A novel method of health indicator construction and remaining useful life prediction based on deep learning

Indexed by:
Web of Science Group

Xianbiao Zhan[a,c], Zixuan Liu[b], Hao Yan[a], Zhenghao Wu[a], Chiming Guo[a], Xisheng Jia[a,*]

[a] Army Engineering University of PLA, China
[b] North China Institute of Aerospace Engineering, China
[c] Hebei Key Laboratory of Condition Monitoring and Assessment of Mechanical Equipment, China

Highlights

- Utilize technologies such as SK and VMD to extract multi-domain features as feature sets.
- Utilize monotonicity, trendiness, and robustness to select features for fusion and construct HI.
- Combining SSAE with LSTM for condition assessment and residual life prediction.

Abstract

The construction of health indicators (HI) for traditional deep learning requires human training labels and poor interpretability. This paper proposes an HI construction method based on Stacked Sparse Autoencoder (SSAE) and combines SSAE with Long short-term memory (LSTM) network to predict the remaining useful life (RUL). Extracting features from a single domain may result in insufficient feature extraction and cannot comprehensively reflect the degradation status information of mechanical equipment. In order to solve the problem, this article extracts features from time domain, frequency domain, and time-frequency domain to construct a comprehensive original feature set. Based on monotonicity, trendiness, and robustness, the most sensitive features from the original feature set are selected and put into the SSAE network to construct HI for state partitioning, and then LSTM is used for RUL prediction. By comparing with the existing methods, it is proved that the prediction effect of the proposed method in this paper is satisfied.

Keywords

stacked sparse autoencoders；health indicators; long short-term memory networks; remaining useful life prediction

1. Introduction

Planetary gearboxes have the advantages of small size, high load-bearing capacity and high transmission efficiency, and have been widely used in mechanical equipment such as helicopters, high-speed trains and heavy vehicles [5]. Due to the complex structure of planetary gearboxes is difficult to observe the health state directly, so the degradation state of the equipment needs to be modelled with the help of health indicators (HI) [6]. An appropriate and effective HI can directly and accurately reflect the degradation state of mechanical equipment and is helpful to the subsequent remaining useful life (RUL) prediction. Currently, the mainstream methods of HI construction mainly include: statistical parameter-based construction methods, multidimensional statistical feature fusion-based construction methods and deep learning-based construction methods [36].

The method of construction based on statistical parameters is to subject the monitored signal to statistical characterization to obtain HI(such as: root mean square (RMS), energy, peak value, etc). Malhi et al. [22] extracted RMS and peak values from the collected vibration signals by continuous wavelet

(*) Corresponding author.
E-mail addresses: X. Zhan (ORCID:0000-0001-6775-9589) zxblzx999@163.com, Z. Liu liuzixuan122395@163.com, H. Yan yhfy_1997@163.com, Z. Wu sttwhao@163.com, C. Guo guochiming@nudt.edu.cn, X. Jia (ORCID: 0009-0009-6766-6601) xs_jia2022@163.com,

transform as HI to predict the RUL of the bearings. Chen et al. [7] used the relative root mean square (RRMS) as HI to describe the degradation characteristics of the bearing. Igba et al. [16] extracted the RMS and peak values of the raw vibration signals for fault detection of wind turbines and concluded that the RMS and peak values could better describe the health status of the equipment. Direct extraction of statistical parameter features is susceptible to interference from other signals and it is difficult to precisely characterize the degradation rules of the equipment. Therefore, the researcher performs signal processing on the raw vibration signal. Related research shows that the statistical features extracted by signal processing can extract weak fault features in the pre-fault period, which helps to find faults earlier and can better describe the degree of degradation of the equipment [26,27,30,4]. Theoretically, a single statistical feature may have some monotonicity and correlation, and it can reflect the degree of degradation of a fault and is simple to calculate. However, the fault information contained in a single statistical parameter may be less complete and does not provide a more comprehensive reflection of the degraded condition of the equipment. And each feature does not contribute equally to the HI construction [24], so the health of the equipment cannot be accurately depicted.

In order to obtain more comprehensive fault information, some scholars have proposed a method based on the fusion construction of multidimensional features. Among them, dimensionality reduction using principal components analysis (PCA) to extract the first principal component as HI, which has received much attention from scholars. Widodo et al. [31] extracted multidimensional features from time-domain signals and used the PCA algorithm to construct HI, which were fed into a relevance vector machine (RVM) for training and prediction. The results obtained outperformed the single-feature prediction results. Liu et al. [18] used PCA to extract features from the original monitoring data so as to reduce the dimension and construct HI. Shi et al. [28] addressed the problem that PCA cannot effectively extract weak faults in time-varying signals. Deep recursive dynamic principal component analysis (RDPCA) was proposed for dimensionality reduction to construct HI. Subsequent scholars have improved on PCA and proposed a variety of new algorithms for feature fusion with reduced dimensionality, achieving good results in both diagnosis and

prediction [1,8]. The existing research literature on multidimensional feature fusion for constructing HI has focused mainly on the study of fusion methods. Although such fusion methods are simple and easy to implement, they are likely to fuse features that are not applicable, which in turn affects the accuracy of health assessment and RUL prediction.

With the rise of neural networks in recent years, many academics have used deep learning to construct HI. Luo et al. [19] used convolutional neural networks (CNN) for feature extraction. The features were then input into a Bi-directional long short-term memory (Bi-LSTM) network with an attention mechanism to construct HI and perform RUL prediction. Yoo et al. [33] converted the vibration signal into an image, which is input to CNN to construct HI, it was experimentally concluded that the image-based RUL prediction has higher accuracy and outperforms the conventional algorithm. Yu et al. [34] proposed a bidirectional recurrent neural network encoder decoder framework to build HI, and the results show that the proposed method has better performance. The use of deep learning to construct HI is an efficient way to assess and predict without the need for human or expert experience. However, the HI constructed by this method has poor interpretability and poor resistance to interference.

It can be seen from the above research literature that the current structure of HI has a good degradation trend and high prediction accuracy. But the following questions still remain:

(1) In the HI modelling process, the above literature does not take into account the monotonicity, correlation and robustness of HI. Therefore, it fails to accurately describe the degradation state of the mechanical equipment and obtain a more accurate model of health degradation.

(2) The original vibration signal contains complex information, which can not directly reflect the working state of mechanical equipment. However, when extracting features from a single domain, it will lead to insufficient feature extraction, which can not reflect the degradation state information of mechanical equipment more comprehensively. Thus, the accuracy of assessment and prediction is reduced.

To address these issues, this paper combines signal processing techniques to extract multi-domain features for fusion. In the process of data fusion, data fusion methods based on deep learning are widely used. In particular, deep learning

algorithms such as autoencoder (AE) have been recognized by experts in fusion and construction of HI due to their clear structure and strong feature extraction ability. Qin et al. [25] extracted multi-dimensional time-domain and frequency-domain features, and then used the degradation trend constrained Variational autoencoder (DTC-VAE) to construct the HI. Zhou et al. [37] designed a distribution contact ratio metric (DCRM) to calculate the distribution distance of the original vibration signal of the Gaussian Mixture model as HI, and made RUL prediction, which achieved good prediction results. She et al. [29] used AE improved algorithm Sparse Autoencoder (SAE) to build HI. The constructed HI is used for RUL, and good results are obtained. SAE model adds a penalty factor to the hidden layer of the original AE model to achieve the sparsity limit [15]. The advantage of the SAE model over the AE model is that the non-linear features can be fully learned. It also reduces the input dimensionality making the calculation easy and simple. In view of this, in this paper, the SAE network model is stacked in order to more fully learn the non-linear features, constituting a Stacked Sparse Autoencoders (SSAE). The SSAE model retains the advantages of the SAE model and also more fully extracts the non-linear features.

