A Novel Bearing Remaining Useful Life Prediction Methodology with Slope-Based Change Point Detection and WOA-Attention-Bilstm Model

Guangqi Qiu [1], Binlu Ye [1], Yingkui Gu [1], Peng Huang [1], He Li [2 *], Zifei Xu [3]

[1] *School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou, China*

[2] *School of Engineering, Liverpool John Moores University, Liverpool, 3 Byrom Street, L3 3AF, UK (e-mail: he.li@centec.tecnico.ulisboa.pt).*

[3] *School of Engineering, University of Liverpool, Liverpool, UK*

Abstract: Appropriate health indicator (HI) and efficient prediction models are critical factors for accurate remaining useful life (RUL) prediction, particularly, when dealing with fluctuations and redundant information in the HI curve. To address these challenges, this study proposed an HI construction method for better characterization of the degradation behavior based on the entropy weight method and kernel entropy component analysis (EWM-KECA). The HI construction method can eliminate the fluctuations in HI and identify the fault change point position in HI. For RUL estimation, a bearing RUL prediction method was developed by integrating slope-based change point detection with a whale optimization algorithm (WOA)-Attention-bidirectional long short-term memory (BiLSTM) model. By eliminating more than 85% of duplicate data that is not useful for RUL prediction, this approach achieves more accurate RUL predictions while reducing computational resource requirements. The reliability and effectiveness of the proposed method are validated using the bearing degradation dataset. The results from comparative analysis and ablation experiments demonstrate that the proposed method consistently achieves superior performance. Compared with models such as CNN-Attention-BiGRU, WOA-CNN-BiGRU, and WOA-Attention-CNN, the mean absolute error, mean absolute percentage error, and root mean squared error values have been reduced by more than 50%, indicating the proposed RUL prediction methodology represents an advanced and effective approach.

Index Terms: Remaining useful life, bearing, health indicator construction, slope change point detection, neural network

---

[*] Corresponding author e-mail: he.li@centec.tecnico.ulisboa.pt (He Li), zifei.xu@liverpool.ac.uk (Zifei Xu)

I. Introduction

EARING is an indispensable component in many mechanical systems, and its performance directly affects the system's reliability and safety [1]. Through accurate RUL prediction, the bearing potential faults are found in advance, and the corresponding maintenance measures are taken to reduce unexpected downtime and damage [2]. However, the damage of bearing during operation is not observed directly, and constructing appropriate HI that accurately reflects the system degradation trend, has become a key issue for RUL prediction [3-5].

For a suitable HI construction, many scholars [6-8] have made a lot of efforts in synthesizing the degradation information of various aspects of the system from the hybrid metrics [9, 10] such as monotonicity, robustness, trendability, and consistency, compared with the single HI such as RMS [11], kurtosis [12] and peak-to-peak value [13], the performance of RUL prediction has been greatly improved. However, in the process of HI construction with this strategy, many problems emerged, which led to inaccurate RUL prediction, which is summarized as follows.

(a) The weights of the prognostic metrics. HI construction by linear combination is one of the most popular approaches, and the weight determination of the prognostic metrics is very important. By the prognostic metrics, the feature parameters that best represent the bearing degradation behavior are selected, but, the different combinations of the prognostic metrics are led to select different sensitive features. It is necessary to develop an algorithm to determine the best weight, to build an optimal combination.

(b) The fluctuations in the HI curve. The fluctuations in the HI curve are from the loss or damage of sensors, sudden changes in the environment, the noise or vibration from other equipment, and so on. These fluctuations indicate the change in system pattern and contain important information about the potential system failure. However, these fluctuations affect the accuracy when developing RUL prediction. Therefore, how to balance the relationship between this potential information and the accuracy of RUL prediction becomes an important problem.

(c) The change point position in the HI curve. According to the bathtub curve, there are two performance transition points: early failure point and degradation beginning point, and these change points are important moments when the state or performance of the mechanical system changes. For

RUL prediction, it is necessary to focus on the latter degeneration beginning point, in other words, RUL prediction is generally carried out in the stage after the degeneration beginning point. Therefore, how to accurately identify the position of the degradation beginning point is very important, which seriously affects the accuracy and computational efficiency of RUL prediction.

To determine the weights of the prognostic metrics, statistical or multi-objective optimization methods have been adopted by many scholars [14-16]. Typically, Wang et al. [17] established a unified framework of weighted normalized square envelope sum to determine the relationships between kurtosis, negative entropy, Gini index, and smoothness index. Li et al. [18] used the self-adjusting analytic hierarchy process (SAAHP) to determine the weight of the monotonicity, correlation, and robustness. Gu et al. [19] found the weights by the convex optimization approach. However, these available methods are subject to human intervention and empirical biases.

For eliminating the fluctuations in the HI curve, the fitting method is a major choice for many scholars [20, 21]. However, this fitting method typically relies on specific mathematical models such as polynomial regression [22], exponential regression [23], and linear regression [24]. In practice, HI data does not fully conform to the assumptions of these models, and inappropriate mathematical models are chosen, resulting in poor-fitting results.

To identify the accurate position of the degradation beginning point, many experts detect the changes in the vibration or sound signals [25, 26]. Li et al. [27] used the principle to determine the degradation beginning point position. Wong et al. [28] proposed a gradient method to identify the degradation beginning point position. However, the method is susceptible to noise interference, leading to inaccurate partitioning.

In addition, for RUL prediction models, the Attention-BiLSTM network is a widely studied model that integrates a BiLSTM network with an attention mechanism [29]. The prediction accuracy is improved due to BiLSTM captures the sequence information comprehensively by processing the data in forward and reverse directions, and the interpretability and effectiveness are improved by the attention mechanism that automatically focuses on important time steps to enhance the focus on key features, but this Attention-BiLSTM model contains a large number of parameters, requiring a large number of experiments for tuning.

To sum up, to address the abovementioned four issues, a bearing RUL prediction method with slope-based change point detection and WOA-Attention-BiLSTM model was proposed. First, a HI construction method with EWM-KECA was proposed to determine the weights of the prognostic metrics and fuse them into one-dimensional HI. Then, a slope-based change point detection was proposed to eliminate the fluctuations in HI and identify the fault change point position in HI. Finally, a WOA-Attention-BiLSTM model was established to develop RUL prediction. This paper aims to develop a comprehensive approach to improve the accuracy, stability, and computational efficiency of RUL prediction by focusing on three aspects: weight optimization, fluctuation handling, and change point localization. To achieve these objectives, a bearing RUL prediction method based on slope-based change point detection and the WOA-Attention-BiLSTM model is proposed. Thus, the contributions are summarized as follows.

(1) The EWM-KECA approach is proposed to construct the HI characterized by the bearing degradation behavior. First, the weights of the prognostic metrics are determined by EWM, then the selected sensitive features are fused into one-dimensional HI retained more information by KECA.

(2) A slope-based change point detection method is proposed to reduce the computational burden and increase the computational efficiency of the model. By locally weighted regression smoothing (LOWESS), the fluctuations in the HI curve are eliminated, and the accurate fault change point position is identified by the maximum second-order difference of the slope between adjacent points. This method serves as a critical step in enhancing the accuracy and reliability of the proposed RUL prediction framework.

(3) A RUL prediction method with a WOA-Attention-BiLSTM model is proposed to improve the prediction accuracy. By adding a WOA [30] optimization algorithm layer into the original Attention-BiLSTM model network structure, the best hyper-parameters are found for establishing the optimal Attention-BiLSTM model, to address issues such as complex parameter selection, high computational cost, strong data dependence, poor model stability, and interpretability.

The remainder of the paper is organized as follows. The details of the proposed bearing RUL prediction method are described in Section 2. Experimental results and discussion are performed in Section 3. A summary of the paper is given in Section 4.

II. The proposed bearing RUL prediction method

Fig. 1 shows the procedure of the proposed bearing RUL prediction method with slope-based change point detection and WOA-Attention-BiLSTM model, including four sequential steps: (1) data acquisition and pre-processing, (2) HI construction with EWM-KECA, (3) slope-based change point detection method, and (4) RUL prediction with WOA-Attention-BiLSTM model.

A: Data acquisition and pre-processing

In this paper, the vibration signal is selected as the monitoring parameter, because the vibration signal contains sufficient operating behavior information and various frequency component information, which is used to monitor bearing states and analyze different types of faults, such as roller faults, inner ring fault or outer ring fault.

The bearing's actual operation is complex and changeable, and the collected bearing vibration signals are inevitably doped with noise, which is similar to useful signals in their form and characteristics. To improve the accuracy and stability of the bearing RUL prediction model, the wavelet threshold de-noising method is employed to de-noise the vibration signal.

Then, 33 serial numbering feature parameters are extracted to construct the feature space in this paper, including 15 time-domain features, 6 frequency-domain features, 8 wavelet packet energy features, and 4 information entropy features. Here, the fifteen extracted time-domain features are mean, variance, root mean square, standard deviation, maximum, minimum, peak, waveform index, kurtosis index, absolute mean, skewness, square root magnitude, margin factor, impulse factor, peak-to-peak value, respectively. The six frequency-domain features are kurtosis frequency, mean frequency, skewness frequency, standard deviation frequency, centroid frequency, and root mean square frequency, respectively. Wavelet packet energy features are a 3-layer decomposition of wavelet packets, thus there are 8 wavelet packet energy features. The four information entropy features are approximate entropy [31], fast sample entropy [32], singular spectrum entropy [33], and spectrum exponential entropy [34], respectively. Furthermore, the above-mentioned 33 serial numbering feature parameters are logarithmic transformations by Eq. (1), then a total of 66 features were obtained in this paper, and the extracted features are named No. 1 to 66 in the above order.

$$F_m = \begin{cases} F_i & m = i \\ \ln(F_i) & m = i + 33 \end{cases} \tag{1}$$

where, $F_m, m = 1, 2, \cdots, 66$ is the extracted features, and the first 33 features are the constructed feature-space, named $F_i, i = 1, 2, \cdots, 33$; while the last 33 features are obtained through logarithmic transformation.

