

Aeroengine Remaining Life Prediction Based on Advanced Health Index Construction and Gray Similarity Multiscale Matching

Shilong Sun^{ID}, Member, IEEE, Haodong Huang^{ID}, Jian Ding^{ID}, Dong Wang^{ID}, Member, IEEE,
and Wenfu Xu^{ID}, Senior Member, IEEE

Abstract— In this article, we address the challenge of predicting the remaining useful life (RUL) of aeroengines, which are critical components of aircraft that operate under increasingly extreme conditions as engine performance enhances. Ensuring the safety and reliability of these engines is paramount. To this end, we introduce a novel RUL prediction methodology that leverages gray similarity multiscale matching. This approach employs the robust capabilities of long short-term memory (LSTM) networks for processing time-series data. First, an LSTM stacked autoencoder (L-SAE) is designed to extract pivotal operational features of the engine, thereby delineating its degradation trajectory. Furthermore, the gray correlation analysis is utilized to assess the similarity between these degradation trajectories, which are complemented by a multitime scale sliding window technique for enhanced similarity matching. Subsequently, kernel density estimation (KDE) is applied to gauge the uncertainty associated with the prediction outcomes. The efficacy and superiority of our proposed method are demonstrated through the validation of the experiment study. Comparative analysis reveals that our method outperforms existing techniques in key evaluation metrics, underscoring its potential applicability to large-scale datasets. This validation confirms not only the method's effectiveness but also its advantage in predicting the RUL of aeroengines with greater accuracy and reliability.

Index Terms— Gray correlation analysis, remaining useful life (RUL), similarity matching, stacked autoencoder.

I. INTRODUCTION

PROGNOSTICS and health management (PHM) technology is used to monitor the operating status of equipment

Received 2 September 2024; revised 28 September 2024; accepted 30 September 2024. Date of publication 16 October 2024; date of current version 30 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52475112, in part by Guangdong Basic and Applied Basic Research Foundation Under Grant 2024A1515012041, in part by the Basic Research Program of Shenzhen JCYJ20220818102415034, and in part by Shenzhen Higher Education Stability Support Plan under Grant GXWD20231130195340002. The Associate Editor coordinating the review process was Dr. Alvaro Hernandez. (Corresponding author: Shilong Sun.)

Shilong Sun, Haodong Huang, and Jian Ding are with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: sunshilong@hit.edu.cn; hhd1340201839@163.com; 13397159127@163.com).

Dong Wang is with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: dongwang4-c@sjtu.edu.cn).

Wenfu Xu is with the Key University Laboratory of Mechanism and Machine Theory and Intelligent Unmanned Systems of Guangdong Shenzhen, Guangdong 518055, China (e-mail: wfxu@hit.edu.cn).

Digital Object Identifier 10.1109/TIM.2024.3481588

or systems in real time, detect potential failures through data analysis and predictive models in advance, and provide optimized maintenance recommendations to maximize equipment reliability and reduce maintenance costs [1]. Remaining useful life (RUL) prediction is one of the key steps in PHM technology. This method can detect abnormal conditions and performance degradation of equipment components through real-time monitoring and data analysis so that actions can be taken in advance to reduce losses before the problem develops into a serious failure.

RUL prediction methods can be divided into model-based methods, data-driven methods, and hybrid methods [2]. The model-based methods mainly construct corresponding physical models for RUL prediction by deeply understanding the operation mechanism and failure mechanism of equipment [3]. The data-driven methods utilize failure operation data of the same type of equipment regularly collected from various real-time monitoring sensors to estimate the RUL of the equipment through data-driven models.

The data-driven methods can be roughly divided into statistical model-based methods, machine learning-based methods [4], and similarity-based methods based on their principles. Among them, statistical model-based methods, predominantly random process models, capture the randomness and variability of equipment operation by performing probability analysis and model fitting on historical data to predict RUL. Existing statistical models with more applications include Wiener processes [5], Levy processes [6], and Gaussian processes [7]. Machine learning-based methods utilize machine learning techniques such as multilayer perceptrons [8], support vector machines [9], [10], and neural networks [11], [12], [13] to learn complex degradation and failure patterns of equipment. Neural networks perform well in processing large amounts of high-dimensional data and often perform well in predicting the RUL of complex systems. However, neural networks suffer from the “black box” effect and require a large amount of labeled data for training.

The initial application of the similarity-based method comes from [14] in which sample-based reasoning and similarity-matching matrix prediction are the basis of the similarity-based prediction (SBP) method. Wang et al. [15] simplified similarity matching and proposed the trajectory similarity prediction method in the PHM turbofan engine RUL prediction compe-

tition in 2008, which is also considered the beginning of the mature application of the SBP method. Similarity matching is the core content of the SBP method, mainly involving two links: similarity matching rules and similarity measurement. The former is used to determine the degradation behavior for similarity measurement, while the latter is used to measure the similarity between different degradation behaviors. Distance methods [16], [17] are widely used in existing research for similarity measurement, while normalized mutual correlation methods [18] and maximum mean difference methods [19] are used to measure correlation. To overcome the impact of time delay on calculation, Wang [20] first used dynamic time window methods for time series matching. On this basis, Yu et al. [21] further proposed a new zero-centered rule to extract more similar reference sample segments. According to the specific scale difference of matching rules, similarity matching rules can also be divided into local similarity-based and global similarity-based matching rules. The local similarity-based matching rule uses part of the degradation trajectory (usually the end) of the equipment under test as the degradation behavior for similarity comparison [19], [22]. The global similarity-based matching rule treats the entire degradation trajectory of the equipment under test as the object for similarity comparison [15]. Besides using degradation trajectories as the object of similarity measurement, sensor signal spectra are also directly used for similarity matching [23].

A major drawback of similarity matching methods in remaining life prediction is their high dependency on data. These methods typically require a large amount of historical data to find samples like the current state, which demands high data quality and representativeness. Particularly in high-dimensional data spaces, the complexity of computing similarity increases significantly, leading to high computational costs and time overhead. Additionally, similarity matching methods may be very sensitive to noise, with anomalies in the data potentially significantly affecting prediction accuracy, especially when data preprocessing is insufficient [24].

Hybrid methods combine physical model knowledge with data-driven techniques for remaining life prediction. A common approach involves using physical knowledge and practical experience to build analytical models, and then employing data-driven techniques (such as particle filters and SVM) to update model parameters based on real-time data. Artificial intelligence is also frequently integrated with statistical models. For instance, Zemouri and Gouriveau [25] combined ANN with autoregressive models for remaining life prediction; Sankavaram et al. [26] developed a hybrid method integrating fault diagnosis and remaining life prediction for automotive and onboard electronic systems; Wang et al. [27] used the sparse learning feature of RVM to train health index generation based on similarity methods; Wang and Mamo [28] combined exponentially weighted moving average control charts with random forests, integrating multiple evolutionary algorithms within the random forests. Although these methods address the limitations of the previous approaches, they are still not well developed, and there are few reported studies so far [29].

