seconds. To ensure complete degradation trajectories, each test started from the healthy condition and was terminated when the maximum vibration amplitude exceeded ten times that of the initial healthy state.

A non-overlapping sliding window is applied, where each one-minute segment is treated as a time window for feature extraction. The preprocessing procedure is illustrated in Figure 2.

Table 2. Operating conditions of XJTU-SY bearing dataset

Operating condition	1	2	3
Speed (rpm/Hz)	2100/35	2250/37.5	2400/40
Load (kN)	12	11	10
Dataset	Bearing 1_1 to 1_5	Bearing 2_1 to 2_5	Bearing 3_1 to 3_5

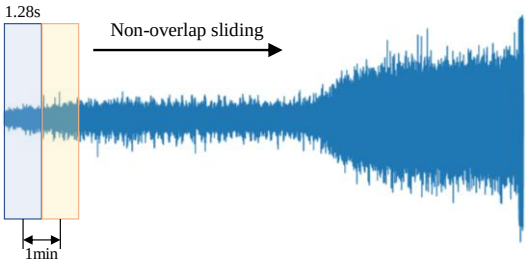

Figure 2. Data preprocessing procedure

3.2. HI Construction

3.2.1. Implementation Details

The hyperparameters of the proposed method are set as follows: $\omega_m^{(n)}=10$, $\omega_m^{(a)}=5$, $\omega_s=1$, $\omega_T^{+}=50$, $\omega_T^{-}=100$, $\tau=20$. The terminal step T is determined by the length of the dataset.

The DRL agent is trained with the SAC algorithm, using a learning rate of 3×10^{-4} with the Adam optimizer and a batch size of 32. The replay buffer size is 10^6, and HER is employed to deal with the sparse terminal reward. The discount factor γ is set to 0.99, and the temperature parameter α is automatically adjusted during training.

Feature vectors of dimension 18 are extracted for each time window, and the HI is updated at every step until the terminal state.

3.2.2. The Constructed HI

Eight bearings (1_1, 1_2, 1_3, 1_5, 2_1, 2_2, 2_3, and 2_4) are selected for analysis. Since failure threshold information is incorporated during training, HI values exceeding the threshold are truncated. Consequently, all constructed HIs can be consistently normalized by dividing them by τ, which facilitates cross-bearing comparisons. As Figure 3 shows, each HI trajectory is clearly divided into two stages, corresponding to the normal and degradation phases, owing to the stage-wise reward design. In addition, abnormal points are automatically identified along the trajectories, providing valuable guidance for early fault detection and subsequent prediction tasks.

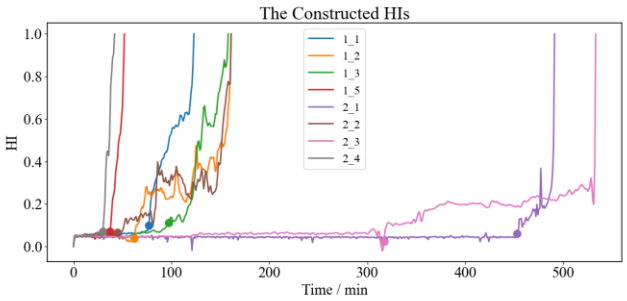

Figure 3. The constructed HIs of UTD-HI

3.3. Performance Evaluation

3.3.1. Evaluation Metrics

In order to quantitatively evaulate the performance of the proposed method, three metrics are measured as follows:

- Trendability

Trendability measures the correlation between the HI and the run-to-failure timeline, and thus evaluates whether the HI follows the degradation process. It is defined as:

$$Tred(H)=\frac{\left|\sum_{t=1}^{T}\left(H_t-\bar{H}\right)\left(t-\bar{t}\right)\right|}{\sqrt{\sum_{t=1}^{T}\left(H_t-\bar{H}\right)^2\sum_{t=1}^{T}\left(t-\bar{t}\right)^2}} \tag{10}$$

where $\bar{H}=\frac{1}{T}\sum_{t=1}^{T}H_t$ and $\bar{t}=\frac{1}{T}\sum_{t=1}^{T}t$.

- Monotonicity

Monotonicity evaluates whether the HI reflects the irreversible degradation process of machinery. It is defined as:

$$Mon(H)=\frac{\left|\sum_{t=2}^{T}\mathrm{I}\left(H_t-H_{t-1}>0\right)-\sum_{t=2}^{T}\mathrm{I}\left(H_{t-1}-H_t>0\right)\right|}{T-1} \tag{11}$$

where $\mathrm{I}(\cdot)$ is an indicator function.