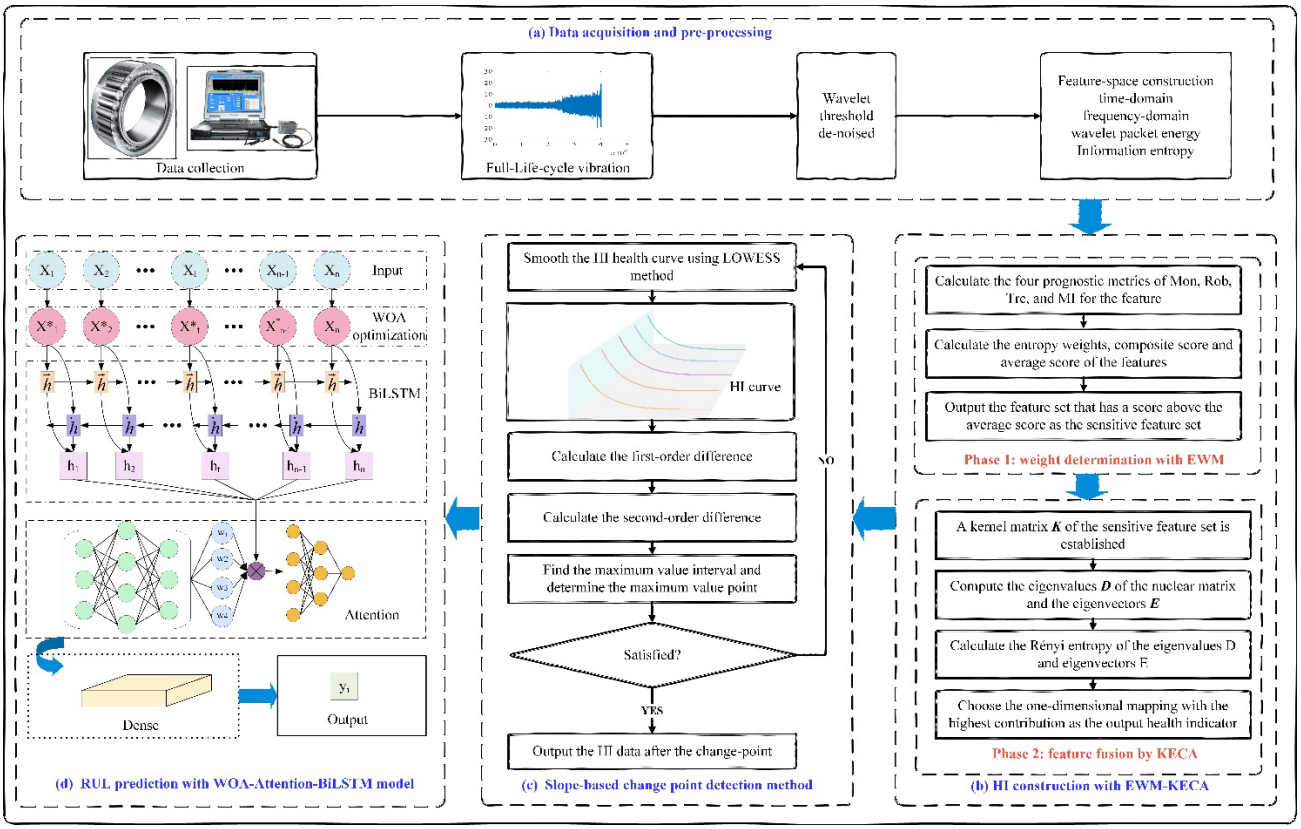


Fig. 1. Procedure of the proposed bearing RUL prediction method

B: HI construction with EWM-KECA

Choosing suitable HI is the premise of accurate RUL prediction, and this paper makes efforts for the effectiveness of HI construction. To address the existing problems in statistical analysis methods such as lack of criteria for bearing parameter selection, insufficient degradation information for bearing feature, and missing of bearing small components, this paper proposes a HI construction method with EWM-KECA, including two phases: (1) Phase 1: weights determination for four prognostic metrics

by EWM, and (2) Phase 2: feature dimension reduction fusion by KECA. The HI construction method, referred to HI construction method with EWM-KECA, is summarized below.

(1) Phase 1: weight determination for four prognostic metrics by EWM

EWM is a comprehensive evaluation method whose calculation process is simple and easy to understand, which is employed to determine the weights of the prognostic metrics with monotonicity (Mon), robustness (Rob), trendability (Tre), and Mutual Information (MI) [35, 36] because avoid the influence of subjective factors by making full use of the collected data. The calculation of EWM is summarized as follows.

HI construction method with EWM-KECA

Input: all the extracted features from the time domain, frequency domain, wavelet packet energy, and information entropy.

Phase 1: weight determination for four prognostic metrics by EWM

a. Calculate a score matrix A.

b. Compute the entropy weight of the prognostic metrics according to Eqs. (3), (4) and (5).

c. Determine the score for each feature with Eq. (6).

d. Obtain the average of the feature composite scores, and compose the sensitive feature set above the average score.

Phase 2: feature fusion by KECA

a. A kernel matrix *K* of the sensitive feature set is established according to Eq. (8).

b. The eigenvalues *D* and eigenvectors *E* of kernel matrix *K* are calculated by nonlinear mapping.

c. Rényi entropy of eigenvalues *D* and eigenvectors *E* is calculated according to Eq. (10).

d. KECA mapping is carried out and the relation between mapping and kernel matrix is established.

e. The one-dimensional map with the highest contribution rate is selected as a HI.

Outputs: one-dimensional health indicator, weights of the prognostic metrics, composite score, the average score

Step 1: Establish the score matrix A.

$$\mathrm{A}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{2}$$

where, m is the number of the extracted features for the vibration signal (m is 66 in this paper), n is the number of the prognostic metrics (n is 4 in this paper), and x_{ij} is the score of the i_{th} feature to the j_{th} prognostic metric, respectively.

Step 2: Calculate the weight P_{ij} of the i_{th} feature in the j_{th} prognostic metric.

$$P_{ij}=\frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \tag{3}$$

Step 3: Find the entropy of each prognostic metric.

$$e_j=-\frac{1}{\ln m}\sum_{i=1}^{m} P_{ij}\ln P_{ij} \tag{4}$$

Step 4: Compute the weight of the i_{th} prognostic metric.

$$\omega_i=\frac{1-e_j}{\sum_{i=1}^{n}(1-e_j)} \tag{5}$$

Step 5: Calculate the score of the i_{th} feature.

$$Score_i=\omega_1\cdot \mathrm{x}_{i1}+\omega_2\cdot x_{i1}+\cdots+\omega_n\cdot x_{in} \tag{6}$$

Step 6: Select the feature that is higher than the average score as the sensitive feature set.

$$F=\left\{Score_i \geq mean(Score_i)\right\} \tag{7}$$

(2) Phase 2: feature fusion by KECA

The selected sensitive features by EWM are high-dimensional data, the HI for RUL prediction is usually one-dimensional data. Thus, KECA is employed to fuse the sensitive features into one-dimensional data. KECA introduces the concept of entropy based on KPCA, and it is a new data transformation method that carries out entropy component analysis in feature space with good

nonlinear processing ability. The feature fusion process of KECA is as follows.

Step 1: Calculate the Euclidean distance of the sensitive features, and establish the kernel matrix $\boldsymbol{K}$ through the Parzen window function.

$$\hat{V}(p)=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int k_\sigma(x,x_i)k_\sigma(x,x_j)\mathrm{d}x=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}k_{\sqrt{2}\sigma}(x_i,x_j)=\frac{1}{N^2}\boldsymbol{I}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{I} \tag{8}$$

where, N is the number of selected features, $\boldsymbol{K}$ is the kernel matrix, $\boldsymbol{I}$ is a $N\times 1$ vector.

Step 2: Compute the eigenvalues and eigenvectors of the kernel matrix $\boldsymbol{K}$ by nonlinear mapping.

$$K=EDE^{\mathrm{T}} \tag{9}$$

where, $D=diag(\lambda_1,\lambda_2,\cdots,\lambda_N)$ and $E=[e_1,e_2,\cdots,e_N]$ are the eigenvalue matrix and eigenvector matrix, respectively.

Step 3: Calculate the Rényi entropy of the eigenvalues and the eigenvectors

$$H(p)=-\log\int p^2(x)\mathrm{d}x \tag{10}$$

where, $p(x)$ is the probability density function.

Step 4: Reorder the eigenvalues and the eigenvectors

$$\xi=\left(\sqrt{\lambda_i}e_i^{\mathrm{T}}\boldsymbol{I}\right)^2 \tag{11}$$

where, ξ is the correlation coefficient of Rényi entropy. Sort the correlation coefficient of Rényi entropy to obtain $\xi_1>\xi_2>\cdots>\xi_N$, and the first p eigenvalues are selected as the main component direction and standardize their eigenvectors.

Step 5: Establish the relation between the mapping and kernel matrix by KECA.

$$\Phi_{\mathrm{eca}}=D^{1/2}E_k^{\mathrm{T}} \tag{12}$$

where, Φ_{eca} is the projection of eigenvalues and eigenvectors., E_k^{T} is a matrix of feature vectors.

Step 6: Output the highest contribution rate mapping as the one-dimensional HI.

C: Slope-based change point detection method

As shown in Fig. 2, in general, the constructed HI has the features of heavy-tail distribution and two stages: a stable operation period and a sharp degradation period. In the first stage, HI is stable and the

degradation features are not obvious. In the second stage, the degradation trend of HI was obvious. Therefore, there is a fault change point between the stable operation period and the sharp degradation period of HI, and the degraded data of the second stage is used in the RUL prediction model to avoid interfering with the accuracy and the computational burden due to non-degraded data. However, so far, the fault change points are generally determined according to the experience of experts or by setting a threshold. These two methods have great subjective consciousness, leading to inaccurate fault change point detection. For this reason, the CUSUM algorithm [37], Cauchy distribution [38], Bayes [39], and genetic algorithm [40] are applied to fault change point detection, but these methods have problems such as a large amount of computation, long time, and difficult threshold determination.