In terms of remaining life prediction, the mainstream RUL prediction methods currently available include: physical model-based RUL prediction and data-driven RUL prediction-based methods. The RUL prediction method based on physical models can obtain more accurate prediction results [38,32]. However, for complex systems, it is more difficult to build mathematical models and requires a great deal of expert knowledge. As a result, an increasing number of scholars are adopting a data-driven RUL prediction-based approach [9]. Fei et al. [14] achieved better prediction results using a Support Vector Machine (SVM). Cheng et al. [10] used the Hilbert-Huang transform to construct the HI of the vibration signal. Then used a deep CNN to predict the RUL of the bearing and achieved a high prediction accuracy. Lin et al. [20] first used the Fuzzy C-means algorithm (FCM) for state partitioning of the whole-life data. Then the HI was constructed using the principal component analysis algorithm. Finally, the RUL was performed using RNN and validated on two different datasets and better prediction results were obtained. Although the above literature has achieved good prediction performance. However, the

prediction results for the long time series aspect are still poor. Therefore, researchers have improved RNN and used LSTM algorithm for RUL prediction. Zhou et al. [39] proposed an entropy based sparsity metric criterion to address the sensitivity issue of constructing HI for mechanical equipment defect monitoring. And on this basis, LSTM is used for regularization prediction. The results are verified on several data sets and good prediction results are obtained. Zhou et al. [40] designed an LSTM based RUL prediction method and validated it on three different types of datasets, with better prediction results than other methods. LSTM can better solve the task in terms of long-time sequences and alleviates the problems of gradient disappearance and explosion that exist in RNN [21].

This paper first uses signal processing techniques to extract multi-domain features. Then superior degradation features are selected by using monotonicity, correlation and robustness, and finally fed into a SSAE network to fuse and construct HI. Based on the HI curves, The period of severe degradation are determined and RUL is performed by using LSTM. The main contributions of this paper are as follows.

(1) In order to more comprehensively depict the degradation status information of equipment, this paper extracts time-domain, frequency-domain features by spectral kurtosis (SK) transformation, and multi-scale fuzzy entropy features by variational mode decomposition (VMD) transformation. Eliminating the drawbacks of traditional features allows for better extraction of fault sensitive features.

(2) In order to be able to construct better HI, this paper selects excellent features with the help of monotonicity, correlation and robustness, and inputs them into the SSAE network model for fusion, getting rid of the shortcomings of the existing mainstream construction of HI.

(3) In order to construct a suitable HI curve, SSAE, which has a clear structure and powerful capability of non-linear feature learning, was chosen for fusion construction of HI and combined with LSTM for RUL. the effectiveness of the proposed algorithm was verified using full-life data.

The remaining chapters are arranged as follows: Section 2 provides an overview of feature extraction in the time, frequency and time-frequency domains and the methods of feature selection; Section 3 describes the theoretical knowledge of the SSAE model and the LSTM model; Section 4 describes

the general framework of the paper; Section 5 presents the experiments of this paper and the experimental analysis; Section 6 is article conclusion.

2. Feature extraction and selection

The accuracy of feature extraction directly affects the accuracy of subsequent classification and regression prediction. In order to provide a more comprehensive portrayal of equipment degradation state information, multi-domain features are extracted as feature sets in this paper. The feature extraction and selection methods are described in detail below.

2.1 Time domain feature extraction

During the degradation of a planetary gearbox, the amplitude and probability distribution of the vibration signal changes, resulting in changes in the statistical parameters that measure the characteristics of the signal [23]. These statistical parameters are simple to understand and have a clear physical meaning, so in this paper 12 statistical features are selected for analysis in the time domain feature extraction, as shown in Table 1. Where x is the sampled time signal, i is the sample index, and N is the number of samples.

Table 1 Statistical characteristics.

Feature	Expression	Feature	Expression
RMS	$(\frac{1}{N}\sum_{i=1}^{N} x_i^2)^{\frac{1}{2}}$	Impulse factor	$\frac{x\frac{1}{N}\sum_{i=1}^{N}|x_i|}{2}_{max}$
Kurtosis	$\frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\overline{x})^4}{\rho^4}$	Margin factor	$\frac{x\frac{1}{N}\sum_{i=1}^{N}|x_i|}{}_{max}$
Skewness	$\frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\overline{x})^3}{\rho^3}$	Mean	$\overline{x}=\frac{1}{N}\sum_{i=1}^{N} x_i$
Peak to Peak(P-P)	$x_{max}-x_{min}$	Standard deviation(Std)	$\sigma=(\frac{1}{N}\sum_{i=1}^{N=1}(x_i-mean))^{\frac{1}{2}}$
Crest factor	x_{max}	Energy	$\sum_{i=1}^{N} x_i^2$
Shape factor	$RMS/(\frac{1}{N}\sum_{i=1}^{N}|x_i|)$	Variance	$\frac{1}{N}\sum_{i=1}^{N=1}(x_i-\overline{x})^2$

2.2 Frequency domain feature extraction

The authentic sign of a planetary gearbox includes a large amount of noise, resulting in that the impulse characteristics are often masked by signals such as noise and hard to extract in the early stages of a fault. Dwyer [41] found that the Kurtosis in its time domain index was sensitive to the shock signal, so in order to extract the transient information of the signal, the Kurtosis was introduced into the frequency domain and the concept of frequency domain Kurtosis was proposed. Subsequently, researcher Antoni [2] gave a definition of Spectrum Kurtosis (SK) and developed the fast SK algorithm. SK is also considered to be the most effective method for detecting impulse characteristics in planetary gearboxes. Reference [2] provides a specific formula for calculating spectral kurtosis.

In this paper, when extracting frequency domain features, the authentic vibration signal is first analyzed by SK algorithm, and then the statistical features in Table 1 are extracted as frequency domain features. Compared with the statistical features directly extracted from the original vibration signal, the statistical features extracted through SK algorithm analysis and processing are more robust and can better extract fault sensitive features [3].

2.3 Time-frequency domain feature extraction

This paper mainly uses Variational Mode Decomposition (VMD) to decompose the raw vibration signal to obtain multiple IMF components, and then extracts the fuzzy entropy of every IMF as time-frequency domain characteristics. The specific calculation process of VMD and multi-scale fuzzy entropy is as follows.

(1) Variational Mode Decomposition

Common time and frequency domain signal processing methods mainly deal with steady state signals, but the collected planetary gearbox vibration signals are non-linear and non-stationary. VMD [13] can effectively process non-stationary and nonlinear signals, and VMD algorithm overcomes the defects of Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD) and Local Mean Decomposition (LMD), such as modal aliasing and slow processing speed. Reference [13] provides a specific calculation process for solving time-frequency domain problems using VMD.

(2) Multiscale fuzzy entropy

Fuzzy entropy can measure the complexity of the signal time series and overcomes the disadvantages of sample entropy. The use of an affiliation function instead of a threshold improves the statistical stability of the algorithm. When a planetary gearbox is abnormal, its signal changes and exhibits a fuzzy entropy value that is different from the normal state.

The specific calculation of fuzzy entropy is as follows [11].

Step1: Define the phase space dimension $[u(1), u(2), \cdots u(M)]$ for the input time series n, reconstruct a set of n dimensional vectors $X_n(i)$ based on the original series as:

$$X_n(i) = [u(i), u(i+1), \cdots, u(i+n-1)] - u_0(i) \quad (1)$$

$$i = 1, 2, \cdots, M - n + 1 \quad (2)$$

$$u_0(i) = \frac{1}{n} \sum_{j=0}^{n-1} u(i+j) \quad (3)$$

Step2: Introduce the fuzzy entropy subordination function $A(x)$ as:

$$A(x) = \begin{cases} 1, x = 0 \\ exp[-In2(\frac{x}{r})^2], x > 0 \end{cases} \quad (4)$$

In the equation, r is the similarity tolerance limit, defined as k times the standard deviation of the original one-dimensional time series. That is $r = k \cdot SD$, SD is the standard deviation of the original data.

Step3: Calculate the similarity A_{ij}^n between vectors $X_n(i)$ and $X_n(j)$ as:

$$A_{ij}^n = \begin{cases} 1, d_{ij}^n = 0 \\ exp\left|-In2(\frac{d_{ij}^n}{r})^2\right|, d_{ij}^n > 0 \end{cases} \quad (5)$$

$$i = 1, 2, \cdots, M - n + 1 \quad (6)$$

In the equation: d_{ij}^n is the maximum absolute distance between the vector $X_n(i)$ and $X_n(j)$. The expression is:

$$d_{ij}^n = \max_{p=1,2,\cdots,n} [|u(i+p-1) - u_0(i)|] - [|u(j+p-1) - u_0(j)|] \quad (7)$$

Step4: For each dimensional vector i, taking the average value yields.

$$C_i^n(r) = \frac{1}{M-n} \sum_{j=0, j\neq i}^{M-n+1} A_{ij}^n \quad (8)$$

Definition:

$$\phi^n(r) = \frac{1}{M-n+1} \sum_{i=1}^{M-n+1} C_i^n(r) \quad (9)$$

Step5: Add 1 to the mode dimension and repeat Step1 to Step4 for 1 set of $n+1$ dimensional vectors to obtain.

$$\phi^{n+1}(r) = \frac{1}{M-n} \sum_{i=1}^{M-n+1} C_i^{n+1}(r) \quad (10)$$

Step6: The original time series fuzzy entropy $FuzzyEn(n, r)$ is:

$$FuzzyEn(n, r) = In\phi^n(r) - In\phi^{n+1}(r) \quad (11)$$

From the above algorithm, it can be seen that the signal is decomposed by VMD to derive multiple IMF components, thus realizing the multiscale of the vibration signal. The fuzzy entropy is extracted from the IMF components, which is the multiscale fuzzy entropy of the original vibration signal.