To accurately assess and predict the health status of aero-engines, this article proposes a RUL prediction method based

on gray similarity multiscale matching. After using the long short-term memory (LSTM) stacked autoencoder (L-SAE) model to extract sensor features and generate degradation trajectories, the gray similarity multiscale matching framework is used to compare the degradation trajectory sets in the training set with the degradation trajectory fragments of the test set samples, thereby determining the reference samples and performing RUL prediction. Then, the uncertainty of the prediction results is evaluated, and the kernel density estimation (KDE) method is used to analyze the prediction results. The main innovations and creativities of this study are given as follows.

- 1) A parameter-based cleaning method for aviation engine sensors was employed, combining the Mann–Kendall (MK) test with information entropy to select and screen performance parameter models.
- 2) An L-SAE model was proposed to fuse preprocessed performance parameters, generating a degradation index that reflects the degradation performance.
- 3) The gray similarity was introduced into the similarity measurement method, and the RUL prediction was achieved by combining it with the similarity matching rules of multiscale sliding time.

The remainder of this article is organized as follows. Section II describes the proposed method's detailed progress, including data preprocessing, building a degradation index, and RUL prediction. Section III illustrates the experiment process. Section IV concludes.

II. PROCEDURE OF THE DEGRADATION INDEX DEVELOPMENT

A. Data Cleaning

In this article, we use the C-MAPSS [30] public dataset to validate our method. Commercial aviation propulsion system simulators generate this dataset. This study focuses on the FD001 data subset, which includes rich signal data from 21 sensors. These sensor signals carry valuable information about the performance degradation during the engine operation. However, not all sensors contain relevant information about equipment degradation. To characterize the engine performance degradation process, it is necessary to identify sensor data that exhibits monotonic changes with service time.

The MK test is a nonparametric statistical method used to detect trends in time series data. All sensor signals are input into the MK test, and for each sensor, the test is applied to all samples. At a specific significance level, it is determined whether the sensor data in each sample exhibits a monotonic trend. If the MK [31] test identifies a significant trend in at least 90 out of 100 measurements for a sensor, it suggests that the sensor merits further investigation. The pass rates for all sensors in the 100 training set samples using the MK test are shown in Fig. 1, where the dark dashed line represents the threshold with a pass rate of 0.9.

In this study, it is considered that when the pass rate of the MK test for a sensor signal is greater than 90%, the data from that sensor contains significant degradation information and is effective for RUL prediction. Based on the above results,

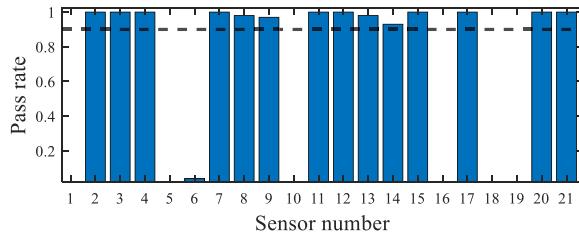


Fig. 1. Sample pass rate of MK test.

sensors with the following numbers, namely 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, and 21, were ultimately selected for further processing.

Information entropy measures the disorder or uncertainty in a system. The higher the information entropy, the greater the uncertainty of a random variable, indicating a larger amount of information. To reduce computational complexity, enhance calculation precision, and facilitate the Comparison of information entropy values across different time series, a windowing approach is employed to partition the time series into multiple windows. This enables entropy calculations for the data within each window, and ultimately, the average entropy value is computed for the entire time series. Considering variations in the lengths of different samples, a window length equivalent to 10% of each time series length is selected.

To reduce computational complexity, enhance calculation precision, and facilitate the comparison of information entropy values across different time series, a windowing approach is employed to partition the time series into multiple windows. This enables entropy calculations for the data within each window, and ultimately, the average entropy value is computed for the entire time series. Considering variations in the lengths of different samples, a window length equivalent to 10% of each time series length is selected.

The definition of information entropy for sensor parameters as continuous time variables is shown in the following equation:

$$H(X) = - \int_{D_X} f(x) \log(f(x)) \quad (1)$$

where $f(x)$ is the probability distribution function of the continuous variable X , and D_X is the domain of X . For the logarithm function in the formula, a uniform base of 2 is used, making the unit of information entropy bit.

From Fig. 2, it can be observed that although there are differences in the information entropy of all sensor parameters, their numerical values are all greater than 0.95, indicating a certain level of complexity. Therefore, it can be considered that the selected 14 sensors all contain a significant amount of degradation information.

B. Data Normalization

Considering that the dataset contains signals from multiple sensors, each of which may use different units of measurement for the physical quantities, and that the measurement ranges of different sensors can vary significantly, it is possible that the readings from certain sensors may be more prominent,

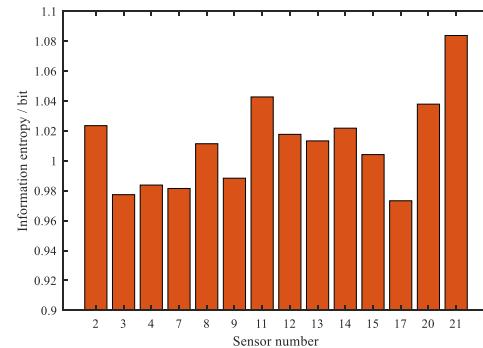


Fig. 2. Information entropy of sensor information.

while the contributions from others may be relatively minor. For these reasons, normalizing the data is essential. Normalization brings all indicators to the same order of magnitude, reducing the bias caused by differences in the magnitude of sensor values and ensuring a more balanced contribution from each sensor to the overall data analysis. Common normalization methods include Z-Score standardization, the Min-Max method, the Softmax function, and the Sigmoid function. To preserve the relative relationships in the original data, this article adopts the Min-Max method to normalize the selected 14-D sensor signals, mapping the data to a specific range through linear transformation. The specific formula for this method is as follows:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

where X represents the original signal data, X_{\max} and X_{\min} correspond to the maximum and minimum values in the original data, respectively, and X_{norm} denotes the signal data obtained after normalization.

C. Smooth Filtering

After normalizing the sensor data following data cleaning and considering the potential presence of noise components in the signal, further smoothing filtering is applied to the data.