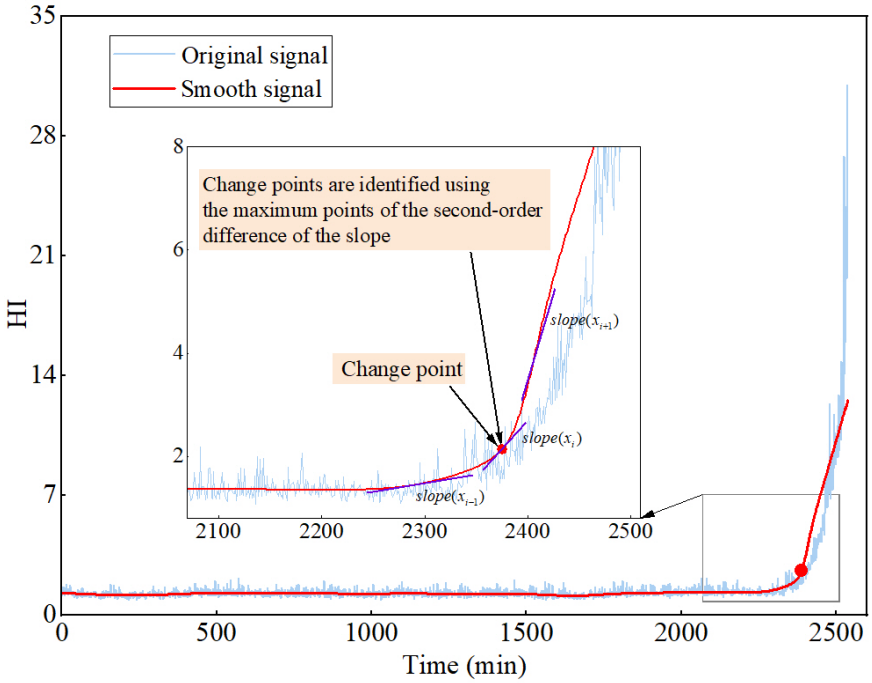


Fig. 2. Slope-based change point detection algorithm

To address the above problems, a slope-based change point detection algorithm is proposed in this paper. When the bearing starts to undergo rapid degradation, the slope of the degradation curve is changed rapidly. Different from other methods, this paper defines the maximum point of local slope increase (decrease) speed change as the failure point. Considering that a curve gradually rises and suddenly accelerates after a certain time, or a curve begins to accelerate and decline, and suddenly

slows down and gradually flattens after a certain point, that is, the point is the slope of the curve. The difference in slope of curves on both sides of a point can reflect the amplitude of slope change on both sides of the point, that is, the first-order difference, but its "turning point" is not the point of maximum slope change, but the point of maximum local slope increase (decrease) speed change. Under the condition of equal spacing, the second-order difference of slope change amplitude reflects the increase (decrease) rate of slope change. By finding the local maximum value of slope change plus (minus) velocity, the position of slope change point can be found.

The slope-based change point detection algorithm, which plays a key role in identifying critical transitions, is summarized in detail, and its step-by-step procedure is outlined as follows.

The slope-based change point detection algorithm

Input: HI

a. Smooth the HI data with locally weighted regression smoothing (LOWESS) according to Eqs. (13), (14) and (15).

b. Select the midpoint of adjacent observation points on the HI curve as the exploration point, and form a sequence of exploration points as the sliding window.

c. Calculate the linear regression coefficient for the data points left or right of each exploration point.

d. Compute the average slope of the left and right sides of the exploration point by the regression coefficient according to Eq. (17).

e. Calculate the first-order difference of the slope with Eq. (18).

f. Calculate the second-order difference of the slope by Eq. (19).

g. Determine the interval with the largest slope change.

h. Determine the position of the fault change point according to Eq. (20)

Outputs: the position of the fault change point

Step 1: Smooth the HI with locally weighted regression smoothing (LOWESS). Assuming that data requiring smoothing processing is determined by x and y, a smoothed value of (x, y) is obtained by performing weighted regression specifying a range of data adjacent to x. Regression coefficients are defined as follows:

$$w_i = \begin{cases} [1-|x-x_i/[d(x)]^3]^3, |x-x_i/[d(x)]| \le 1 \\ 0, |x-x_i/[d(x)] \ge 1 \end{cases} \tag{13}$$

where, $d(x)$ is the furthest distance between x and x_i in the specified span width.

The first-order polynomial fit with the formula $y = ax + b$ is used to estimate the value of $\hat{y}$ at x.

The slope a and constant b are defined as:

$$a = \sum w_i^2 (x-\overline{x})(y-\overline{y}) / [\sum w_i^2 (x-\overline{x})^2] \tag{14}$$

$$b = \overline{y} - a\overline{x} \tag{15}$$

where, $\overline{x}$ and $\overline{y}$ are weighted averages of x and y, respectively.

The robust weight function is defined to calculate the new weight using the residual of the estimated formula. The robust weighting is:

$$\delta_i = \begin{Bmatrix} \left(1-|e_i|^2\right)^2, e_i \le 1 \\ 0, otherwise \end{Bmatrix} \tag{16}$$

where, $e_i = \dfrac{r_i}{6Median(|r_1|,|r_2|,\cdots,|r_n|)}$ and $r_i = y_i - \hat{y}_i$ is the error of each y value.

Repeat the new weighting method in Eqs. (13)-(15) to obtain a smooth value for any point (x, y) after several iterations. In theory, the number of iterations should be less than 4.

Step 2: Select the midpoint of adjacent observation points on the curve as an exploration point to form an exploration point sequence as a sliding window, and make linear regression on n data points left exploration point to obtain regression coefficient, which is recorded as $y_l^{(n)}(x_i)$; Similarly, makes linear regression on n data points right exploration point to obtain regression coefficient, which is recorded as $y_r^{(n)}(x_i)$. When n= 2, 3, 4, Weighted averages of $y_l^{(n)}(x_i)$ and $y_r^{(n)}(x_i)$ are taken, with the weight being n^2, the weighted average calculation formula is obtained as follows:

$$\overline{y}_j(x_i) = \frac{1}{2^2+3^2+4^2}[2^2 \times y_j^{(2)}(x_i) + 3^2 \times y_j^{(3)}(x_i) + 4^2 \times y_j^{(4)}(x_i)] \tag{17}$$

where, 2, 3, and 4 represent different window sizes. $y_j^{(2)}(x_i)$ is the regression coefficient obtained by performing a linear regression on the two data points before and after point x_i, $y_j^{(3)}(x_i)$ is the

regression coefficient obtained by performing a linear regression on the three data points before and after point x_i, and so on.

Step 3: As x_i increases, the curve slope is increased, and the first-order difference is

$$\Delta S(x_i) = \overline{y}_r(x_i) - \overline{y}_l(x_i) \tag{18}$$

Step 4: As x_i increases, the slope changes rapidly at first, then gradually decreases, the second-order difference is

$$\Delta^2 S(x_i) = \Delta S(x_{i-1}) - \Delta S(x_i) \tag{19}$$

Step 5: Along the sequence x_i, the maximum value in the sequence is obtained, and its corresponding interval is (x_{i-1}, x_i), which is the interval where the slope change point is located. Combining the interval (x_{i-2}, x_{i-1}) and (x_i, x_{i+1}) and their corresponding values, the accurate value of the change point position is obtained.

$$x^* = x_{i-1} + \frac{\Delta^2 S(x_i) - \Delta^2 S(x_{i-1})}{\left[\Delta^2 S(x_i) - \Delta^2 S(x_{i-1})\right] + \left[\Delta^2 S(x_i) - \Delta^2 S(x_{i+1})\right]} \times (x_i - x_{i-1}) \tag{20}$$

where, x^* is the precise value of the slope change point.

To evaluate the accuracy of change point detection methods, this paper defines an error formula.

$$\text{Error} = \left(\frac{|\text{ Detected Position} - \text{Actual Position }|}{\text{Total Distance}} \right) \times 100\% \tag{21}$$

where, Detected Position represents the detected change point position, Actual Position represents the actual change point position, and Total Distance represents the total length of the data sequence.

D: RUL prediction with WOA-Attention-BiLSTM model

By inputting the HI after the fault change point into BiLSTM, the computational burden is reduced and the computational efficiency is improved due to eliminating the interference of useless information. Although information is processed in both directions by BiLSTM, there is still an issue of information loss. To alleviate the aforementioned problem with BiLSTM, the addition of an Attention Mechanism

is a common practice. BiLSTM-Attention combines the bidirectional features of BiLSTM and the attention mechanism, enabling important information to be captured more effectively. However, the BiLSTM-Attention model has more parameters and a more complex structure, making the optimization process of the model more challenging, especially when insufficient data leads to potential overfitting. To address these issues, the WOA algorithm is utilized in this paper for optimizing the parameters of the BiLSTM-Attention model.

The structure of the WOA-Attention-BiLSTM model is depicted in Fig. 3, and the network structure parameters are shown in Table 1, including an input layer, two LSTM layers, an attention layer, and a fully connected output layer. In the input layer, vibration time series data is used as input. The first BiLSTM layer learns and encodes the input sequence, where each time step contains information about the input sequence. The second BiLSTM layer further learns and encodes the output sequence of the first BiLSTM layer to generate deeper representations. The attention mechanism is introduced to weigh the output sequence of the second BiLSTM layer by learning the attention weights of each time step, thereby better highlighting the important time steps. By fully connected layers, the weighted time step representation is nonlinear and transformed into a more abstract feature representation. Finally, the final model output is generated through the output layer.

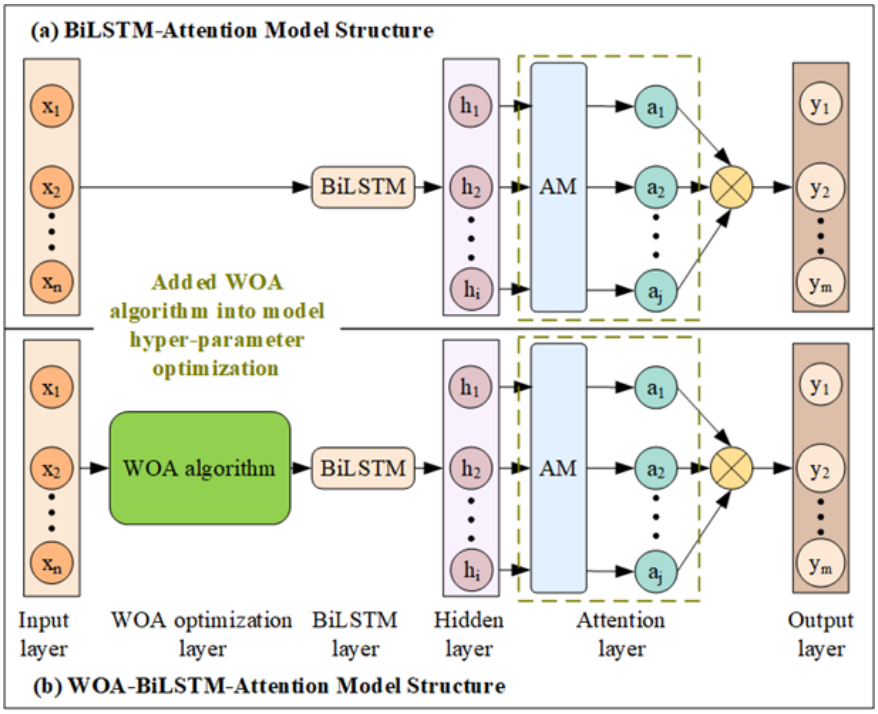


Fig. 3. Proposed WOA-Attention-BiLSTM model structure

Table 1. Network structure parameter

Parameter	Value	Parameter	Value
BiLSTM neurons	512	Optimizer	Adam
Dropout Rate	0.3	Epochs	800
Hidden Layer Neurons	256	Attention Mechanism	64
Hidden Layer Dropout	0.001		

The Attention-BiLSTM model is optimized using the WOA algorithm to reduce redundant calculations, improve the overall computational efficiency of the model, and lower the computational complexity. The specific optimization process is as follows:

Step 1: Initialize the parameters of the whale algorithm W_i, including the maximum number of iterations T, whale population N, optimality seeking dimension, and upper and lower limits of parameters.