2.4 Normalization

Because each feature has a different magnitude, each feature is normalized by the equation (21):

$$F_i = \frac{f_i - \overline{f}}{\sigma} \quad (12)$$

In the equation, F_i is the normalized feature; $\overline{f}$ is the original mean of the feature; σ is the standard deviation of the feature.

2.5 Feature selection

The selection of feature quality has a direct impact on the accuracy of health state assessment and remaining life prediction. The selection of superior features reduces the dimensionality of the data, eliminates irrelevant data, and also improves prediction accuracy, among other things. Therefore, this paper first uses the locally weighted scatterplot smoothing (Loess) method [17] to smooth the time domain characteristic curves, the frequency domain characteristic curves and the fuzzy entropy curves of each IMF component. LOESS not only eliminates noise, but also captures important feature trends. Monotonicity, correlation and robustness are then selected as criteria for evaluating the goodness of the features.

2.5.1 Monotonicity

Monotonicity measures the tendency of a feature to be monotonically increasing or monotonically decreasing. A good predictive feature should be monotonically related to the degradation process and is calculated as [12].

$$Mon_i = \left| \frac{\#dF_i^{positive}}{T-1} - \frac{\#dF_i^{negative}}{T-1} \right| \qquad (13)$$

In the equation, Mon_i is the monotonicity of the i-th feature; $\#dF_i^{positive}$ is the number of positive derivatives of the feature; $\#dF_i^{negative}$ is the number of negative derivatives of the feature; and T is the number of data points collected.

2.5.2 Correlation

Correlation indicates the correlation between characteristics and time, that is, how characteristics change with time. The linear curve has a strong correlation with time, while the nonlinear curve will reduce its correlation with time with the increase of nonlinearity. The specific calculation formula is [12]:

$$Corr_i = \frac{N\sum_{n=1}^{N} f_n t_n - \sum_{n=1}^{N} f_n \sum_{n=1}^{N} t_n}{\sqrt{N\sum_{n=1}^{N} f_n^2 - (\sum_{n=1}^{N} f_n)^2}\sqrt{N\sum_{n=1}^{N} t_n^2 - (\sum_{n=1}^{N} t_n)^2}} \qquad (14)$$

Equation: $Corr_i$ is the correlation of the i-th feature; N is the number of samples; f_n is the i-th feature of the n-th feature; t_n is the time series.

2.5.3 Robustness

Robustness reflects the ability of a feature to maintain its original trend in the face of disturbances such as external perturbations and internal accuracy. After decomposing the feature into a trend component and a residual error component, the robustness of the feature is assessed using the residual error component [12].

$$Rob_i = \frac{1}{N}\sum_{n=1}^{N} exp\left(-\left|\frac{resF_i}{F_i}\right|\right) \qquad (15)$$

$$F_i^{original} = F_i^{smoothed} + resF_i \qquad (16)$$

In the equation, Rob_i is the robustness of the i-th feature; $resF_i$ is the residual error; $F_i^{original}$、$F_i^{smoothed}$ is the feature vector before and after smoothing.

In order to select the best feature to represent the degradation trend of the planetary gearbox, a comprehensive indicator (CI) was constructed using the three intrinsic attributes mentioned above. 0.4, 0.3 and 0.3 were assigned to the weights of monotonicity, correlation and robustness, respectively [12]. The ranking was then done in ascending order according to the size

of CI.

$$CI_i = \omega_1 Mon_i + \omega_2 Corr_i + \omega_3 Rob_i \qquad (17)$$

$$s.t. \begin{cases} \omega_n > 0 \\ \sum_n \omega_n = 1, n = 1,2,3 \end{cases} \qquad (18)$$

3. HI construction and life prediction

3.1 SSAE Network

In order to learn the non-linear features more fully, the SAE network model was stacked to form the SSAE, as shown in Figure 1.
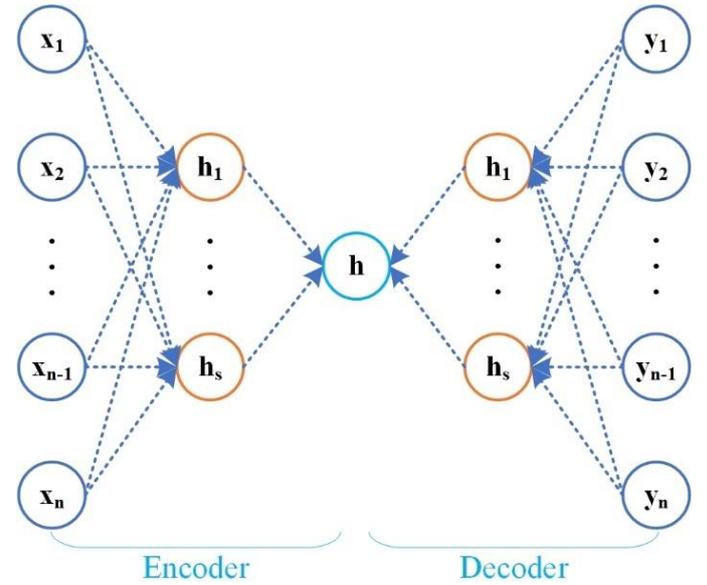

Fig. 1. SSAE structure.

The SSAE model uses the hidden layer features of the first layer of the SAE model as input features for the second layer of the SAE model. Its training process focuses on inputting optimized features into the SSAE model and then learning the non-linear features of the full-life data layer by layer for global tuning. The parameters of SAE mainly include reconstruction error, weight regularization term and penalty factor, and the commonly used penalty factor is Kullback-Leibler (KL) scatter. The formula for SAE is as follows [35]:

Reconstruction errors:

$$J_{error} = \frac{1}{Z}\sum_{z=1}^{Z}(y_i^z - x_i^z)^2 \qquad (19)$$

KL scatter:

$$KL(\rho\|\hat{\rho}_j) = \rho\,log\frac{\rho}{\hat{\rho}_j}KL(\rho\|\hat{\rho}_j) = \rho\,log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\,log\frac{1-\rho}{1-\hat{\rho}_j} \qquad (20)$$

Weight regularization term:

$$J_{weight} = \frac{\lambda}{2}\sum_{l=1}^{L}\sum_{i=1}^{C_l}\sum_{j=1}^{C_{l-1}}\left(w_{ji}^{(l)}\right)^2 \qquad (21)$$

Loss function:

$$J_{loss} = J_{error} + \beta \sum_{j=1}^{m} KL(\rho \| \hat{\rho}_j) + J_{weight} \qquad (22)$$

Where Z is the number of training samples, L is the number of network layers, C_l denotes the number of neuron nodes in layer l. ρ , $\hat{\rho}_j$ denotes the sparsity coefficient and the average activation value of neurons in the hidden layer, respectively, λ is the weight decay coefficient, β is the sparsity penalty factor, and m is the number of hidden neurons model.