The exponential moving average (EMA) method assigns different weights to all observed values based on the time elapsed between the observation moment and the current moment. However, the exponential weighted average method with a fixed smoothing parameter still has certain limitations in practical use. To address the issues mentioned above, this chapter proposes an improved exponential weighted average filtering method based on dynamically adjustable smoothing parameters. For different stages of changes in the time series, adaptive smoothing parameters are employed, using smaller values when the data changes are minor and larger values when the data changes are significant. The new expression is as follows:

$$E_t = \alpha_t \cdot x_t + (1 - \alpha_t) \cdot E_{t-1} \quad (3)$$

where α_t is the dynamically adjustable smoothing parameter, which is designed to determine whether to use a smaller smoothing parameter α_{\min} or a larger smoothing parameter α_{\max} based on the rate of change within the window of the

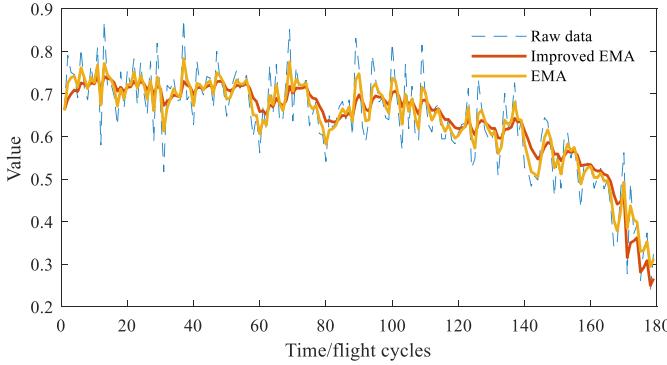


Fig. 3. Comparison of signal filtering effects.

original data. The specific calculation formula is as follows:

$$\alpha_t = \begin{cases} \alpha_{\min}, & \text{rate}_t < R \\ \alpha_{\max}, & \text{rate}_t \geq R \end{cases} \quad (4)$$

$$\text{rate}_t = \frac{|x_{t-N+1} - x_t|}{x_t} \quad (5)$$

where N is the size of the sliding window, rate_t represents the rate of change of the original signal within the sliding window of size N at time t , and R is the threshold used to evaluate the rate of change within the window.

The effects of processing using both the original EMA method, and the improved EMA method are shown in Fig. 3. For the original EMA method, the α was set to 0.5, while for the improved method, the sliding window size was set to 10, the rate of change threshold R was set to 0.4, and the smoothing parameters $\alpha_{\max} = 0.8$ and $\alpha_{\min} = 0.2$ were applied. As observed in the figure, both filtering methods significantly reduced the impact of noise on the original signal. However, in terms of smoothness and denoising effectiveness, the improved method outperformed the original EMA method, providing more effective smoothing and noise reduction for the original signal.

D. L-SAE Build Degradation Index

During the operation of the equipment, multiple sensors monitor the equipment simultaneously, resulting in multiple observations forming a multidimensional time series. Using dimensional reduction methods to fuse multidimensional data into a degradation index to monitor degradation helps to consider comprehensively monitoring information from multiple perspectives.

A stacked autoencoder is a type of neural network model that learns to encode and decode data. It consists of multiple stacked autoencoders, each of which learns to encode and decode its input data. The stacked autoencoder learns how to map high-dimensional input data to low-dimensional representations during training. Compared to ordinary autoencoders, the multilayer structure of stacked autoencoders enables them to learn multilevel representations of data, from low-level features to high-level abstract features, which helps the model better understand and represent data.

Given that sensor data is temporally correlated, LSTM layers are used to capture the temporal dependencies of the

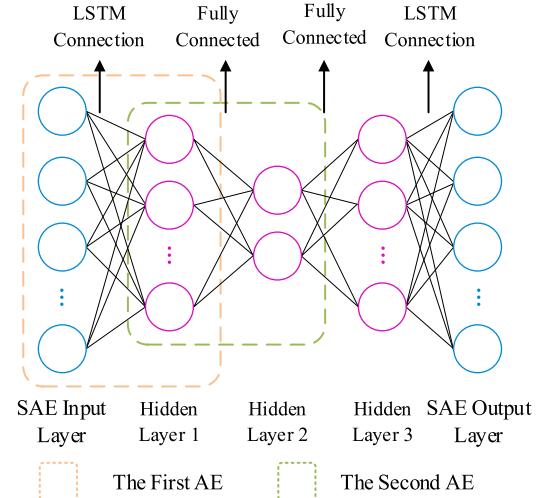


Fig. 4. L-SAE structural diagram.

data. To enhance the mapping ability of the autoencoder, a two-stacked structure consisting of LSTM neural networks, also known as the L-SAE model, is adopted, as observed in Fig. 4.

The hidden layers 1 and 3 use the ReLU function as the activation function, which enhances the sparse nature of the network and improves training speed. The output layer uses the Tanh function as the activation function. The mean squared error function is selected as the loss objective function for the L-SAE model, and Adam is used as the optimizer for model training.

Suppose the autoencoder only learns and reconstructs multidimensional sensor data in normal states. In that case, its performance will be good when reconstructing normal data, but it may produce large errors when reconstructing degraded data. That is to say, the reconstruction error can reflect the degradation level of the equipment. The greater the degradation, the larger the reconstruction error, and vice versa. Therefore, the reconstruction error can be used as a health index. This method helps to eliminate the differences in health indices between individuals caused by unknown factors, improving the robustness of the index to noise.

In general, turbofan engines are in normal operation at the initial stage, and this initial state occupies about 20% of the entire life cycle [32]. Therefore, this article chooses to use the data of the first 20% of the operating cycles of each sample in all training sets in the FD001 dataset as the training set data to train the L-SAE model in an unsupervised manner. The purpose of this strategy is to utilize the data of the initial normal operation state to learn through the L-SAE model, thereby better capturing the operational characteristics and degradation trends of the engine.

Assuming that the entire life cycle data of the i th training sample is $X^i = \{x_1^i, x_2^i, \dots, x_n^i\}$, and only the normal operating cycle data $\{x_1^i, x_2^i, \dots, x_{\text{normal}}^i\}$ is used to train the L-SAE model, the degradation index of the i th training unit at time t can be obtained by reconstructing the error $\varepsilon_t^i = \|x_t^i - \hat{x}_t^i\|_2$, where x_t^i is entire life cycle data of the i th training sample, and \hat{x}_t^i is reconstruction sample. The calculation formula of \hat{x}_t^i

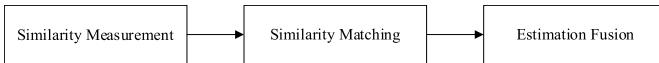


Fig. 5. Similarity matching RUL prediction process.

is as follows:

$$\hat{X} = f_d(W'Z + b_d) \quad (6)$$

where $f_d(\cdot)$ is the activation function of the decoder, $\theta_d = \{W', b_d\}$ represents the network parameters of the decoder, and W' and b_d are the weight and bias of the decoder, respectively.

III. PROCEDURE FOR RUL PREDICTION

After the building of the degradation index, the next step is the prediction of RUL with the developed degradation index. This section introduces the procedure of RUL prediction, as shown in Fig. 5.

A. Similarity Measurement Based on GRA

The similarity measurement method must effectively capture the resemblance between equipment states, ensuring that the selected metric accurately reflects the similarity between degradation behaviors. In this article, we employ gray relational analysis (GRAs) for similarity measurement. GRA, a multifactor statistical method, evaluates the influence of various factors in systems with incomplete information, ranking them based on their impact. It is particularly useful for handling incomplete or unstable data due to its adaptability and interpretability. GRA quantifies similarity by comparing trend patterns rather than absolute values. The general steps of GRA are as follows.