Step 2: Initialize the position and size of the whale population, where the learning rate is a decimal and the other parameters are integers.

Step 3: Initialize the global optimal solution, define the initial value of the global optimal solution as infinity, then traverse each whale in the population, calculate its fitness value, and update the position of the global optimal solution and the optimal solution.

$$fitness = \frac{\sum_{i=1}^{n}\left(y_i - \square y_i\right)}{n} \tag{22}$$

where, fitness is the adaptability function for each whale.

Step 4: Perform iterative operations. In each iteration, each whale in the population is traversed, updating its position based on its position and speed. In this process, the exploration strategy and attack strategy in WOA are adopted, and the updated position is calculated by a random number, according to Eqs. (23) and (24), update the whale's position.

$$X^{j+1} = X^* - A \cdot D \tag{23}$$

$$D = \left| C \cdot X^* - X^j \right| \tag{24}$$

where, X^j represents the current position of the whale, X^{j+1} represents the next position of the whale under the influence of prey, which is calculated according to Eq. (27). X^* represents the current optimal position of the whale, i.e. the position of the prey; D represents the distance between the whale and the prey. A and C are obtained as follows.

$$A = 2ar_1 - a \tag{25}$$

$$C = 2r_2 \tag{26}$$

With each iteration, the convergence factor linearly decreases from 2 to 0, in which r_1 and r_2 take values in the range of $[0,1]$.

$$X^{j+1} = \begin{cases} D \cdot e^{bl} \cdot \cos(2\pi l) + X^* & p \geq 0.5 \\ X^{j+1} = X_{\text{rand}} - A \cdot D & p < 0.5 \end{cases} \tag{27}$$

where, the range of p is $[0,1]$, when $p \geq 0.5$, the position is updated in a spiral shape, with b being the logarithmic spiral shape constant, and l taking a value within the range of $[0,1]$, when $p < 0.5$, a random position is chosen for updating, and X_{rand} is a random position of the whale.

Step 5: The updated positions are bounded to ensure that the whale's position is within the search space. Calculate fitness values for updating positions, and update global optimal solutions and positions of optimal solutions by the fitness values

$$fitness(W_i) < fitness(X^*), X^* = W_i \tag{28}$$

Step 6: At the end of each iteration, the fitness of the global optimal solution value is recorded, and the current iteration number, the global optimal solution, and the position of the optimal solution are output. Finally, the global optimal solution, the process of fitness value change, and the position of the optimal solution in each iteration are returned as results.

To better reflect the error between the predicted and actual values for different RUL models, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean

Absolute Percentage Error (MAPE), and coefficient of determinability (R^2) is employed as the evaluation indicators. The calculation formulas are as follows.

$$\mathrm{MAE}=\frac{1}{u}\sum_{i=1}^{u}\left|y_i-f\left(x_i\right)\right| \tag{29}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{u}\sum_{i=1}^{u}\left(y_i-f\left(x_i\right)\right)^2} \tag{30}$$

$$MAPE=\frac{100\%}{u}\sum_{i=1}^{u}\left|\frac{\hat{y}_i-y_i}{y_i}\right| \tag{31}$$

$$R^2=1-\frac{\sum_{i=1}^{u}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{u}(\overline{y_i}-y_i)^2} \tag{32}$$

where, u is the training sample number.

III. Experimental results and discussion

A: Experimental data description

XJTU-SY bearing dataset is used to verify the accuracy and effectiveness of the proposed RUL prediction method [41]. Three working conditions were set up in the test, and five rolling bearings with LDKUER204 were tested under each working condition. Two acceleration sensors (PCB352C33) were used to collect the bearing vibration data. The signal was sampled every 1min at a sampling frequency of 25.6kHz and a sampling time of 1.28s, i.e.32768 vibration signal data points were collected every 1 min. The specific information of the 15 tested bearings is shown in Table 2, including the sample number, rated life, actual life and failure location, and so on.

Table 2. Description of XJTU-SY experimental data

Conditions	Dataset	Samples	Rated life	Actual life	Failure location
Condition 1 (Speed 2100 r/min, Load 12KN)	Bearing1-1	123	5.60~9.68 h	2 h 3 min	Outer ring
	Bearing1-2	161		2 h 41 min	Outer ring
	Bearing1-3	158		2 h 38 min	Outer ring
	Bearing1-4	122		2 h 2 min	Cage
	Bearing1-5	52		52 min	Inner ring, outer ring

Condition 2 (Speed 2250 r/min, Load 11KN)	Bearing2-1	491	6.79~11.73 h	8 h 11 min	Inner ring
	Bearing2-2	161		2 h 41 min	Outer ring
	Bearing2-3	533		8 h 53 min	Cage
	Bearing2-4	42		42 min	Outer ring
	Bearing2-5	339		5 h 39 min	Outer ring
Condition 3 (Speed 2400 r/min, Load 10KN)	Bearing3-1	2538	8.47~14.63 h	42 h 18min	Outer ring
	Bearing3-2	2496		41 h 36min	Inner ring, outer ring, cage, roller
	Bearing3-3	371		6 h11 min	Inner ring
	Bearing3-4	1515		25 h 15min	Inner ring
	Bearing3-5	114		1 h 54 min	Outer ring

B: Bearing RUL prediction

All the samples are divided into the training set and the testing set with a ratio of 7:3.

COMPUTER ENVIRONMENT: All algorithms were performed on the MATLAB 2022 Software platform on a personal computer equipped with a 12th Gen Intel(R) Core (TM) i5-8300 CPU, and 16.0G of RAM.

1) Comparative analysis of HI construction

Fig. 4 shows the results of the HI construction process, which contains the following steps: (a) wavelet de-noised of the raw bearing signal, (b) feature extraction from the de-noised signal, (c) the weight determination of prognostic metrics with EWM, (d) the fitness value of all the features, (e) the selected sensitive features, (f) one-dimensional HI obtained using KECA.

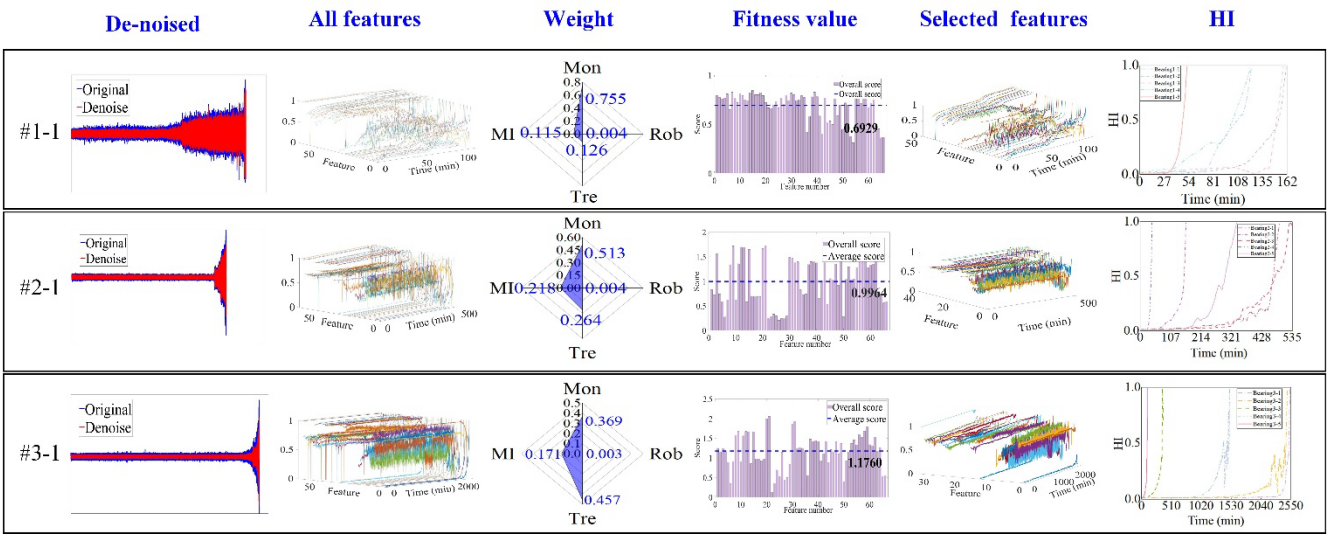


Fig. 4. Results of HI construction process

As seen in Fig.4 (a), the full-life-cycle signal shows the features of trend rise and two stages, and the burr in the raw signal is eliminated and the fluctuation of the signal is reduced by wavelet threshold de-noised method.

To mine the implicit degradation information in the vibration signal, 33 degradation features containing time domain, frequency domain, information entropy, and wavelet energy features were extracted from the de-noised vibration signals by referring to the feature preprocessing method in Section 2.1. To further improve the characterization performance of the extracted degradation features, the 33 degradation features were logarithmic transformations, to obtain 66 degradation features in total. It can be observed from Fig.4 (b), that the 66 extracted features in bearing #1-1, bearing #2-1, or bearing #3-1, are all chaotic, and their regularities are not obvious if such data is directly used for RUL prediction, the accuracy is seriously affected. Therefore, it is necessary to select the sensitive features.

Before the sensitive feature selection, the weight of four prognostic metrics such as Mon, Rob, Tre, and MI is determined by EWM, and the fitness value is computed for the sensitive feature selection. Four prognostic metrics such as Mon, Rob, Tre, and MI are calculated for 66 extracted features, and the available weight determination methods are often determined manually by experts' experience, which is subjective. Therefore, the EWM method is proposed in this paper to determine the weights of the four prognostic metrics.