3.2 LSTM

LSTM is a prediction of time series, by constructing storage units to store long-term memory information, especially in life span prediction, LSTM has been better used. The LSTM controls the state of the memory cell by linking the three cells of the oblivion gate, the input gate and the output gate through point multiplication. The forget gate f_t is used to control whether information in the memory unit is saved or discarded. The function of the input gate i_t is used to estimate whether to let the input information into the current memory cell state or not. The output gate O_t serves much the same purpose as the input gate and is used to determine whether the current signal will be output to the next layer. The specific structure is shown in Figure 2.
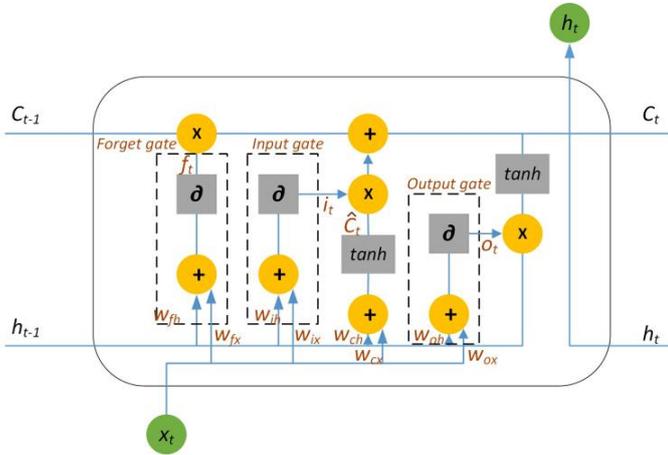

Fig. 2. LSTM structure.

$$f_t = sigmoid(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \qquad (23)$$
$$i_t = sigmoid(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \qquad (24)$$
$$O_t = sigmoid(W_{oh}h_{t-1} + W_{ox}x_t + b_i) \qquad (25)$$
$$\hat{C}_t = tanh(W_{ch}h_{t-1} + W_{cx}x_t + b_c) \qquad (26)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \qquad (27)$$
$$h_t = O_t \cdot tanh(C_t) \qquad (28)$$

The above equations illustrate the principles of calculating forgetting, input and output gates, where $\hat{C}_t$ is the memory cell, C_t is the memory cell and h_t is the hidden state. W is the weight matrix of the three gate cells and b is the threshold. $sigmoid$ and $tanh$ are the activation functions and "" represents the dot product.

3.3 Predictive performance evaluation indicators

For a more intuitive comparative analysis of the forecasting models, the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) were chosen as evaluation metrics and defined as follows.

$$RMSE = \sqrt{\frac{\sum_{K=1}^{k}(HI_{actual}^k - HI_{predicted}^k)^2}{k}} \qquad (29)$$

$$MAE = \frac{1}{k}\sum_{k=1}^{k}|HI_{actual}^k - HI_{predicted}^k| \qquad (30)$$

In the equation, HI_{actual} is the actual HI value; $HI_{predicted}$ is the predicted HI value and k is the sample size.

3.4 Remaining useful life prediction framework based on health indicators

This paper takes the perspective of HI construction. In order that the weak degradation features can be extracted more comprehensively and adequately, the full-life vibration signals from four sensors of the planetary gearbox were collected. Then, the degradation features of multiple domains were extracted using signal processing techniques, and finally the excellent features were selected by combining mathematical properties such as monotonicity, correlation and robustness to input into the SSAE network model to construct the HI. And the RUL prediction was carried out using LSTM. The specific implementation steps are as follows.

Step1: Time domain feature extraction. 12 statistical features for each way of the original vibration signal are extracted.

Step2: frequency domain feature extraction. In order to effectively detect and extract pulse features, the SK algorithm is first used to process each way of the original vibration signal, and then 12 statistical features are extracted as frequency domain features respectively.

Step3: Time-frequency domain feature extraction. In order to extract the weak degradation features, the VMD algorithm is used to process each way of the original vibration signal, decompose it to obtain multiple IMF components, and then extract the fuzzy entropy of each IMF element as time-

frequency domain features.

Step4: Monotonicity, trending and robustness are used to select superior features in multiple domains for input into the SSAE network model for data fusion to construct suitable and accurate HI.

Step5: Construction of HI. determination of thresholds and

degradation times is based on HI curves.

Step6: Input the constructed HI as input and labels into the LSTM for RUL prediction.

Step7: Output the prediction results.

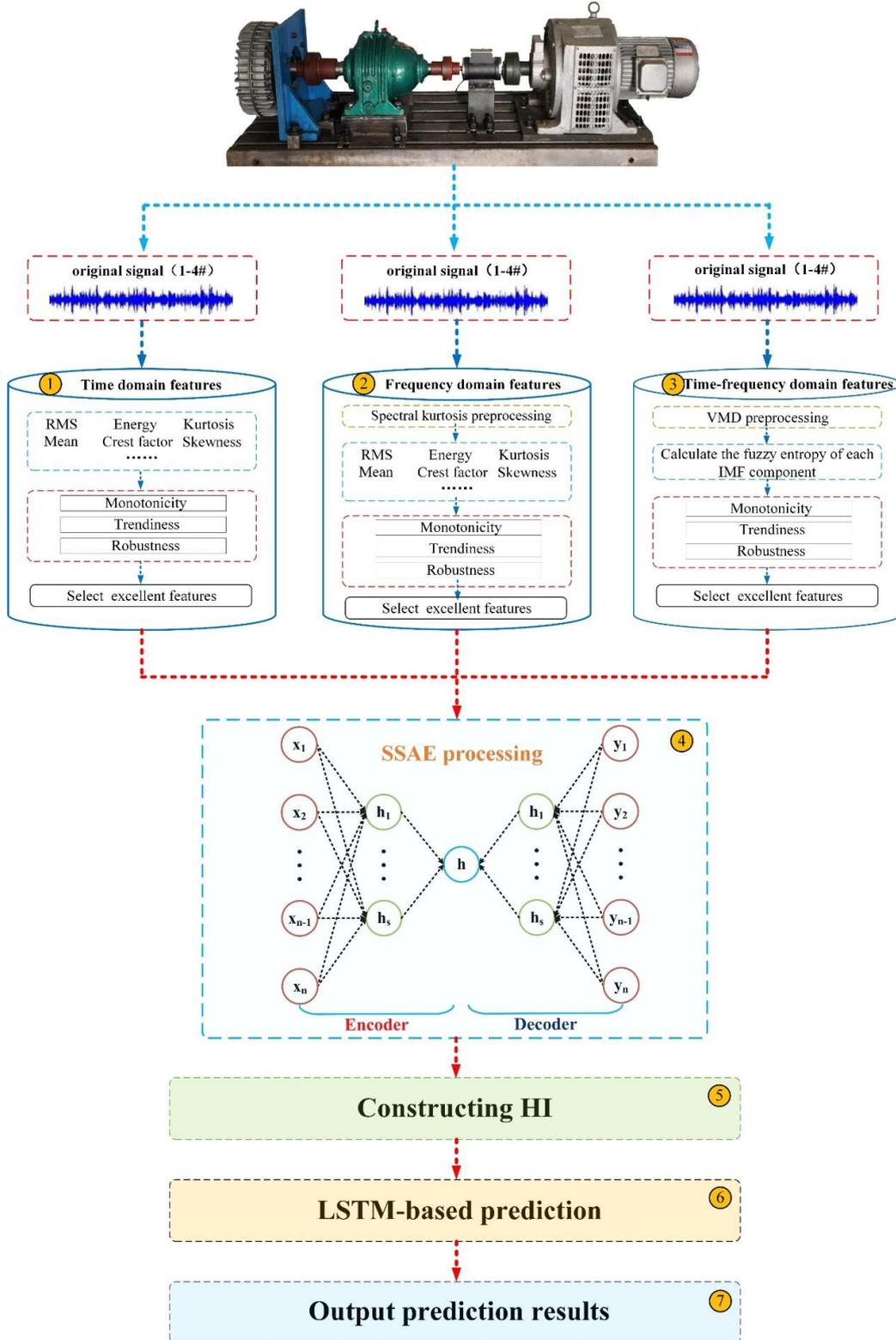The flow chart for the calculations in this paper, is shown in Figure 3.

Fig. 3. Structural framework of this paper.

4. Experimental results and analysis

4.1 Description of the experiment

To investigate the performance degradation process of planetary gearboxes as well as to validate the proposed method of constructing HI based on SSAE and performing RUL prediction using LSTM in this paper. The group built a planetary gearbox test rig and carried out 1000 hours of accelerated degradation experiments on planetary gearboxes, collecting data on the entire process of performance degradation of planetary gearboxes from intact to wear failure.