- 1) *Positive Exponential Transformation*: Convert the minimization, intermediate, and interval indices uniformly into maximization indices. In a maximization index, a larger value represents a better outcome.
- 2) *Identify and Analyze the Sequence*: It is necessary to determine the main sequence and the subsequence objects. The main sequence typically encompasses the primary trends or characteristics that need to be analyzed, while the subsequences refer to the data sequences that constitute the factors capable of explaining or influencing the overall behavior of the system.
- 3) *Preprocess the Variables*: To eliminate the impact of dimensions and reduce the range of variables to simplify calculations, preprocessing operations are necessary. The method of mean normalization is used to remove dimensions from the main sequence and all subsequences, which can be calculated using the following equation:

$$f(x(k)) = \frac{x(k)}{\bar{x}} = y(k) \quad (7)$$

where $x(k)$ represents the comparison sequence, that is, the subsequence, and $\bar{x} = (1/n) \sum_{k=1}^n x(k)$ denotes the mean of the corresponding subsequence.

- 4) The correlation coefficients between each index in the subsequence and the parent sequence can be computed using the following equation (8), as shown at

the bottom of the next page, where $x_0(k)$ represents the parent sequence, $x_i(k)$ represents the i th subsequence, $\min_i \min_k |x_0(k) - x_i(k)|$ represents the distance on the dimension closest to the parent sequence among all dimensions of all subsequences. ρ is a coefficient controlling the discriminant ability, also known as the resolution coefficient, generally taking values in the range $[0, 1]$, where a smaller ρ implies stronger discriminative ability, with 0.5 often considered appropriate.

- 5) Take the mean of the correlation coefficients for each subsequence to obtain the final gray correlation degree

$$\rho_{oi} = \frac{1}{m} \sum_{k=1}^m \xi_i(k). \quad (9)$$

In the above equation, the result is obtained by taking the average of the total sum. If each index plays a different role in the comprehensive evaluation, a weighted average of the correlation coefficients can be calculated

$$\rho'_{oi} = \frac{1}{m} \sum_{k=1}^m w_k \cdot \xi_i(k). \quad (10)$$

Overall, GRA reveals the similarity and degree of association between different sequences by calculating and comparing their relational coefficients.

B. Similarity Matching

To identify similar degradation behaviors in the training set compared to the test samples, you could traverse all lengths of degradation behaviors. However, this approach significantly increases the computational load. To simplify the search, it is common to select degradation behaviors of fixed length for the test samples.

Given a certain number of training set reference samples, where the health index of the i th reference sample is denoted as $H^i = \{h_1^i, h_2^i, \dots, h_{T_i}^i\}$, and T_i is the full lifespan length of the i th reference sample, representing the duration until failure or damage occurs. The health index of the j th test sample is $Z^j = \{z_1^j, z_2^j, \dots, z_{t_j}^j\}$, where t_j is the last recorded moment for this sample, at which point the test sample has not yet experienced failure or damage.

In extracting degradation information from samples, the concept of a sliding time window is used. Assuming the window size is τ , it is generally considered that the last degradation trajectory segment of the test sample contains more degradation information and better represents the degradation state of the sample. Therefore, the last trajectory segment $Z_\tau^j = \{z_{t_j-\tau+1}^j, z_{t_j-\tau+2}^j, \dots, z_{t_j}^j\}$ of the test sample is selected as the object for similarity matching. For reference samples, it is usually assumed that the overall information of the reference samples needs to be considered. Therefore, $T_i - \tau + 1$ trajectory segments of length τ are divided from the reference sample, specifically represented as $H_{t_m}^i = \{h_m^i, h_{m+1}^i, \dots, h_{m+\tau}^i\}$, where $m = 1, 2, \dots, T_i - \tau + 1$. These trajectory segments constitute the reference segments of the i th reference sample.

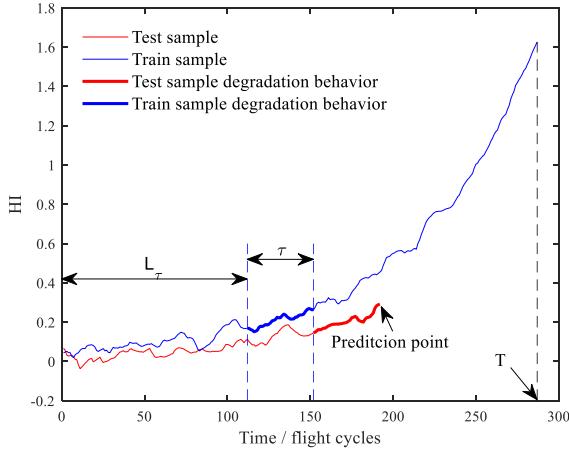


Fig. 6. Schematic of similar matching RUL prediction.

The specific similarity matching prediction method can be seen in Fig. 6. Here, we use the reconstruction error ε_t^i as a HI. This approach helps eliminate the differences in health indices between individuals caused by unknown factors, thereby enhancing the index's robustness against noise. In this work, the reconstruction error is employed to measure the device's health condition, making the constructed health index more representative. The red and blue curves represent the degradation behaviors of the test sample and the reference sample, respectively. For the test sample, a segment from the end of the trajectory is selected (highlighted in red) and compared with multiple trajectory segments from the reference sample to find similar matches, thereby pinpointing the highlighted blue segment. This allows for the prediction of the RUL of the test sample.

Assuming a sliding window group $M = \{\tau_1, \tau_2, \dots, \tau_n\}$, where n is the number of different-sized sliding windows in the group and $0 < \tau_1 < \tau_2 < \dots < \tau_n$. Considering the presence of small-sized samples in the test set, to ensure that each test sample can be processed by the sliding window group $\{\tau_1, \tau_2, \dots, \tau_n\}$, the minimum sliding window size must be smaller than the minimum length of the test sample, that is $\tau_1 \leq T_{\min}^{\text{test}}$. Additionally, it is required to ensure that the sliding window does not exceed the length of any test sample and that the maximum sliding window length is smaller than the minimum length of the reference sample. This ensures that each reference sample can be processed by the sliding window group $M = \{\tau_1, \tau_2, \dots, \tau_n\}$, so the maximum sliding window size must also satisfy $\tau_n \leq \min(T_{\max}^{\text{test}}, T_{\min}^{\text{train}})$. To better describe the overall degradation of reference and test samples, the set $\{\tau_1, \tau_2, \dots, \tau_n\}$ is defined as an arithmetic progression.

Different sliding window sizes are used because the right size is essential for retaining effective degradation information. A size that is too small may miss important details, while one that is too large may add redundant data. Since mechanical equipment degradation occurs in stages with varying rates,

a multiscale sliding window group is employed to account for these differences. This method helps capture key degradation information more comprehensively, reducing information loss and redundancy, and improving similarity matching accuracy and robustness.