From Fig.4 (c) and Fig.4 (d), the weights of four prognostic metrics such as Mon, Rob, Tre and MI in bearing #1-1, bearing #2-1, or bearing #3-1 are determined as [0.755, 0.004, 0.126, 0.115], [0.513,

0.004, 0.264, 0.218], and [0.369, 0.003, 0.457, 0.171], respectively. Then, the features that exceed the average fitness value are selected as the sensitive features. Here, the average fitness values for bearing #1-1, bearing #2-1, or bearing #3-1 are 0.6929, 0.9964, and 1.1760, respectively, and the selected sensitive features are shown in Fig.4 (e).

It can be observed that, compared with the raw features in Fig.4 (b), the selected sensitive features in Fig.4 (e) show better degradation features and regularity, and the performance of RUL prediction using such selected sensitive features is better.

From Fig. 4(f), by KECA, the selected sensitive features are constructed into one-dimensional HI, and it can be observed that the constructed HI curves that contain 15 bearings under three working conditions show the features of rising trend and two stages as with the raw signal, indicating that the running behavior of the bearing is well characterized.

To verify the accuracy of the proposed HI construction method, eight methods such as AHP-KECA, FCE-KECA, EWM-PCA, EWM-UMAP, AHP-PCA, AHP-UMAP, FCE-PCA, and FCE-UMAP are constructed, from Analytic Hierarchy Process (AHP), Fuzzy Comprehensive Evaluation (FCE), Principal Component Analysis (PCA) and UMAP, for comparison with the proposed EWM-KECA approach, the comparison results are shown in Table 3 and Fig. 5.

Table 3. Comparison of HI construction methods

Metrics	EWM-KECA	AHP-KECA	FCE-KECA	EWM-PCA	EWM-UMAP	AHP-PCA	AHP-UMAP	FCE-PCA	FCE-UMAP
					#1-1				
Mon	0.230	0.033	0.016	0.016	0.049	0.016	0.016	0.016	0.131
Rob	0.977	0.827	0.929	0.842	0.784	0.842	0.757	0.842	0.624
Tre	0.962	0.512	0.390	0.384	0.750	0.384	0.591	0.384	0.096
MI	4.322	4.062	4.280	4.176	4.158	4.322	4.610	4.112	4.320
					#2-1				
Mon	0.053	0.033	0.020	0.045	0.041	0.045	0.037	0.045	0.037
Rob	0.943	0.935	0.943	0.611	0.739	0.611	0.685	0.611	0.800
Tre	0.697	0.142	0.097	0.477	0.438	0.475	0.224	0.477	0.305
MI	7.759	7.157	6.935	6.587	8.087	6.575	6.627	6.587	6.206

	#3-1								
Mon	0.040	0.007	0.010	0.010	0.004	0.010	0.015	0.010	0.008
Rob	0.926	0.923	0.922	0.736	0.748	0.736	0.719	0.736	0.687
Tre	0.409	0.329	0.337	0.019	0.502	0.019	0.558	0.019	0.323
MI	6.347	6.315	6.309	6.167	6.840	6.168	5.998	6.168	5.895

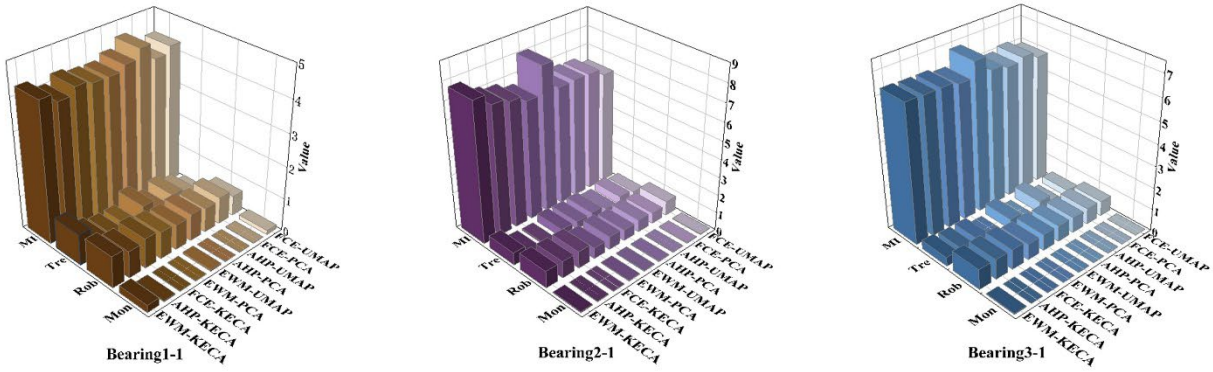


Fig. 5. Comparative results of HI construction methods

From Table 3 and Fig. 5, it can be observed that EWM-KECA has achieved the best performance with the maximum values of the Mon, Rob, and Tre prognostic metrics in bearing #1-1, bearing #2-1, or bearing #3-1. For Mon prognostic metric, the minimum value of bearing #1-1 appears in AHP-PCA, FCE-KECA, AHP-UMAP, and FCE-PCA, and the minimum values of bearing #2-1, or bearing #3-1 appear in FCE-KECA and EWM-UMAP, respectively. For Rob prognostic metric, the minimum values of bearing #1-1 and bearing #3-1 all appeared in FCE-UMAP, and the minimum values of bearing #2-1 appeared in EWM-PCA, AHP-PCA, and FCE-PCA. For the Tre prognostic metric, the minimum values of bearing #1-1 and bearing #2-1 all appeared in FCE-UMAP, and FCE-KECA, respectively, and the minimum values of bearing #3-1 appeared in EWM-PCA. AHP-PCA, and FCE-PCA.

For the MI prognostic metric, the minimum values of bearing #1-1, bearing #2-1, and bearing #3-1 appeared in AHP-KECA, AHP-PCA, and FCE-UMAP, respectively. The maximum value of bearing #1-1 appeared in AHP-UMAP, and the maximum values of bearing #2-1 and bearing #3-1 all appeared in EWM-UMAP. EWM-KECA has achieved the second-maximum value with small errors of -6.25%, -4.06%, and -7.21% than the maximum value in bearing #1-1, bearing #2-1, or bearing #3-1,

respectively. Generally, it is acceptable compared with all other methods.

To sum up, it is concluded that EWM-KECA has obtained the best performance among these HI construction methods in bearing #1-1, bearing #2-1, and bearing #3-1.

2) Comparative analysis of change point detection

To verify the effectiveness of the slope-based change point detection method, three generated data are employed to comparative analysis with six other fault change point detection algorithms such as Cauchy outlier detection algorithm(CODA) [42], CUSUM Maximum Cumulative Sum (CUSUM-MCS) [43], CUSUM Minimum Mean Squared Error (CUSUM-MMSE) [44], Hidden Markov Model (HMM) [45], shape-based multiple segmentation algorithm(SMSA) [46] and time-varying Kalman Kilter (TVKF) [47]. The three generated data contain 600 data points with a time interval of 1s between each consecutive data point. The change point positions of the three generated data are 299th second, 200th second, and 300th second, respectively. The comparative results for seven fault change point detection algorithms are shown in Table 4 and Fig. 6. Here, the errors are computed by Eq. (21), Detected Position represents the detected change point position, Actual Position represents the actual change point position, and Total Distance represents the total length of the data sequence.

Table 4. Comparison of fault change point detection algorithms

Models	Data 1	Error	Data 2	Error	Data 3	Error
Proposed method	299	0	200	0	300	0
CODA	226	0.120	284	0.140	302	0.003
CUSUM-MCS	393	0.158	226	0.120	434	0.223
CUSUM-MMSE	457	0.265	490	0.483	471	0.285
HMM	299	0	231	0.052	—	—
SMSA	299	0	150	0.083	525	0.420
TVKF	299	0	247	0.078	388	0.147

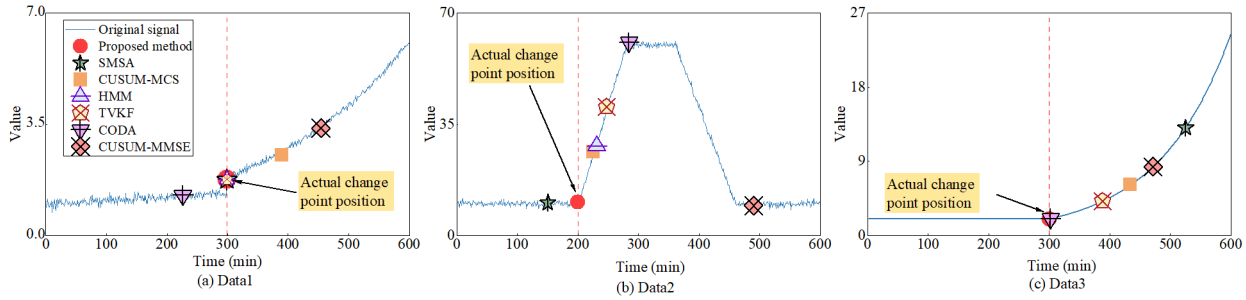


Fig. 6. Comparative results for change point detection methods

As can be seen from Table 4 and Fig. 6, for the three generated data, the proposed slope-based change point detection has all recognized the accurate change point position with zero error. For the first generated data, HMM, SMSA, and TVKF recognize the accurate change point position with zero error, but for the second generated data and the third generated data, HMM, SMSA, and TVKF have achieved worse performance results. Especially, for the third generated data, the largest error appears in SMSA and HMM fails to detect the change point of such data. For the first generated data and the second generated data, the largest error appears in CUSUM-MMSE. Through the three comparative results with these three generated data, it is concluded that the proposed slope-based change point detection algorithm has achieved the best detection accuracy. Therefore, this paper adopts the slope-based change point detection to identify the change point, and the HI data after this identified change point are imported into the RUL prediction model to eliminate the interference of the stable data before this identified change point, to increase the computational efficiency and the accuracy of RUL prediction model.

From Fig. 4(f), the constructed HI shows the features of a heavy tail and two stages. Next, the constructed HI is imported to identify the change point position, and the results are shown in Table 5 and Fig. 7. As seen in Table 5 and Fig. 7, since the HI data before the identified change point position is relatively stable and the degradation is not obvious, these data are generally discarded in the RUL prediction, while the HI data after the identified change point position often shows good degradation features and is retained as the input of the RUL prediction model.