The planetary gearbox test stand consists of three speed control motors, a magnetic powder brake, a speed and torque sensor and a planetary gearbox. Four vibration sensors are arranged on the planetary gearbox. The sensor test stand and the layout of the sensors are shown in Figure 4. The planetary gearbox used in this experiment is a single-pole planetary gearbox, model NGW11. This single-pole planetary gearbox consists of a gear ring, a sun wheel and three planetary wheels, which are connected by a planetary carrier.
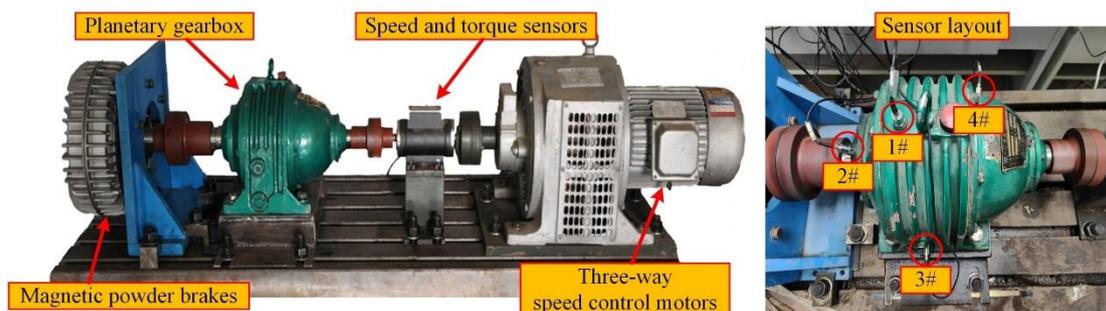

Fig. 4. Planetary gearbox test rig.

The vibration signal from the planetary gearbox was acquired at a sampling frequency (Fs) of 20 kHz; the sampling time was 12 s; the signal was acquired at 5 min intervals; the total duration of the acquisition was 1000 h. The speed and load current parameters during the experiments were: 1000 rpm and 1A load current. The measured experimental results are shown in Figure 5. After the failure, the sun wheel, the gear ring and the planetary wheel all experienced wear to varying degrees.
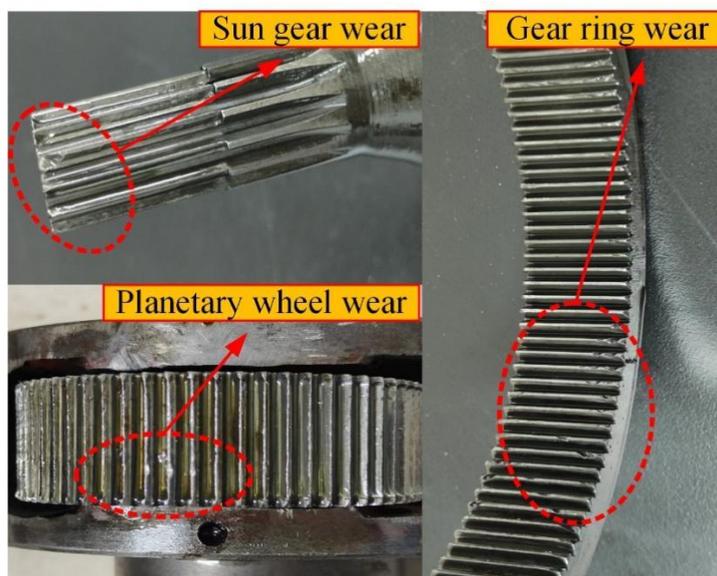

Fig. 5. Gear wear.

4.2 Lifetime data

In this experiment, four raw vibration signals from the planetary gearbox were collected, as shown in Figure 6. The 4 sensors are in different locations but the performance degradation process is similar. Initially the acceleration values are relatively flat and gradually increase with the operation of the planetary gearbox. In addition, vibration signals are collected at different positions on the planetary gearbox. This gives a more complete reflection of the degree of degradation of the planetary gearbox.
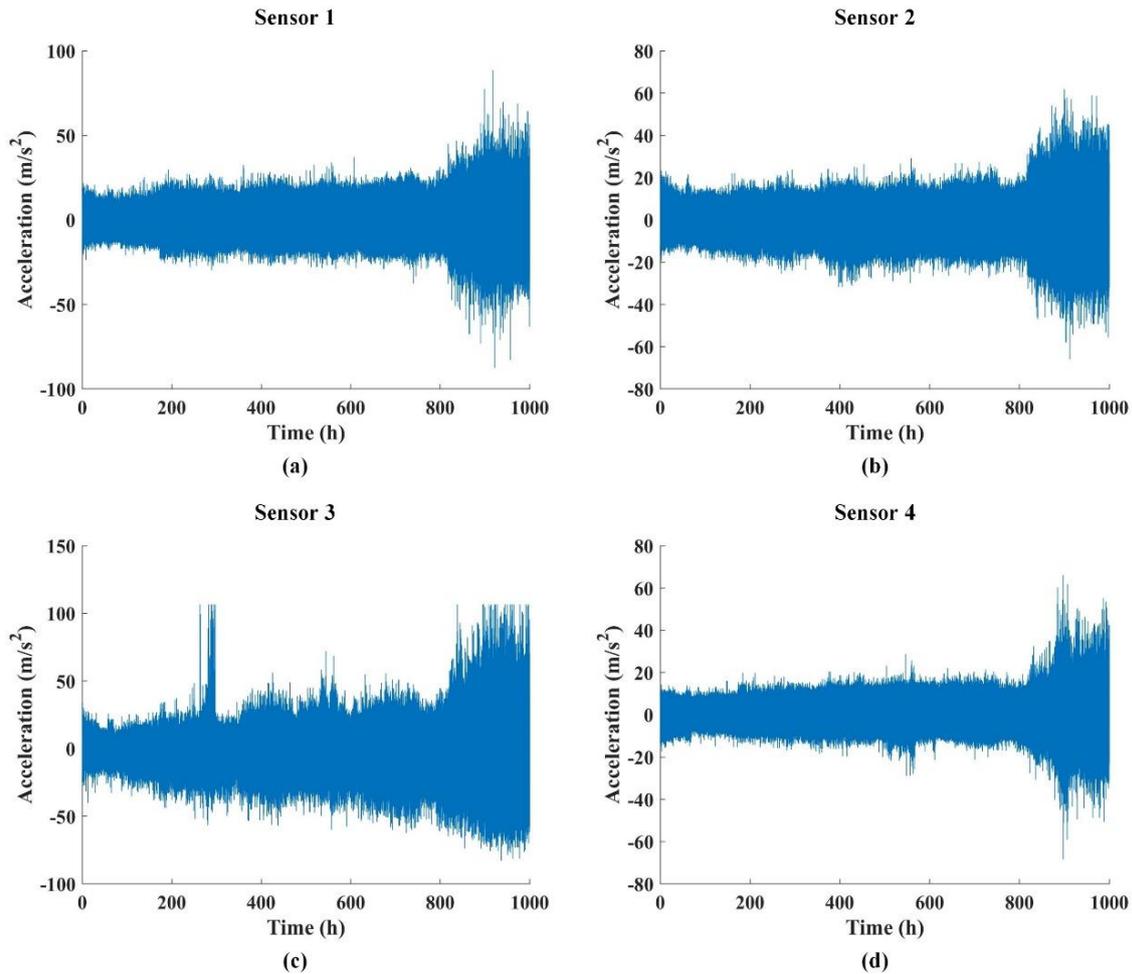
Fig. 6. Four full life vibration signals, (a) sensor 1; (b) sensor 2; (c) sensor 3; (d) sensor 4.

4.3 Feature extraction and selection

In this paper, the signals from 4 sensors of the planetary gearbox are collected, and only the 1st sensor is used as an example for specific analysis.

4.3.1 Time domain feature extraction and selection

Directly extract 12 statistical features from the original vibration signal. Monotonicity, correlation and robustness were calculated for each feature as well as the overall value. This is shown in Table 2. The monotonicity of all 4 indicators, RMS, Peak2Peak, Energy and Std, is greater than 0.28, and the correlation and robustness indicators are greater than 0.8, which is higher than the other characteristics. The CI values for all four metrics are greater than 0.6, with larger values indicating better performance of the degraded features. Therefore, the first four excellent features with larger CI values were selected as input to the SSAE model. To further eliminate the volatility generated by noise and to obtain more intuitive information

about the health of the equipment, LOESS is used to smooth each feature. In LOESS, points closer to the fitted point are given more weight and points further away are given less weight. The 4 sensitive time domain features and the smoothed feature curves are given in Figure 7.