For the FD001 sub-dataset, considering that the minimum length of the test samples is 31, and the minimum time scale should not exceed the minimum length of the samples, τ_1 is set to 31. Additionally, if the time scale is greater than 100, the extracted segments may contain too much interference noise. Therefore, the maximum length of the sliding window is set to be less than 100. For uniform partitioning, the interval between time scales is ultimately set to 20, i.e., $\Delta\tau = 20$. Considering the above conditions, the final set of multiscale time windows is determined to be (31, 51, 71, 91). This design considers the characteristics of the test sample length, ensuring that the selected sizes can cover the time range of the samples and avoid introducing too much noise with overly small or large windows.

C. Estimation Fusion

After using multiple sizes of sliding time windows to process the trained health indicators and establish a degradation model reference library, each size is used to extract segments from the test degradation trajectory and match them with the reference library. By calculating gray relational degrees, similar trajectory segments are identified in the degradation model reference library and treated as reference samples.

At each time scale, a series of similar segments satisfying the conditions can be selected using the following equation, where I represents the set of all training units in the degradation model reference library, λ is the adjustment factor used to adjust the similarity requirements for reference segments, and W_k is the set of training units that satisfy the conditions:

$$\rho_{\tau_k}^{i,j} \geq \lambda \max_{s \in I} (\rho_{\tau_k}^{s,j}), \quad i \in W_k. \quad (11)$$

Suppose, at time scale τ_k , the starting point of the degradation trajectory segment for the i th training unit of the j th test sample is denoted as $L_{\tau_k}^{i,j}$, where T_i represents the full lifespan length of the i th reference sample training unit. The RUL of the i th reference sample at time scale τ_k can be calculated using the following equation, where the parameters are as indicated and can be referenced:

$$\text{RUL}_{\tau_k}^{i,j} = T_i - L_{\tau_k}^{i,j} - \tau_k + 1. \quad (12)$$

At each scale, a series of predicted RUL values $\{\text{RUL}_{\tau_k}^{i,j} | i \in W_k\}$ can be obtained. Generally, it is believed that the higher the similarity between the degradation trajectory segments of the test sample and the reference sample, the more similar the degradation performance of the two samples. Therefore, the RUL values of these reference sample degradation trajectory segments will be closer to the actual RUL

$$\zeta_i(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \cdot \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \cdot \max_i \max_k |x_0(k) - x_i(k)|} \quad (8)$$

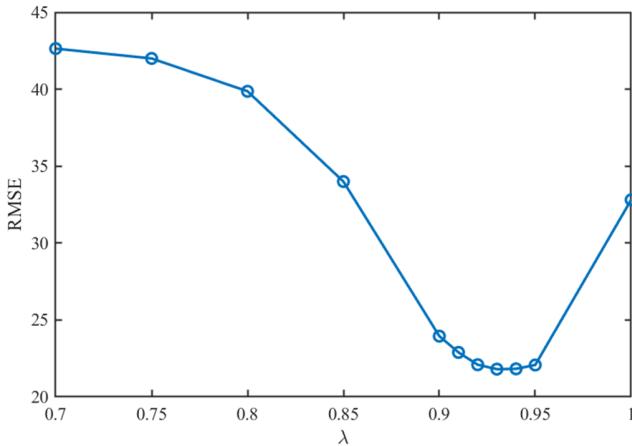


Fig. 7. Result of adjusting factor parameter iteration.

value of the test sample. The weight of the RUL $RUL_{\tau_k}^{i,j}$ for the i th reference sample determined by the j th test sample at time scale τ_k is obtained based on gray similarity using the following equation, where $\rho_{\tau_k}^{i,j}$ represents the similarity corresponding to the degradation trajectory segment of the i th reference sample training unit found for the j th test sample at time scale τ_k :

$$w_{\tau_k}^{i,j} = \frac{1/\rho_{\tau_k}^{i,j}}{\sum_{s \in I} 1/\rho_{\tau_k}^{s,j}}, \quad i \in W_k. \quad (13)$$

At each time scale, a series of weights can be obtained according to the above formulae. These weights, along with the RULs of the training units that satisfy the conditions, can be used to calculate the RUL of the j th test sample. Specifically, see the following equation:

$$RUL^j = \sum_{\tau_k \in M} \sum_{i \in W_k} w_{\tau_k}^{i,j} \cdot RUL_{\tau_k}^{i,j}. \quad (14)$$

After performing multitime scale similarity matching on all turbofan engine fan blades in the training set, it is necessary to determine an appropriate adjustment factor λ to select a certain amount of reference samples that meet the similarity coefficient requirements. If this value is too large, the final number of reference samples may be too small, leading to higher errors. If the value is too small, some reference samples with relatively low similarity may be used to predict the final results, introducing errors. Considering that the numerical value of this parameter does not have a significant impact on other parameters, an iterative optimization approach is adopted to determine its value. After iterative optimization, the adjustment factor is determined to be set at 0.93. The specific iteration process can be seen in Fig. 7.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The entire process of RUL prediction of aeroengine is shown in Fig. 8.

To further demonstrate the effectiveness and feasibility of the proposed method in this article, we employ penalty functions and root mean square error (RMSE) for evaluation.

The specific formulas are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (\Delta_j)^2} \quad (15)$$

$$Score = \sum_k S_j, S_j = \begin{cases} e^{-\frac{\Delta_j}{b_1}} - 1, & \Delta_j \leq 0 \\ e^{-\frac{\Delta_j}{b_2}} - 1, & \Delta_j > 0. \end{cases} \quad (16)$$

In the formula, Δ_j represents the absolute error, and Δ_j is set to $EstimateRUL_j - TrueRUL_j$, k is the sample number of test dataset, and $b_1 = 10, b_2 = 13$.

In addition, a new evaluation metric, namely the composite indicator (CI), has been introduced. This metric comprehensively considers the impact of both RMSE and penalty scores on the predictive results. Its aim is to balance the importance of accuracy and timeliness when assessing model performance. Specifically, the design intent of this metric is to emphasize the importance of RMSE for accuracy while highlighting the emphasis of penalty scores on timeliness. Following equation gives the specific formula:

$$CI = 0.3 \times RMSE + 0.7 \times Score. \quad (17)$$

Typically, the engine unit operates normally in the early stages and then undergoes linear degradation. According to research findings, considering the negligible losses of a turbofan engine in the initial stages of operation, it can be assumed that the RUL of the engine remains constant before the onset of degradation. In this article, referring to previous studies [32], a value of $R_{early} = 125$ is adopted. By setting RUL labels during the early degradation phase, the RUL labels of the training set are adjusted accordingly, as illustrated in Fig. 9.