Table 5. Units for Magnetic Properties

Bearing	Position	Bearing	Position	Bearing	Position
#1-1	69	#2-1	447	#3-1	2388
#1-2	111	#2-2	95	#3-2	2031
#1-3	128	#2-3	279	#3-3	233
#1-4	76	#2-4	29	#3-4	933
#1-5	35	#2-5	223	#3-5	81

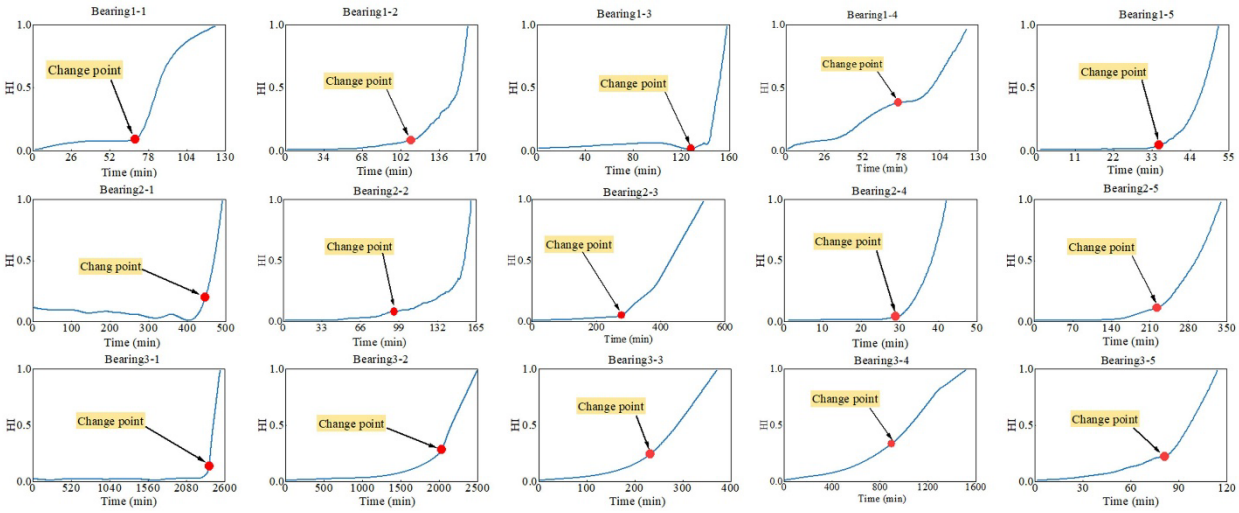


Fig. 7. Results for bearing change point detection

To further verify the effect of the slope-based change point detection, the three minutes before the change point (three minutes preceding) and the three minutes after the change point (three minutes following) were treated as new change points, the full-cycle data and the processed data that only retained the degraded data after the change point position (or the processed data that eliminated the stable data before the change point position) are imported into the LSTM model respectively, and the comparison results are shown in Table 6 and Fig. 8.

Table 6. Comparative results with or without change point detection

Data	**Bearing #1-1**				**Bearing #2-1**				**Bearing #3-1**			
	R^2	**MAE**	**MAPE**	**Time**	**R^2**	**MAE**	**MAPE**	**Time**	**R^2**	**MAE**	**MAPE**	**Time**
Full cycle data	0.944	0.019	0.053	17	0.907	0.025	0.032	23	0.978	0.176	0.117	96
Three	0.932	0.017	0.019	14	0.843	0.039	0.044	21	0.970	0.267	0.052	41

minutes before												
Processed data	0.998	0.005	0.010	13	0.984	0.022	0.028	17	0.998	0.135	0.027	36
Three minutes after	0.963	0.013	0.014	12	0.834	0.042	0.063	16	0.980	0.240	0.047	34

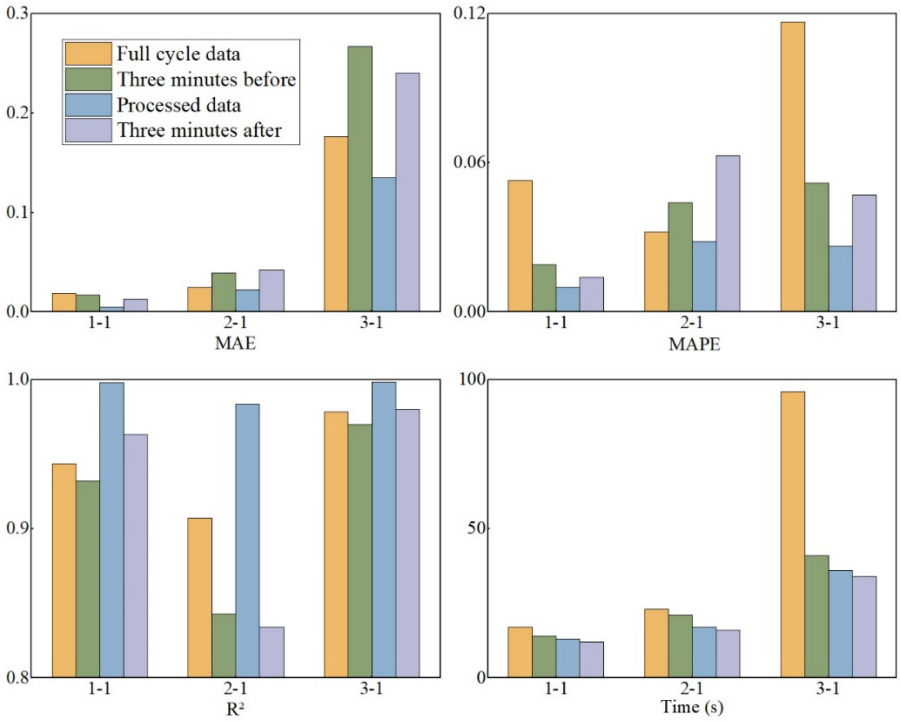


Fig. 8. Comparative results with or without change point detection

As can be seen from Table 6 and Fig. 8, all evaluation indicators such as R^2, MAE, and MAPE for the processed data are better than the other three groups of data, proving the validity and practicability of the slope-based change point detection method. In terms of the time cost, the processed data is more time-saving than the full-cycle data due to eliminating the stable data before the change point position. The larger the amount of data, the more time-saving. Thus, it is concluded that the RUL prediction model embedded with the change point detection algorithm effectively reduces the computational burden and improves the computational efficiency.

3) Comparative analysis of RUL Prediction Models

For simplicity, the following sections will only use bearing1-1 data for case validation. Three different optimization algorithms such as Moth-Flame Optimization (MFO) [48], Red tailed hawk algorithm (RTH) [49], and Wild horse optimization (WHO) [50] are employed to verify the superiority of the WOA optimization algorithm, the comparison results are shown in Table 7 and Fig. 9(a), and the prediction error for four algorithms are shown in Fig. 9(b).

From Table 7 and Fig. 9(a), it can be seen more intuitively that the histograms of WOA algorithm with MAE, MAPE, and RMSE are all the lowest than three other algorithms such as MFO, RTH, and WHO, and From Fig.9 (b), it can be seen that the prediction result of WOA algorithm is the closest to the actual prediction value, indicating WOA optimization algorithm shows the best performance in all evaluation indicators such as MAE, MAPE and RMSE than three other optimization algorithms.

Compared to MFO, the three evaluation indicators with MAE, MAPE, and RMSE of WOA have decreased by 58.24%, 57.44%, and 58.91%, respectively. Compared to RTH, the three evaluation indicators with MAE, MAPE, and RMSE of WOA have decreased by 48.91%, 48.07%, and 48.64%, respectively. Compared to WHO, the three evaluation indicators with MAE, MAPE, and RMSE of WOA have decreased to 76.93%, 76.35%, and 77.31%, respectively.

These results show that the WOA optimization algorithm is significantly better than other optimization algorithms in all evaluation indicators and the prediction error, proving its effectiveness. Thus, WOA is chosen to optimize the Attention-BiLSTM model.

Ablation experiments are performed to verify the accuracy of the WOA-Attention-BiLSTM model, the comparative models are WOA-Attention-LSTM, WOA-LSTM, WOA-BiLSTM, Attention-LSTM, and Attention-BiLSTM, and MAE, MAPE, and RMSE are used to evaluate the prediction performance, the comparative results of ablation experiments are shown in Table 7 and Fig. 9 (c), and the prediction error is shown in Fig. 9 (d).

As can be seen from the data in Table 7 and Fig. 9 (c), it can be seen more intuitively that the histograms of WOA-Attention-BiLSTM with MAE, MAPE, and RMSE are all the lowest than the other models such as WOA-Attention-LSTM, WOA-LSTM, WOA-BiLSTM, Attention-LSTM, and Attention-BiLSTM, and From Fig.9 (d), it can be seen that the prediction result of WOA-Attention-BiLSTM is

the closest to the actual prediction value, indicating WOA-Attention-BiLSTM shows the best performance in all evaluation indicators in ablation experiments.

Compared to WOA-Attention-LSTM, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 38.46%, 38.37%, and 50.77%, respectively. Compared to WOA-LSTM, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 63.64%, 61.76%, and 73.10%, respectively. Compared to WOA-BiLSTM, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 60.00%, 59.48%, and 67.65%, respectively. Compared to Attention-LSTM, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 61.90%, 60.90%, and 70.14%, respectively. Compared to Attention-BiLSTM, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 65.22%, 67.00%, and 68.61%, respectively. Thus, by the results of ablation experiments, the WOA-Attention-BiLSTM model outperformed all other models, demonstrating the superiority of the WOA-Attention-BiLSTM model in significantly enhancing the prediction accuracy and reducing errors.

To verify the superiority of the proposed WOA-Attention-BiLSTM model, the STOA RUL prediction methods such as WOA-BP, WOA-ELM, WOA-GRU, WOA-BiTCN, CNN-Attention-BiGRU, WOA-CNN-BIGRU, and WOA-Attention-CNN are chosen for the comparative test, the results are shown in Table 7 and Fig. 9 (e), and the prediction error is shown in Fig. 9 (f).

As can be seen from the data in Table 7 and Fig. 9 (e), it can be seen more intuitively that the histograms of WOA-Attention-BiLSTM with MAE, MAPE, and RMSE are all the lowest than the STOA models such as WOA-BP, WOA-ELM, WOA-GRU, WOA-BiTCN, CNN-Attention-BiGRU, WOA-CNN-BIGRU and WOA-Attention-CNN and From Fig.9 (f), it can be seen that the prediction result of WOA-Attention-BiLSTM is the closest to the actual prediction value, indicating WOA-Attention-BiLSTM shows the best performance in all evaluation indicators such as MAE, MAPE and RMSE than the STOA models.