Table 2. Comparison of intrinsic properties of time-domain features.

| Features	Monotonicity	Correlation	Robustness	CI
RMS	0.282	0.885	0.838	0.630
Peak2Peak	0.283	0.883	0.834	0.628
Energy	0.282	0.836	0.832	0.613
Std	0.280	0.823	0.842	0.612
var	0.227	0.756	0.765	0.547
marginfactor	0.226	0.754	0.752	0.542
crestfactor	0.219	0.676	0.742	0.513
shapefactor	0.230	0.622	0.758	0.506
Kurtosis	0.231	0.612	0.664	0.475
impulsefactor	0.116	0.596	0.566	0.395
skewness	0.118	0.589	0.546	0.388
Mean	0.124	0.566	0.493	0.367
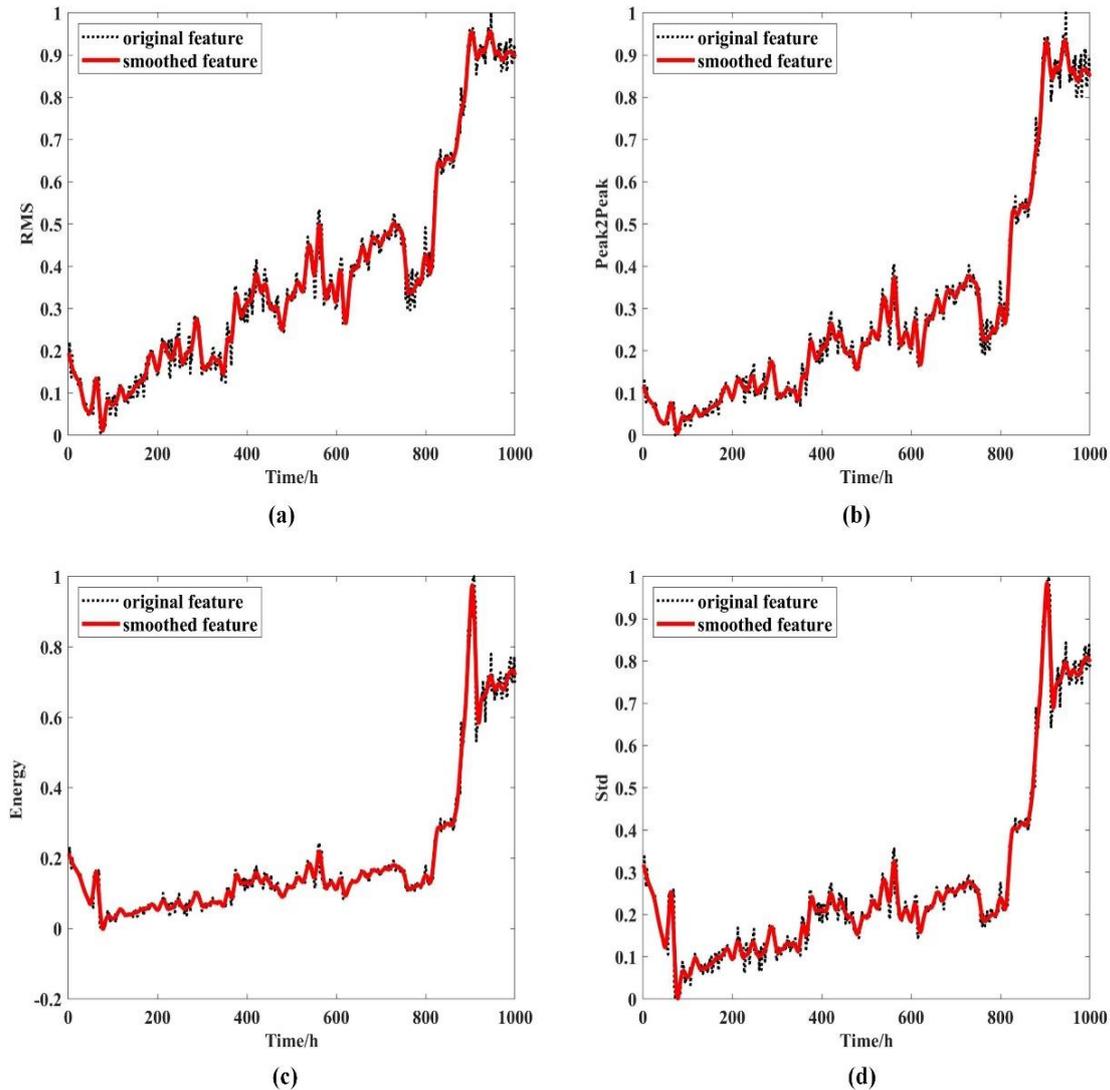
Fig. 7. Time domain characteristic curves: (a) Std features; (b) RMS features; (c) Peak2Peak features; (d) Energy features.

4.3.2 Extraction and selection of frequency domain feature

The 12 statistical features were extracted after the SK transformation. Monotonicity, correlation and robustness as well as overall CI values were calculated for each feature, as shown in Table 3. SK-Kurtosis, SK-impulsefactor, SK-Peak2Peak and SK-crestfactor all had CI values greater than 0.5 and were higher than the remaining features. Therefore, the first four excellent features with larger CI values were selected as inputs to the SSAE model. Figure 8 shows the results of the LOESS optimisation of the four excellent features, with the smoothing operation enabling a smoother degradation feature of the planetary gearbox to be obtained. A better overview of the current and future health of the planetary gearbox can be obtained.

Table 3. Comparison of the intrinsic properties of the frequency domain features.

| Features	Monotonicity	Correlation	Robustness	CI
SK-Kurtosis	0.241	0.768	0.723	0.544
SK-impulsefactor	0.235	0.754	0.705	0.532
SK-Peak2Peak	0.232	0.775	0.682	0.530
SK-crestfactor	0.246	0.752	0.605	0.506
SK-var	0.219	0.744	0.621	0.497
SK-Std	0.216	0.644	0.620	0.467
SK-Mean	0.242	0.652	0.511	0.446
SK-RMS	0.212	0.516	0.612	0.423
SK-Energy	0.112	0.566	0.625	0.402
SK-skewness	0.142	0.592	0.556	0.401
SK-marginfactor	0.134	0.591	0.539	0.393
SK-shapefactor	0.101	0.562	0.495	0.358
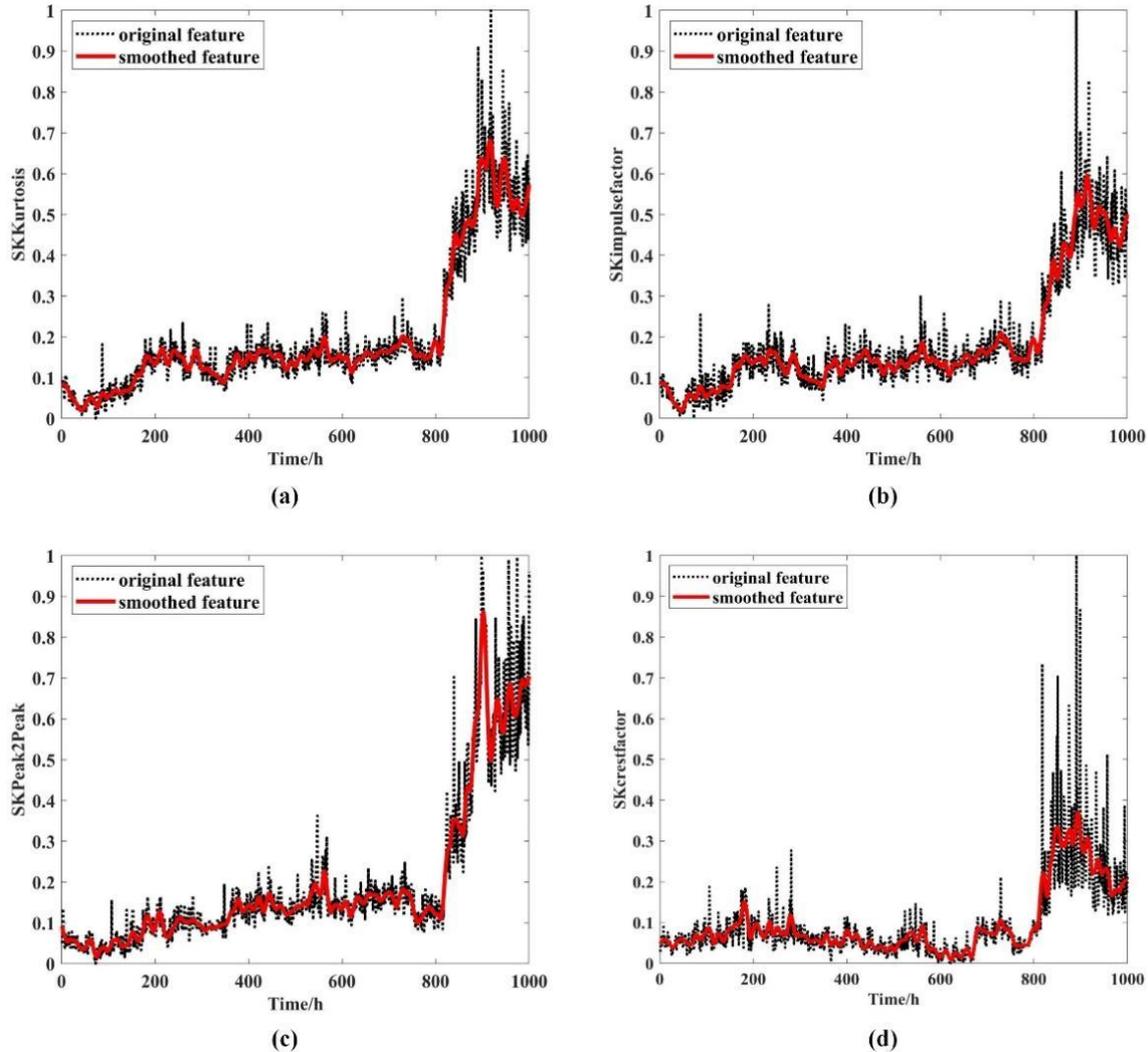
Fig. 8 Frequency domain feature curves: (a) SKKurtosis feature; (b) SKimpulsefactor feature; (c) SKPeak2Peak feature; (d) SKcrestfactor feature.