- Hybrid metric

Since trendability and monotonicity capture different aspects of HI quality, a hybrid metric (HM) is used to provide a comprehensive evaluation (Guo, Yu, Duan, Gao, and Zhang 2022):

$$HM=\frac{Tred+Mon}{2} \tag{12}$$

3.3.2. Comparison Results

To evaluate the effectiveness of the proposed UTD-HI method, comparisons are made with three representative HI construction approaches: (1) RMS-HI, which directly uses the root mean square of raw signals as the health indicator;

(2) PCA-HI, where the first principal component after principal component analysis is used as the HI; and (3) AE-HI, which employs the reconstruction error of an autoencoder as the HI. The evaluation is conducted on eight bearings using trendability, monotonicity, and a hybrid metric, as reported in Tables 3–5.

Table 3. Trendability of different methods

Bearing	RMS-HI	PCA-HI	AE-HI	UTD-HI
1_1	0.8632	0.8106	0.5701	**0.8698**
1_2	0.9014	0.8081	0.8824	**0.9079**
1_3	0.7560	**0.8198**	0.4436	0.7972
1_5	**0.7584**	0.6448	0.4527	0.7284
2_1	0.3731	**0.4048**	0.2573	0.3766
2_2	**0.8838**	0.6829	0.4497	0.8394
2_3	**0.8821**	0.7815	0.8236	0.8619
2_4	**0.7909**	0.7416	0.7224	0.7491

Table 4. Monotonicity of different methods

Bearing	RMS-HI	PCA-HI	AE-HI	UTD-HI
1_1	0.2459	0.1311	0.0943	**0.3984**
1_2	**0.2000**	**0.2000**	0.0766	0.1304
1_3	**0.3885**	0.0573	0.0422	0.3538
1_5	0.2157	0.1765	0.1127	**0.4615**
2_1	0.0245	0.0163	0.0357	**0.0713**
2_2	0.225	0.0125	0.0898	**0.235**
2_3	0.0301	0.0225	0.0298	**0.0507**
2_4	0.2195	0.2683	0.1494	**0.381**

Table 5. Hybrid metric of different methods

Bearing	RMS-HI	PCA-HI	AE-HI	UTD-HI
1_1	0.5546	0.4709	0.3322	**0.6341**
1_2	**0.5507**	0.50405	0.4795	0.5192
1_3	0.5723	0.43855	0.2429	**0.5755**
1_5	0.4871	0.4107	0.2827	**0.5950**
2_1	0.1988	0.2106	0.1465	**0.2240**
2_2	**0.5544**	0.3477	0.26975	0.5322
2_3	0.4561	0.402	0.4267	**0.4563**
2_4	0.5052	0.505	0.4359	**0.5651**

The results show that RMS-HI achieves the highest trendability, as it is highly sensitive to signal amplitude growth during degradation; however, its monotonicity is limited and it lacks the capacity to integrate multi-feature information. PCA-HI and AE-HI, as fusion-based and reconstruction-based methods respectively, exhibit more unstable performance. In particular, PCA-HI may introduce counterproductive effects when irrelevant features are incorporated, thereby reducing discriminative power, while AE-HI relies heavily on high-quality healthy data that are difficult to obtain in practice, leading to inferior results in most cases. In contrast, although UTD-HI does not always surpass RMS-HI in trendability, it achieves a better trade-off between trendability and monotonicity. As reflected in the hybrid metric, UTD-HI demonstrates superior overall performance in the vast majority of cases and exhibits greater robustness across different bearings and operating conditions. These results demonstrate that the proposed method effectively learns to assign feature weights, suppresses low-quality contributions, and generates health indicators that evolve smoothly and reliably with degradation, thereby offering a more practical solution for health assessment under unsupervised conditions.

4. CONCLUSION

This study proposes an unsupervised terminal-dominant health indicator construction framework within a DRL paradigm. The method adaptively assigns feature weights under the guidance of stage-aware and smoothness rewards, while HER is introduced to address sparse terminal rewards. Experimental results on the XJTU-SY bearing dataset demonstrate that the constructed HIs not only exhibit superior monotonicity and trendability but also achieve the best performance on the hybrid metric compared with RMS-, PCA-, and AE-based baselines. These findings confirm that UTD-HI can effectively capture degradation processes and distinguish between normal and abnormal stages without relying on labeled data. Therefore, the proposed method provides a reliable foundation for downstream tasks such as remaining useful life prediction and maintenance decision optimization in complex industrial environments. Future research may focus on improving generalization across diverse operating conditions, for example, through meta-reinforcement learning or transfer learning.