Compared to WOA-BP, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 93.30%, 93.43%, and 92.35%, respectively. Compared to WOA-ELM, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 84.75%, 83.62%, and 86.82%, respectively. Although the performance of WOA-ELM is a little better than that of WOA-BP, there is still a significant gap

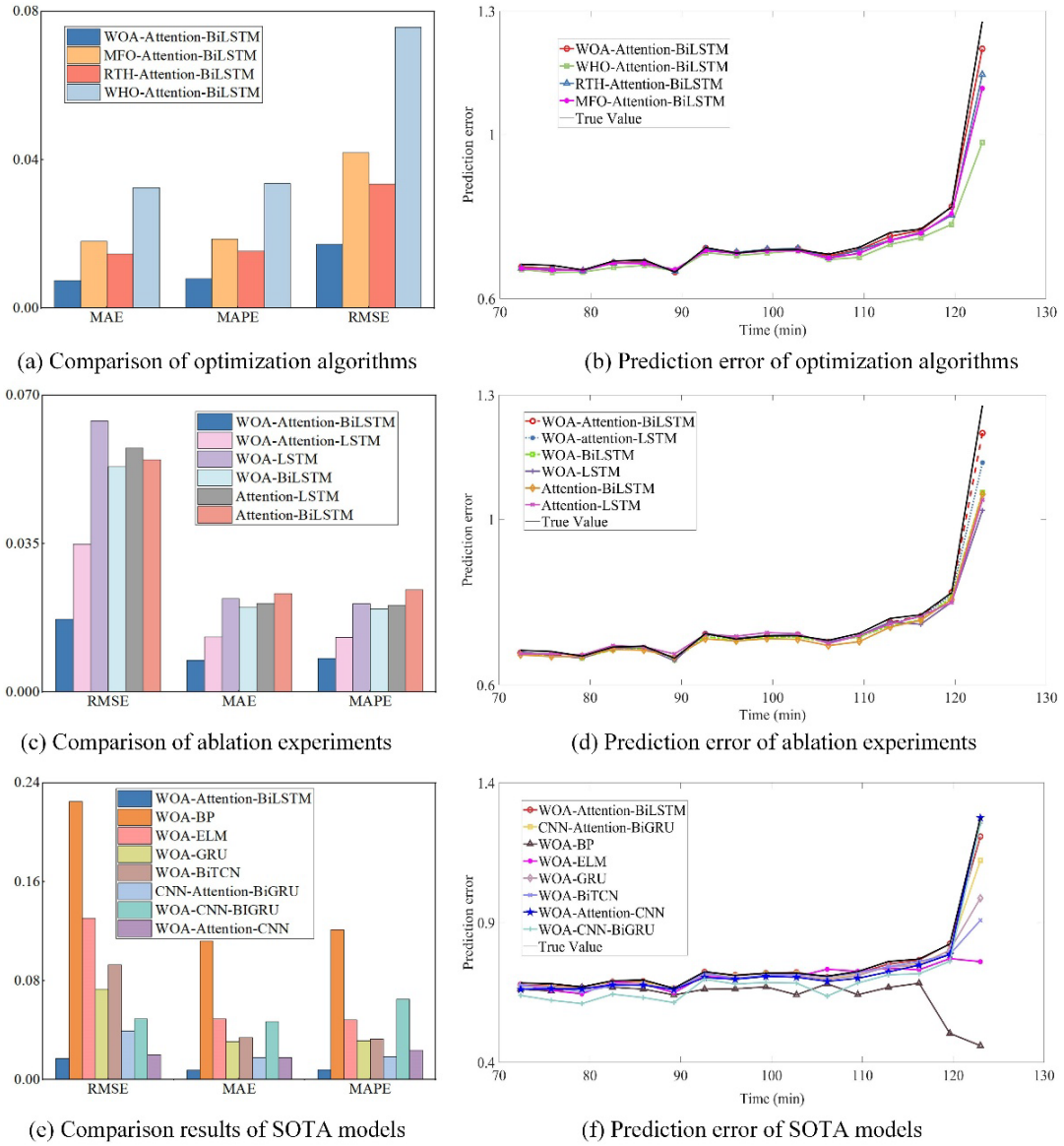
compared with WOA-Attention-BiLSTM. Compared to WOA-GRU, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 75.45%, 74.54%, and 76.50%, respectively. The WOA-GRU method performs better, but it still falls short of the performance level of WOA-Attention-BiLSTM. Compared to WOA-BiTCN, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 77.84%, 75.91%, and 81.46%, respectively. This indicates that the WOA-BiTCN method has a larger error and lower prediction accuracy. Compared to CNN-Attention-BiGRU, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 57.59%, 57.23%, and 56.45%, respectively. Although the performance of CNN-Attention-BiGRU is already quite good, WOA-Attention-BiLSTM still has a higher accuracy. Compared to WOA-CNN-BiGRU, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 82.98%, 87.69%, and 65.31%, respectively. Although WOA-CNN-BiGRU performs well, there is still a significant improvement in the prediction accuracy with WOA-Attention-BiLSTM. Compared to WOA-Attention-CNN, the MAE, MAPE, and RMSE of WOA-Attention-BiLSTM have reduced by 55.56%, 66.67%, and 60%, respectively. While WOA-Attention-CNN achieves a competitive performance, WOA-Attention-BiLSTM still demonstrates superior predictive accuracy.

In summary, the WOA-Attention-BiLSTM model significantly outperforms the SOTA RUL prediction models and has achieved the highest prediction accuracy with the minimum values of MAE, MAPE, and RMSE, demonstrating the effectiveness and superiority of the WOA-Attention-BiLSTM model in reducing prediction errors and improving model performance.

Table 7. Comparative analysis of the optimization algorithms, ablation experiments and SOTA RUL prediction models

	Models	MAE	Relative error	MAPE	Relative error	RMSE	Relative error
Optimization algorithms	WOA-Attention-BiLSTM	0.008	0	0.008	0	0.017	0
	MFO-Attention-BiLSTM	0.018	58.24%	0.019	57.44%	0.042	58.91%
	RTH-Attention-BiLSTM	0.015	48.91%	0.015	48.07%	0.033	48.64%
	WHO-Attention-BiLSTM	0.033	76.93%	0.034	76.35%	0.076	77.31%
Ablation experiments	WOA-Attention-BiLSTM	0.008	0	0.008	0	0.0172	0
	WOA-Attention-LSTM	0.013	38.46%	0.013	38.37%	0.0349	50.77%

	WOA-LSTM	0.022	63.64%	0.021	61.76%	0.0639	73.10%
	WOA-BiLSTM	0.020	60.00%	0.020	59.48%	0.0531	67.65%
	Attention-LSTM	0.021	61.90%	0.020	60.90%	0.0576	70.14%
	Attention-BiLSTM	0.023	65.22%	0.024	67.00%	0.0548	68.61%
SOTA RUL prediction models	WOA-Attention-BiLSTM	0.008	0	0.008	0	0.017	0
	WOA-BP	0.112	93.30%	0.121	93.43%	0.225	92.35%
	WOA-ELM	0.049	84.75%	0.049	83.62%	0.130	86.82%
	WOA-GRU	0.031	75.45%	0.031	74.54%	0.073	76.50%
	WOA-BiTCN	0.034	77.84%	0.033	75.91%	0.093	81.46%
	CNN-Attention-BiGRU	0.018	57.59%	0.019	57.23%	0.039	56.45%
	WOA-CNN-BIGRU	0.047	82.98%	0.065	87.69%	0.049	65.31%
	WOA-Attention-CNN	0.018	55.56%	0.024	66.67%	0.020	60%



(a) Comparison of optimization algorithms

(b) Prediction error of optimization algorithms

(c) Comparison of ablation experiments

(d) Prediction error of ablation experiments

(e) Comparison results of SOTA models

(f) Prediction error of SOTA models

Fig. 9. RUL prediction models: the comparative results, the prediction error

IV. Conclusion

This paper focused on the problem of HI construction, fault change point detection, and parameter optimization in the bearing RUL prediction model, and efforts were made to establish a novel bearing RUL prediction method. The effectiveness of the proposed RUL prediction method was verified by the XJTU-SY bearing dataset. The main conclusions are as follows.

(1) Through the comparative results of HI construction, change point detection, and the optimization algorithms, it is found that the proposed HI construction with EWM-KECA, the slope-based change point detection technique, and the WOA optimization algorithm have all achieved the best performance, avoiding the unilaterality and subjectivity and better characterizing the bearing degradation behavior, to improve the computing efficiency and reduce the computing burden for RUL prediction.

(2) By the comparative results of the ablation experiments and the SOTA prediction models, it is observed that the histograms of the proposed WOA-Attention-BiLSTM model are all the lowest with the smallest prediction error than other models, verifying the superiority of the proposed WOA-Attention-BiLSTM model.

(3) It is foreseeable that in the future era of big data, the RUL prediction model embedded with a change point detection algorithm will be more and more attention. Due to the features of the heavy-tail distribution of the bearing full-life-cycle data, the stable data without degradation trend before the change point position accounting for more than 90% will be eliminated for RUL prediction, thus significantly and efficiently improving the computing power of the model.

References

[1] Lei YG, Li NP, Guo L, Li NB, Yan T, Lin J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 2018, 104: 799-834.

[2] Gao H, Li YB, Zhao Y, Song Y. Dual Channel Feature-Attention-based Approach for RUL Prediction Considering the Spatiotemporal Difference of Multisensor Data. IEEE Sensors Journal, 2023, 23(8): 1.

[3] Duan Y, Cao XG, Zhao JB, Li M, Yang X. A Spatio-temporal Fusion Autoencoder-based Health Indicator Automatic Construction Method for Rotating Machinery Considering Vibration Signal Expression. IEEE Sensors Journal, 2023, 23(20): 1.

[4] Zeng M, Wu F, Cheng YW. Remaining Useful Life Prediction via Spatio-Temporal Channels and Transformer. IEEE Sensors Journal, 2023, 23(23): 29176-29185.

[5] Zhuang JC, Jia MP, Huang CG, Michael Beer, Feng K. Health prognosis of bearings based on transferable autoregressive recurrent adaptation with few-shot learning. Mechanical Systems and Signal Processing, 2024, 211: 111186.