4.3.3 Extraction of time-frequency domain feature

In this paper, the signal is decomposed by VMD to derive multiple IMF components, and the fuzzy entropy of every IMF element is extracted as a time-frequency domain feature. The VMD decomposition requires that the value of the decomposition layer K be determined first. In this paper, the central frequency method is used to determine the value of K. Over-decomposition is considered to have occurred when the values of the centre frequencies are similar. After decomposing the signal by VMD, the centre frequencies of each modal component at different K values are shown in Table 4. When the number of modal components is 8, the centre frequencies 4038Hz and 4498Hz are close to each other and modal mixing may occur, So the number of modes should be selected as 7. The fuzzy entropy of each IMF is calculated to form a subset of

features: FuzzyEn1 to FuzzyEn7, where the spatial dimension is set to $n = 2$ and $r = 0.2SD$ is the similarity tolerance limit.

Table.4. Center frequency corresponding to different K.

Number of Modes	Central Frequency /Hz
2	7032098
3	45012132735
4	259 939 15503538
5	253 909 136319793980
6	251 901 1311171133424335
7	248 890 12451578212835644386
8	247 889 1237157321323299940384498
9	238 851 10281328160621333300040454499

Monotonicity, correlation and robustness and overall CI values were calculated for each IMF fuzzy entropy. As shown in Table 5. The g values of FuzzyEn2, FuzzyEn4, FuzzyEn3 and FuzzyEn6 were all greater than 0.5. Therefore, the first four

excellent features with larger *CI* values were selected as inputs to the SSAE model. Figure 9 shows that the features become smoother, eliminating the volatility generated by noise and obtaining more intuitive information on degraded features. In addition, the smooth time sequence is used as enter to the subsequent prediction mannequin to achieve more accurate RUL results. The validity of the prediction model is further improved.

Table 5. Comparison of intrinsic properties of time-frequency domain features.

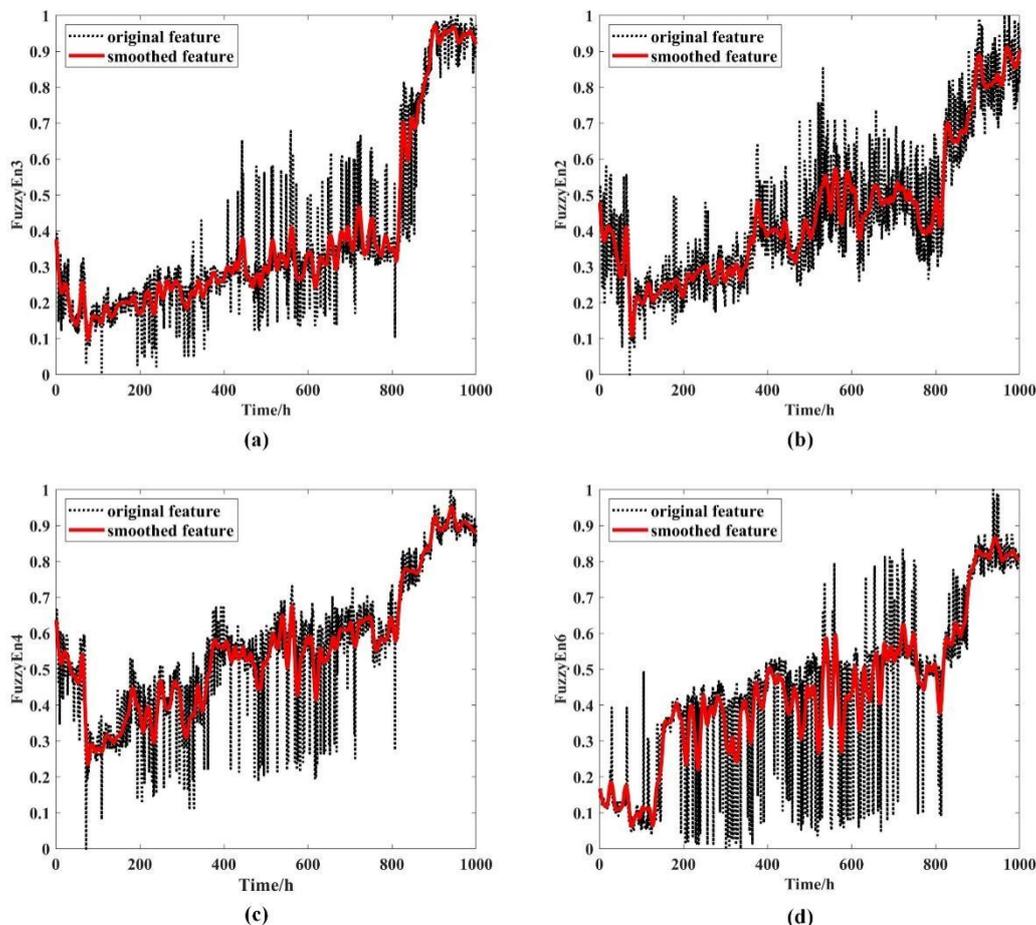| Features	Monotonicity	Correlation	Robustness	*CI*
FuzzyEn3	0.278	0.845	0.687	0.571
FuzzyEn2	0.273	0.754	0.576	0.508
FuzzyEn4	0.276	0.739	0.582	0.507
FuzzyEn6	0.269	0.734	0.577	0.500
FuzzyEn5	0.238	0.661	0.549	0.458
FuzzyEn7	0.171	0.664	0.535	0.428
FuzzyEn1	0.132	0.637	0.526	0.402

Fig. 9. Time-frequency domain feature curves; (a) FuzzyEn3 feature; (b) FuzzyEn2 feature; (c) FuzzyEn4 feature; (d) FuzzyEn6 feature.

In order to prove the superiority of VMD decomposition for extracting time-frequency domain features of original vibration signals. This paper compares and analyses the signal processing methods of EMD, EEMD and CEEMD. Firstly, each signal processing method is used to decompose the original vibration signal to obtain multiple IMF components, and then the IMF component fuzzy entropy of each signal processing method is calculated. Finally, the top 4 superior features of each signal

processing method were selected using monotonicity, correlation and robustness and the *CI* values were calculated. This is shown in Figure 10. The values of the first four excellent features selected by VMD decomposition are larger than those obtained by other signal decomposition, which verifies the superiority of the original vibration signal decomposed by VMD in this paper as time-frequency domain features. In addition, the EMD decomposition yields large values of *CI* for

the first excellent feature, but subsequent features start to become less good and may appear to be modally confounded. CEEMD decomposes the signal, and the *CI* values of the extracted time-frequency domain features are relatively stable, verifying that CEEMD can decompose the original vibration signal better. However, the *CI* value is still lower than that of the VMD method. Therefore, the VMD method is chosen to be better.
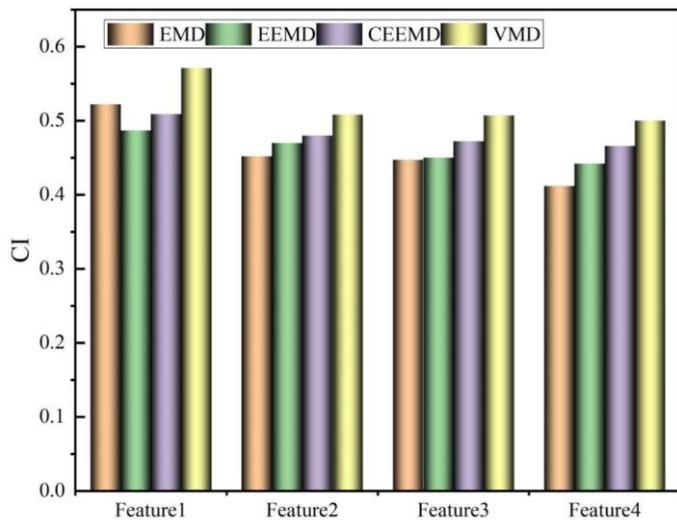

Fig. 10 Comparison of values for different signal processing methods.