To analyze the predictive performance of the multiscale sliding window RUL prediction method, this article selects test samples 13, 16, 60, 93, and 99 from the FD001 dataset with different operating cycle lengths as typical examples. This is done to validate the proposed method's prediction accuracy for test samples with varying operating cycles. A comparison is made between single-scale prediction and multiscale sliding window prediction methods, using the relative error as the evaluation metric. The results of the comparison are shown in Table I. From Table I, it can be observed that although the samples have different operating cycle lengths, the predictive results of the multiscale model are superior to those of the single-scale model. Additionally, a comparison analysis of the number of reference samples reveals that the proposed multisample weighted method outperforms the weighted results obtained by selecting the most similar sample at each scale and then combining the predictions for the four samples. This emphasizes the importance of an adequate number of reference samples. Consequently, the multi-time-scale sliding window framework overcomes the accuracy limitations of a single fixed-scale matching, and the fusion of multiscale prediction results enhances the generalization capability of the method, contributing to improved generalization and accuracy in RUL prediction.

To further analyze the differences in the predictive performance of the SBP method at various operating durations, RUL predictions are made at fixed intervals starting from halfway

TABLE I
COMPARISON OF PREDICTION RESULTS BETWEEN SINGLE-SCALE AND MULTI-SCALE SIMILARITY MATCHING

Test Sample	Operating Cycle	Actual RUL	Prediction Error					
			Single Scale				Multi-scale	
			31	51	71	91	4 Sample Weighting	Weighted Diversity
#13	195	95	0.45	0.56	0.61	0.61	0.56	0.19
#16	113	84	0.37	0.37	0.38	0.38	0.38	0.07
#60	147	100	0.52	0.52	0.32	0.73	0.15	0.08
#93	244	85	0.79	0.72	0.72	0.84	0.39	0.07
#99	97	117	0.43	0.43	0.56	0.32	0.43	0.17

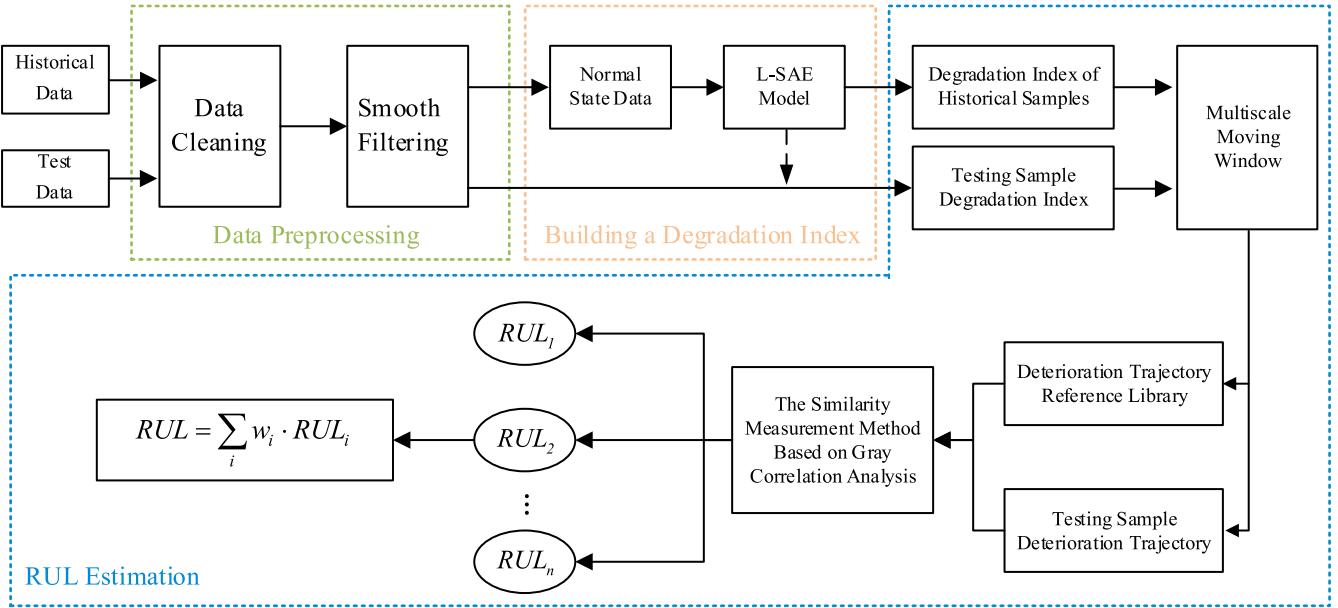


Fig. 8. Flowchart of aeroengine remaining life prediction based on L-SAE health index construction and gray similarity multiscale matching.

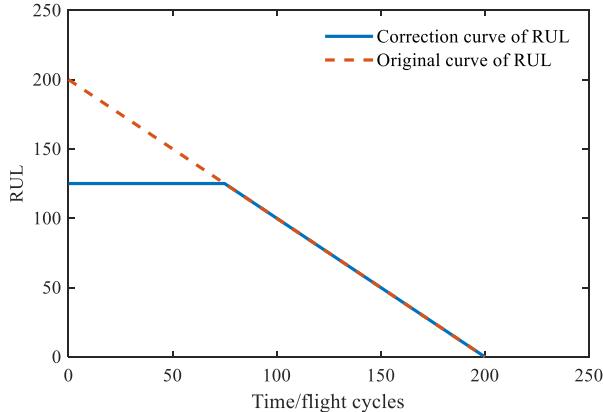


Fig. 9. Schematic of RUL label correction.

through the life cycle until the end of the life cycle. Given that the maximum sliding window size in the multiscale sliding window is 91, and to ensure an ample amount of data is available as the proportion of service life gradually increases, only test samples with actual operating cycle lengths exceeding 190 are considered. The selected evaluation metrics include mean absolute error (MAE), RMSE, and penalty scores. The evaluation metrics for the predicted results of test samples vary

TABLE II
EVALUATION METRICS FOR RUL PREDICTION RESULTS UNDER DIFFERENT RUNNING TIMES

The Proportion of Running Time to Total Time	MAE	RMSE	Penalty Scores
50%	56.01	64.48	4.83e+03
60%	51.80	59.96	4.13e+03
70%	42.70	49.87	1.58e+03
80%	27.60	33.61	439.25
90%	15.29	21.68	115.89
100%	11.51	16.02	103.08

with the operating duration, as shown in Table II, where MAE and RMSE exhibit a clear monotonic decrease trend.

From Table II, it can be observed that as the operating duration increases, the scores of all three evaluation metrics show a decreasing trend. In other words, as the equipment or system is used for a more extended period, the predictive errors for RUL gradually decrease. This trend aligns with the general pattern observed in early predictions, where data from the early stages of equipment operation contain less degradation information. With the increase in operating time, more degradation information accumulates in the data, leading

TABLE III
COMPARISON OF PREDICTION RESULTS OF RUL PREDICTION METHODS

Methods	RMSE	Score	CI
CNN-LSTM[33]	21.02	5585	973
DCNN	22.36	10412	7295
SVR[34]	24.87	1670	1176
Concurrent VAE[35]	25.22	7645	5359.066
AE[36]	24.32	4159	2918.596
Supervised[37]	24.98	3532	2479.894
ED-SBP (Ours)	21.78	964	681

TABLE IV
COMPARISON OF DIFFERENT DATA SMOOTH FILTERING METHODS

Data smooth filtering	MSE
Raw data	552.25
EMA	537.32
Improved EMA	474.36

TABLE V
PREDICTION RESULTS OF HI COMPARISON EXPERIMENT

Methods	RMSE	Score
PCA	41.6	15437
ISOMAP	43.03	8132
L-SAE	21.78	964

to more accurate RUL predictions. This implies that after a certain period of system operation, the prediction method can provide more accurate estimates of future service life. This study reflects that in the early stages of equipment operation, the accuracy of the prediction model may be relatively low due to limited degradation information. However, over time, as the equipment experiences more operating conditions, the prediction model can better learn and adapt to the true operating state of the system, thereby improving the accuracy of predictions.