[6] Zhuang JC, Cao YD, Ding YF, Jia MP, Feng K. An autoregressive model-based degradation trend prognosis considering health indicators with multiscale attention information. Engineering Applications of Artificial Intelligence, 2024, 131: 107868.

[7] Li Y, Wang HJ, Li JW, Tan J. A 2-D Long Short-Term Memory Fusion Networks for Bearing Remaining Useful Life Prediction. IEEE Sensors Journal, 2022, 22(22): 21806-21815.

[8] Huang P, Gu Y, Qiu G. A novel feature dimensionality reduction method for gearbox fault diagnosis with HMSDE, DANCo-DDMA and KELM. Nonlinear Dynamics, 2024, 112: 14071-14091.

[9] Yi CC, Li SH, Huang T, Xiao H, Jiang YF. On a Prediction Method for Remaining Useful Life of Rolling Bearings via VMD-Based Dispersion Entropy and GAN. IEEE Sensors Journal, 2023, 23(22): 1.

[10] Moradi M, Broer A, Chiachío J, Benedictus R, Loutas TH, Zarouchas D. Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data. Engineering Applications of Artificial Intelligence, 2023, 117: 105502.

[11] Lei YG, Li NP, Gontarz S, Lin J, Radkowski S, Dybala J. A model-based method for remaining useful life prediction of machinery. IEEE Transactions on Reliability, 2016, 65(3): 1314-1326.

[12] Suh SH, Lukowicz P, Lee Y O. Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks. Knowledge-Based Systems, 2022, 237: 107866.

[13] Xu WY, Jiang QS, Shen YH, Xu FY, Zhu QX. RUL prediction for rolling bearings based on Convolutional Autoencoder and status degradation model. Applied Soft Computing, 2022 130:

109686.

[14] Behera S, Misra R. A multi-model data-fusion based deep transfer learning for improved remaining useful life estimation for IIOT based systems. Engineering Applications of Artificial Intelligence, 2023, 119: 105712.

[15] Qin Y, Yang JH, Zhou JH, Pu HY, Mao YF. A new supervised multi-head self-attention autoencoder for health indicator construction and similarity-based machinery RUL prediction. Advanced Engineering Informatics, 2023, 56: 101973.

[16] Xu F, Wang L. Constructing a health indicator for bearing degradation assessment via an unsupervised and enhanced stacked autoencoder. Advanced Engineering Informatics, 2022, 53: 101708.

[17] Wang D, Peng Z, Xi L, The sum of weighted normalized square envelope: A unified framework for kurtosis, negative entropy, Gini index and smoothness index for machine health monitoring. Mechanical Systems and Signal Processing, 2020, 140: 106725.

[18] Li YX, Liu J, Liu BN, Zhang FY, Yuan XH, Zhang YC. Fusing sensitive degradation features with uncertainty analysis for RUL prediction of rotating machines. Measurement Science and Technology, 2024, 35(3): 035123.

[19] Gu X, Yu YX, Hou KX, Guo L, Gao HL. A new method for rolling bearing condition evaluation method under varying speeds. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1.

[20] Zhu JJ, Jiang QS, Shen YH, Xu FY. Res-HSA: Residual hybrid network with self-attention mechanism for RUL prediction of rotating machinery. Engineering Applications of Artificial Intelligence, 2023, 124: 106491.

[21] Francesco DA, Federico ND. Polynomial approximation of derivatives through a regression–interpolation method. Applied Mathematics Letters, 2024, 152: 109010.

[22] Chen JT. Estimating a regression function in exponential families by model selection. Bernoulli, 2024, 30(2): 1669-1693.

[23] Wang YA; Huang Q, Yao ZG, Zhang Y. On a class of linear regression methods. Journal of Complexity, 2024, 82: 101826.

[24] Zhang C, Zhang YB, Hu CX, Liu ZB, Cheng LY, Zhou Y. A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network. IEEE Access, 2020, 8: 36293-36312.

[25] Xiang S, Qin Y, Luo J, Pu HY. Spatiotemporally multidifferential processing deep neural network and its application to equipment remaining useful life prediction. IEEE Transactions on Industrial Informatics, 2022, 18(10): 7230-7239.

[26] Wen L, Yang G, Hu LX, Yang CS, Feng K. A new unsupervised health index estimation method for bearings early fault detection based on Gaussian mixture model. Engineering Applications of Artificial Intelligence, 2024, 128: 107562.

[27] Li X, Zhang W, Ding Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction (Article). Reliability Engineering and System Safety, 2019, 182: 208-218.

[28] Wong, JT, Bertozzi AL. A hybrid prognostics technique for rolling element bearings using adaptive predictive models. IEEE Transactions on Industrial Electronics, 2018, 65(2): 1577-1584.

[29] Liu J, Yang Z, Xie JS, Wang RJ, Liu SH, Xi DR. A feature fusion-based method for remaining useful life prediction of rolling bearings. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1.

[30] Mirjalili, Seyedali, Lewis. The whale optimization algorithm. Advances in engineering software, 2016.01.008.

[31] Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America, 1991, 88(6): 2297-2301.

[32] Liu WF, Jiang Y, Xu YS. A super fast algorithm for estimating sample entropy. Entropy, 2022, 24(4): 524.

[33] Hassani H, Thomakos D. A review on singular spectrum analysis for economic and financial time series. Statistics and its interface, 2010, 3(3).

[34] Savvidy K, Savvidy G. Spectrum and entropy of C-systems MIXMAX random number generator. Chaos Solitons & Fractals, 2016, 91(1): 33-38.

[35] Sui WT, Zhang D, Qiu XM, Zhang W, Yuan L. Prediction of bearing remaining useful life based

on mutual information and support vector regression model. IOP Conference Series: Materials Science and Engineering, 2019, 533: 012032.

[36] Gu YK, Bi QP, Qiu GQ. Practical health indicator construction methodology for bearing ensemble remaining useful life prediction with ISOMAP-DE and ELM-WPHM. Measurement Science and Technology, 2022, 33(2): 025007.

[37] Montes DOV, Jeske DR, Zhang Q, Rendon C, Marvasti M. A cusum change-point detection algorithm for non-stationary sequences with application to data network surveillance. Journal of Systems and Software, 2010, .83(7): 1288-1297.

[38] Ebner B, Eid L, Klar B. Cauchy or not Cauchy? New goodness-of-fit tests for the Cauchy distribution. Statistical Papers, 2024, 65(1): 45-78.

[39] Chai J, Lu QY, Hu Y, Wang SY, Lai KK, Liu HT. Analysis and Bayes statistical probability inference of crude oil price change point. Technological Forecasting and Social Change, 2018, 126: 271-283.

[40] Whitley D. A genetic algorithm tutorial. Statistics and Computing, 1994, 4(2): 65-85.

[41] Wang B, Lei YG, Li NP, Yan T. Deep separable convolutional network for remaining useful life prediction of machinery. Mechanical Systems and Signal Processing, 2019, 134: 106330.

[42] Wei DL, Li F, Weng SY. Cauchy noise removal via convergent plug-and-play framework with outliers detection. Journal of Scientific Computing, 2023, 96(3): 76.

[43] Lee S, Lee S. Exponential family QMLE-based CUSUM test for integer-valued time series. Communications in Statistics - Simulation and Computation, 2023, 52(5): 2022-2043.

[44] Marion R, Reynolds Jr, Park C. CUSUM charts for detecting special causes in integrated process control. Quality and Reliability Engineering International, 2010, 26(3): 199-221.

[45] Erivwo O, Makis V, Kwon, Roy. Bayesian change point prediction for downhole drilling pressures with hidden Markov models. Applied Stochastic Models in Business and Industry, 2023, 1.

[46] Zhuang D, Yan QJ, Liu SZ, Ma TF, Liu YB. A shape-based multiple segmentation algorithm for change-point detection. Computers & Industrial Engineering, 2023, 176.

[47] Abewickrema W, Yildirimoglu M, Kim J. Multivariate time-varying Kalman filter approach for cycle-based maximum queue length estimation. Transportation Research Part C: Emerging

Technologies, 2023, 154: 104238.

[48] Mirjalili S, Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. Knowledge-Based Systems, 2015, 89: 228-249.

[49] Ferahtia S, Houari A, Rezk H, Djerioui A, Machmoum M, Motahhir S, Ait-Ahmed M. Red-tailed hawk algorithm for numerical optimization and real-world problems. Scientific reports, 2023, 13(1): 12950.

[50] Zhou S, Shi Y, Wang DJ, Xu XZ, Xu MM, Deng Y. Election Optimizer Algorithm: A new meta-heuristic optimization algorithm for solving industrial engineering design problems. Mathematics, 2024, 12(10): 1513.

Guangqi Qiu is an Associate Professor of the School of Mechanical and Electrical Engineering, at Jiangxi University of Science and Technology. He received his Ph.D in Chemical Process Machinery from South China University of Technology in 2017. His research interests include fault diagnosis, and RUL prediction. He has published more than 30 journal papers.

Binlu Ye is graduated from Jiangxi University of Science and Technology in 2022 and is currently pursuing a master's degree in mechanical engineering at Jiangxi University of Science and Technology. Her current research direction is remaining useful life prediction of rolling bearings.

Yingkui Gu is a Professor of the School of Mechanical and Electrical Engineering, at Jiangxi University of Science and Technology. He received his Ph.D in Mechanical Engineering from Dalian University of Technology in 2005. His research interests include fault diagnosis, reliability engineering, and optimization design. He has published more than 100 journal papers and 3 books.

Peng Huang is a lecturer of the School of Mechanical and Electrical Engineering, at Jiangxi University of Science and Technology. He received his Ph.D degree from University of Electronic Science and Technology of China in 2021. His research interests include reliability analysis, reliability-based optimization design, and risk evaluation.

He Li works at the School of Engineering, Liverpool John Moores University, Liverpool. Before this, he received

his B.S. and Ph.D. degrees in Center for System Reliability and Safety at the University of Electronic Science and Technology of China, China. His research mainly focuses on the failure, risk, reliability, availability, and maintainability of complex systems such as floating offshore wind turbines.

Zifei Xu is a Senior Research Fellow in the Department of Civil and Environmental Engineering at the University of Liverpool. Awarded the prestigious Marie Curie Fellowship in 2022, he specializes in diagnostics, prognostics, and health management for complex structures. His research aims to enhance the reliability and safety of engineering systems through advanced methodologies in structural health monitoring.