ACKNOWLEDGEMENT

This work was supported by the Joint Funds of the National Natural Science Foundation of China (No. U23A20620) and the National Natural Science Foundation of China (No. 52275130).

REFERENCES

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, *104*, 799-834.

Wang, D., Tsui, K. L., & Miao, Q. (2017). Prognostics and health management: A review of vibration based bearing and gear health indicators. *IEEE Access*, *6*, 665-676.

Djeziri, M. A., Benmoussa, S., & Zio, E. (2020). Review on health indices extraction and trend modeling for remaining useful life estimation. In *Artificial intelligence*

techniques for a scalable energy transition: advanced methods, digital technologies, decision support tools, and applications (pp. 183-223). Cham: Springer International Publishing.

Yan, T., Wang, D., Kong, J. Z., Xia, T., Peng, Z., & Xi, L. (2021). Definition of signal-to-noise ratio of health indicators and its analytic optimization for machine performance degradation assessment. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-16.

Meng, J., Yan, C., Chen, G., Liu, Y., & Wu, L. (2021). Health indicator of bearing constructed by rms-CUMSUM and GRRMD-CUMSUM with multifeatures of envelope spectrum. *IEEE Transactions on instrumentation and measurement*, *70*, 1-16.

Zhong, J., Wang, D., Guo, J. E., Cabrera, D., & Li, C. (2020). Theoretical investigations on kurtosis and entropy and their improvements for system health monitoring. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-10.

Yan, T., Wang, D., Xia, T., Zheng, M., Peng, Z., & Xi, L. (2023). Entropy-maximization oriented interpretable health indicators for locating informative fault frequencies for machine health monitoring. *Mechanical Systems and Signal Processing*, *198*, 110461.

Guo, J., Wang, Z., Li, H., Yang, Y., Huang, C. G., Yazdi, M., & Kang, H. S. (2024). A hybrid prognosis scheme for rolling bearings based on a novel health indicator and nonlinear Wiener process. *Reliability Engineering & System Safety*, *245*, 110014.

Buchaiah, S., & Shakya, P. (2022). Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. *Measurement*, *188*, 110506.

González-Muñiz, A., Diaz, I., Cuadrado, A. A., & Garcia-Perez, D. (2022). Health indicator for machine condition monitoring built in the latent space of a deep autoencoder. *Reliability Engineering & System Safety*, *224*, 108482.

Ye, Z., & Yu, J. (2021). Health condition monitoring of machines based on long short-term memory convolutional autoencoder. *Applied Soft Computing*, *107*, 107379.

Ni, Q., Ji, J. C., & Feng, K. (2022). Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network. *IEEE Transactions on Industrial Informatics*, *19*(2), 1301-1311.

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., ... & Levine, S. (2018). Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*.

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., ... & Zaremba, W. (2017). Hindsight experience replay. *Advances in neural information processing systems*, *30*.

Wang, B., Lei, Y., Li, N., & Li, N. (2018). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, *69*(1), 401-412.

Guo, L., Yu, Y., Duan, A., Gao, H., & Zhang, J. (2022). An unsupervised feature learning based health indicator construction method for performance assessment of machines. *Mechanical Systems and Signal Processing*, *167*, 108573.

BIOGRAPHIES

Zeqi Wei Received the B.S. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2022. He is currently working toward the Ph.D degree mechanical engineering in the Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China.

His current research is focused on deep reinforcement learning, maintenance strategy optimization, and machinery health monitoring.

Zhibin Zhao received the B.S. degree in Tsien Hsue-Shen honor class and the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2015 and 2020, respectively.

Currently, he is a Lecturer in mechanical engineering with the Department of Mechanical Engineering, Xi'an Jiaotong University. His research interests include sparse signal processing and machine learning algorithms for machinery health monitoring and healthcare.

Dr. Zhao is an Associate Editor of IEEE Transactions on Instrumentation and Measurement.

Ruqiang Yan received the Ph.D. degree in mechanical engineering from the University of Massachusetts at Amherst, Amherst, MA, USA, in 2007. He is currently a Professor of mechanical engineering with Xi'an Jiaotong University.

His research interests include data analytics, machine learning, and energy-efficient sensing and health diagnosis of large-scale, complex, dynamical systems.

Dr. Yan is a Fellow of ASME (2019) and IEEE (2022). He is also the Editor-in-Chief of the IEEE Systems Journal, an Associate Editor-in-Chief of the IEEE Transactions on Instrumentation and Measurement, an Associate Editor of the IEEE Sensors Journal, and Editorial Board Member of Chinese Journal of Mechanical Engineering and Journal of University of Science and Technology of China.