To further validate the effectiveness of the proposed method, three evaluation metrics were employed to compare the predictive performance of this study with several methods from more recent literature. The comparative results are presented in Table III.

In the previous Section II, we proposed using the improved EMA for smoothing the data. Here, we compared the original data with the EMA prediction results, and the experimental results are as follows.

As shown in Table IV, the proposed improved EMA demonstrates its advantages by reducing the MSE and enhancing prediction accuracy.

Additionally, we compared the PCA and ISOMAP methods with the L-SAE model proposed in this study, while keeping other variables unchanged. The RUL prediction was conducted using a similarity-based matching approach, with RMSE and the penalty function as evaluation indices. The experimental results are shown in Table V.

By comparing RMSE and the penalty function, it can be observed that the prediction indices obtained through the L-SAE dimensionality reduction are lower than those of other

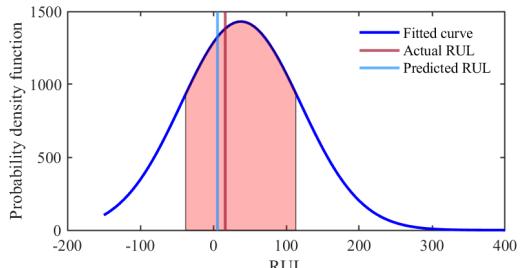


Fig. 10. Evaluation of uncertainty in engine RUL prediction.

methods. This indicates that in the task of RUL prediction, L-SAE outperforms other dimensionality reduction methods, leading to a more accurate estimation of the equipment's remaining life in practical applications.

To depict the uncertainty of RUL prediction results, this section employs the KDE method that can fit the distribution of RUL, which is based on acquired initial RUL and corresponding weights. As a nonparametric method, KDE does not require prior knowledge and can be used to estimate unknown probability density functions. Given the convenient mathematical properties of Gaussian kernels and their common usage in the absence of prior knowledge about RUL distribution, this section adopts the widely used Gaussian KDE method.

As an example, uncertainty assessment for the predicted results of the 19th sample in the test set is illustrated in Fig. 10. The red and light blue vertical lines represent the true RUL and the predicted RUL, respectively. The blue curve shows the probability density function of the predicted RUL, while the pink-shaded area indicates the range covered by the 95% confidence interval of the prediction. According to the model's prediction, the predicted RUL for engine sample No. 19 in the test set is 67, with a 95% confidence interval of (35, 101). A 95% confidence interval means that in repeated sampling and interval construction, we expect 95% of these intervals to contain the true parameter value. In this experiment, since the true RUL is 87, we consider our prediction to be reliable.

In terms of computational complexity, we conducted experiments on a machine equipped with an i9-13900K CPU and an Nvidia RTX 4080 GPU. The AE component was implemented in Python using PyTorch, while the RUL component was developed in MATLAB. The Python execution took 5 s, and the MATLAB execution took 13 s, resulting in a total runtime of 18 s. Additionally, we ran the program on a machine without GPU acceleration, which took approximately 20 s. This demonstrates the advantage of our method in practical applications, where real-time performance is a critical factor, particularly in RUL prediction for aircraft engines.

V. CONCLUSION

This article employs a parameter-based approach for cleaning sensor data from aviation engine sensors. Utilizing the MK test-based parameter selection method, the study filters out complex information in the dataset by considering the monotonic performance of sensors, excluding sensor data unrelated to engine degradation. Subsequently, a degradation trajectory construction method based on L-SAE is applied to the filtered

sensor data, involving normalization and smoothing for noise reduction. The L-SAE model is then utilized to fuse and reduce the dimensionality of the multidimensional data, forming a degradation index and thereby constructing the engine's degradation trajectory. Finally, this article proposes and validates a RUL prediction method based on gray similarity multiscale matching. Through gray correlation analysis and multiscale matching techniques, the method comprehensively considers the dynamic changes in the engine's operational state, leading to improved prediction accuracy and stability. Experimental results demonstrate the superiority of the proposed approach.

REFERENCES

- [1] Y. Hu, X. Miao, Y. Si, E. Pan, and E. Zio, "Prognostics and health management: A review from the perspectives of design, development and decision," *Rel. Eng. Syst. Saf.*, vol. 217, Jan. 2022, Art. no. 108063.
- [2] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2020.
- [3] J. Chiachío, M. L. Jalón, M. Chiachío, and A. Kolios, "A Markov chains prognostics framework for complex degradation processes," *Rel. Eng. Syst. Saf.*, vol. 195, Mar. 2020, Art. no. 106621.
- [4] B. Xue, F. Xu, X. Huang, Z. Xu, and X. Zhang, "Improved similarity based prognostics method for turbine engine degradation with degradation consistency test," *Appl. Intell.*, vol. 52, no. 9, pp. 10181–10201, Jul. 2022.
- [5] B. Yan, X. Ma, G. Huang, and Y. Zhao, "Two-stage physics-based Wiener process models for online RUL prediction in field vibration data," *Mech. Syst. Signal Process.*, vol. 152, May 2021, Art. no. 107378.
- [6] P. Wang, R. X. Gao, and W. A. Woyczyński, "Lévy process-based stochastic modeling for machine performance degradation prognosis," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12760–12770, Dec. 2021.
- [7] P. Shankar Kumar, L. A. Kumaraswamidhas, and S. K. Laha, "Bearing degradation assessment and remaining useful life estimation based on kullback-leibler divergence and Gaussian processes regression," *Measurement*, vol. 174, Apr. 2021, Art. no. 108948.
- [8] I. Sanz-Gorrachategui et al., "Remaining useful life estimation for LFP cells in second-life applications," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [9] H. Zhao, H. Liu, Y. Jin, X. Dang, and W. Deng, "Feature extraction for data-driven remaining useful life prediction of rolling bearings," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [10] Z. Pan, Z. Meng, Z. Chen, W. Gao, and Y. Shi, "A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106899.
- [11] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9521–9530, Dec. 2019.
- [12] J. Wu, M. Wu, Z. Chen, X. Li, and R. Yan, "Degradation-aware remaining useful life prediction with LSTM autoencoder," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [13] Y. Ding, P. Ding, X. Zhao, Y. Cao, and M. Jia, "Transfer learning for remaining useful life prediction across operating conditions based on multisource domain adaptation," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 5, pp. 4143–4152, Oct. 2022.
- [14] J. Liu, D. Djurdjanovic, J. Ni, N. Casoetto, and J. Lee, "Similarity based method for manufacturing process performance prediction and diagnosis," *Comput. Ind.*, vol. 58, no. 6, pp. 558–566, Aug. 2007.
- [15] T. Wang, J. Yu, D. Siegel, and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," in *Proc. Int. Conf. Prognostics Health Manage.*, Oct. 2008, pp. 1–6.
- [16] S. Khaleghi et al., "Developing an online data-driven approach for prognostics and health management of lithium-ion batteries," *Appl. Energy*, vol. 308, Feb. 2022, Art. no. 118348.
- [17] S. Schwartz, J. J. M. Jiménez, R. Vingerhoeds, and M. Salaün, "An unsupervised approach for health index building and for similarity-based remaining useful life estimation," *Comput. Ind.*, vol. 141, Oct. 2022, Art. no. 103716.
- [18] B. Cui and J.-C. Créput, "NCC based correspondence problem for first- and second-order graph matching," *Sensors*, vol. 20, no. 18, p. 5117, Sep. 2020.
- [19] H. Cai, X. Jia, J. Feng, W. Li, L. Pahren, and J. Lee, "A similarity based methodology for machine prognostics by using kernel two sample test," *ISA Trans.*, vol. 103, pp. 112–121, Aug. 2020.
- [20] T. Wang, "Trajectory similarity based prediction for remaining useful life estimation," Ph.D. dissertation, Univ. Cincinnati, 2010.
- [21] W. Yu, I. Y. Kim, and C. Mecheske, "An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme," *Rel. Eng. Syst. Saf.*, vol. 199, Jul. 2020, Art. no. 106926.
- [22] H. Cai, J. Feng, W. Li, Y.-M. Hsu, and J. Lee, "Similarity-based particle filter for remaining useful life prediction with enhanced performance," *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106474.
- [23] W. Sun, Y. Zhou, X. Cao, B. Chen, W. Feng, and L. Chen, "A two-stage method for bearing fault detection using graph similarity evaluation," *Measurement*, vol. 165, Dec. 2020, Art. no. 108138.
- [24] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
- [25] R. Zemouri and R. Gouriveau, "Towards accurate and reproducible predictions for prognostic: An approach combining a RRBF network and an AutoRegressive model," *IFAC Proc. Volumes*, vol. 43, no. 3, pp. 140–145, 2010.
- [26] C. Sankavaram et al., "Model-based and data-driven prognosis of automotive and electronic systems," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2009, pp. 96–101.
- [27] P. Wang, B. D. Youn, and C. Hu, "A generic probabilistic framework for structural health prognostics and uncertainty management," *Mech. Syst. Signal Process.*, vol. 28, pp. 622–637, Apr. 2012.
- [28] F. Wang and T. Mamo, "Hybrid approach for remaining useful life prediction of ball bearings," *Qual. Rel. Eng. Int.*, vol. 35, no. 7, pp. 2494–2505, Nov. 2019.
- [29] D. An, N. H. Kim, and J. H. Choi, "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews," *Rel. Eng. Syst. Saf.*, vol. 133, pp. 223–236, Jan. 2015.
- [30] A. Saxena, K. Goebel, D. Simon, and N. H. W. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manag.*, 2008, pp. 1–9.
- [31] M. G. Kendall, *Rank Correlation Methods*, 4th ed. London, U.K.: Charles Grifin, 1975.
- [32] T. Xia, J. Shu, Y. Xu, Y. Zheng, and D. Wang, "Multiscale similarity ensemble framework for remaining useful life prediction," *Measurement*, vol. 188, Jan. 2022, Art. no. 110565.
- [33] S. Yu, Z. Wu, X. Zhu, and M. Pecht, "A domain adaptive convolutional LSTM model for prognostic remaining useful life estimation under variant conditions," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 130–137.
- [34] K. T. P. Nguyen and K. Medjaher, "An automated health indicator construction methodology for prognostics based on multi-criteria optimization," *ISA Trans.*, vol. 113, pp. 81–96, Jul. 2021.
- [35] T. Wang, D. Guo, and X.-M. Sun, "Remaining useful life predictions for turbofan engine degradation based on concurrent semi-supervised model," *Neural Comput. Appl.*, vol. 34, no. 7, pp. 5151–5160, Apr. 2022.
- [36] T. Krokotsch, M. Knaak, and C. Güthmann, "Improving semi-supervised learning for remaining useful lifetime estimation through self-supervision," *Int. J. Prognostics Health Manage.*, vol. 13, no. 1, pp. P1–P19, Jan. 2022.
- [37] M. Natsumeda and T. Yairi, "Consistent pretext and auxiliary tasks with relative remaining useful life estimation," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 6879–6888, Apr. 2024.



Shilong Sun (Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2018.

He is currently an Assistant Professor with Harbin Institute of Technology, Shenzhen, China. His nurtures keen interests in vibration energy harvesting design, fault diagnosis and prognosis, decision-making with artificial intelligence, and deep learning for industrial data. Now, he focuses on the remaining equipment life estimation research with deep learning and smart energy harvesting techniques.



Haodong Huang was born in Jingmen, China. He received the bachelor's degree in mechanical engineering from the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, China, in 2021, and the master's degree from Harbin Institute of Technology (Shenzhen), Shenzhen, China, in 2024, where he is currently pursuing the Ph.D. degree in mechanical engineering with the School of Mechanical Engineering and Automation.

His research interests include mechanical fault diagnosis and humanoid robots.



Dong Wang (Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2015.

He was a Senior Research Assistant, a Post-Doctoral Fellow, and a Research Fellow with the City University of Hong Kong. He is currently an Associate Professor with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China, where he is also with the State Key Laboratory of Mechanical System and Vibration. His research interests include sparse and complex measures, signal processing, prognostics and health management, condition monitoring and fault diagnosis, statistical learning and machine learning, statistical process control, and nondestructive testing.

Dr. Wang is an Editorial Board Member of Mechanical Systems and Signal Processing. He is an Associate Editor of IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, IEEE SENSORS JOURNAL, and *Journal of Dynamics Monitoring and Diagnostics*.



Jian Ding was born in Wuhan, China. She received the B.S. degree in mechanical engineering from the School of Mechanical Engineering and Automation, Dalian University of Technology of China, Dalian, China, in 2020. She is currently pursuing the M.S. degree in mechanical engineering with School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China.

Her research interests include mechanical fault diagnosis and condition monitoring.



Wenfu Xu (Senior Member, IEEE) received the B.E. and M.E. degrees in control engineering from Hefei University of Technology, Hefei, China, in 2001 and 2003, respectively, and the Ph.D. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2007.

He was a Research Associate with the Department of Mechanical and Automation Engineering at Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. His research interests include reconfigurable robots, space robots, and bionic robots.