In this paper, the *CI* value of each domain is calculated , and the *CI* values are ranked. The larger the *CI* value, the better the feature reflects the degradation state of the planetary gearbox and is more suitable for HI construction and prediction. In this experiment, the vibration signals of four sensors were collected. For each sensor, four time domain features, four frequency domain features and four time-frequency domain features were selected based on the above experimental analysis. As a result, 12 superior features were extracted from each signal, and a total of 48 sensitive features were extracted from the 4-way signals for input into the SSAE model for fusion to construct the HI.

4.4 HI construction and assessment

4.4.1 HI construction

This section uses the SSAE model to fuse superior features from multiple domains into one-dimensional features to construct HI. As the number of SSAE network layers is not more, the better. Too many layers will cause network instability and increase computing time. Too few layers will cause insufficient high-dimensional feature extraction. Therefore, in this paper, a 2-

layer SSAE network was constructed, in which the number of nodes in the hidden layer was set to 48 and 1 respectively; the L2 regularization weight decay coefficients were both set to 0.0002; the sparse penalty weight parameters were set to 0.0002 and 0.002 respectively; and the sparsity parameters were both set to 0.01.

To demonstrate the superiority of the SSAE model, it is compared with PCA, KPCA, AE and Isometric Mapping (Isomap) in the construction of HI curves. Figure 11 shows a graph of the five methods for constructing the HI method. While the PCA, KPCA, AE and Isomap models also obtained results that approximated those of SSAE, the health curves for PCA, KPCA and AE were more volatile compared to the SSAE model. In particular, the HI curves constructed by the KPCA and Isomap algorithms showed large fluctuations in the first 200 hours. The model constructed by PCA is less volatile than the rest of the models, but still weak in monotonicity and correlation. The AE model fluctuates considerably after 400 hours, which seriously affects the accuracy of health assessment and life expectancy prediction. Therefore, through experimental analysis, the proposed HI based on SSAE fusion construction in this paper works best.
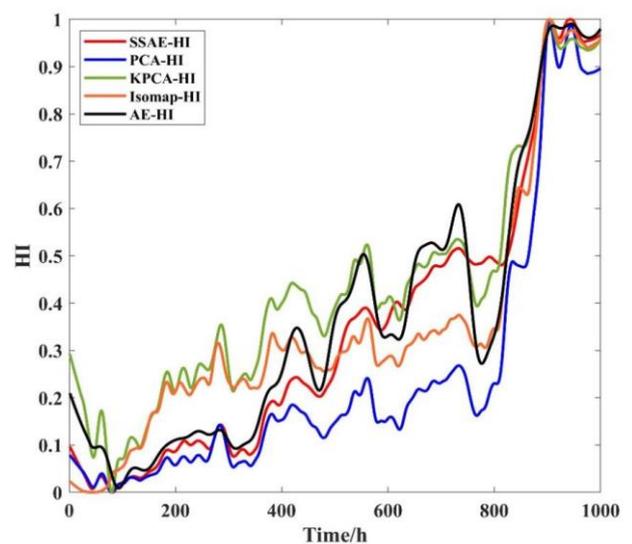

Fig. 11 Comparison of curves for different methods of constructing HI.

In addition, to further illustrate the advantages of the method SSAE proposed in this paper, a quantitative analysis of the five methods was carried out. Monotonicity, correlation and robustness and *CI* values were calculated for each model , as shown in Table 6. Of the four models, the SSAE model had a *CI*

value of 0.69, the highest value relative to the remaining four algorithms, and higher than the *CI* value of any single feature. It shows that the SSAE model is more suitable for fusion building HI and better able to characterize the health of the equipment, further improving the accuracy and validity of subsequent predictive models.

Table 6. Evaluation indicator scores for the HI construction method.

Features	Monotonicity	Correlation	Robustness	*CI*
SSAE-HI	**0.376**	**0.922**	**0.876**	**0.690**
AE-HI	0.198	0.714	0.820	0.539
KPCA-HI	0.236	0.794	0.807	0.575
PCA-HI	0.283	0.895	0.839	0.633
Isomp-HI	0.279	0.823	0.831	0.608

4.4.2 Health status classification

The degradation deepens as the planetary gearbox is operated. But the HI curve, constructed on the basis of real experimental data, is not a completely idealized monotonically rising curve, but has certain fluctuations. This is because there are various influencing factors in the real operation of the equipment, such as the complexity of the working environment. HI curves constructed using SSAE fusion. As shown in Figure 12. In order to avoid phenomena such as false alarms, combined with the HI curve, two warning lines are set up in this paper, namely warning line ① and warning line ②. Based on the HI curve it can be concluded that below the warning line ①, the HI degradation curve is relatively smooth. Between the warning line ① and the warning line ②, there is a significant degradation of the HI curve. Above the warning line ②, the HI curve starts to accelerate and degenerate, and the slope increases until the HI value is 1, and the planetary gearbox suffers from wear failure.

In this paper, the planetary gearbox is divided into 4 states from the beginning to the wear failure based on the HI curve constructed by fusion. They are normal operation (0-375 hours). At this time, the gearbox has experienced initial running in and can be used normally, with a long remaining service life; Initial recession (375-840 hours), when the gearbox can be used

normally without any constraints; Severe recession (840 - 910 hours), when the gearbox is no longer suitable for further use and it can only last for a limited period of time; Wear and tear failure (910-1000 hours), when the gearbox is already showing some wear and tear and could be damaged at any time. The SSAE model fusion constructs health indicators that first stabilize for a period of time and then gradually increase until the wear failure. The running time from the start of performance to normal, initial degradation is long, but the time from severe degradation to planetary gearbox wear failure is short. Therefore, if the health of the planetary gearbox can be accurately assessed and effective measures taken before serious degradation occurs, planetary gearbox failures can be avoided, reducing economic losses and avoiding safety incidents.
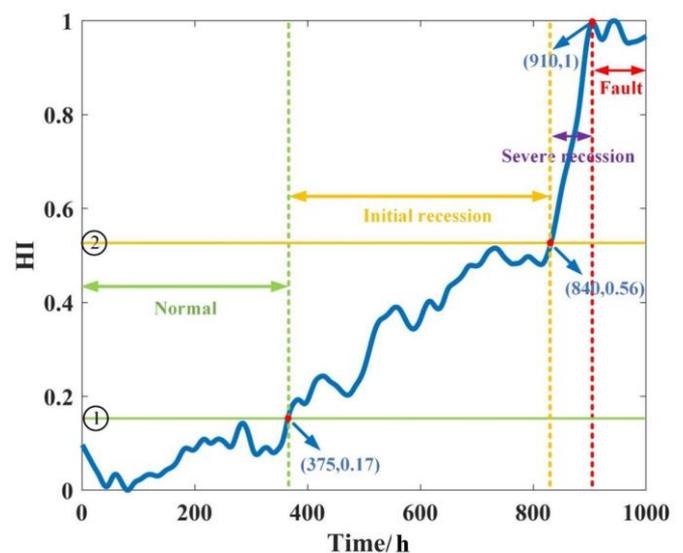

Fig. 12. Health status classification.

4.5 Remaining useful life prediction

4.5.1 RUL prediction at different times

After obtaining HI through SSAE model fusion, it is necessary to select an excellent prediction model for RUL prediction. In this paper, the LSTM network is chosen as the prediction model, and the details are shown in Table 7. The parameters of the LSTM neural network are initialized using standard initialization methods, and then the gradient descent algorithm is selected to update the parameters.

Table 7. LSTM model parameters.

Parameters	Number of hidden units	Number of outputs	Dropout	Number of iterations	Number of batches	Learning Rate
Numerical	300	1	0.2	100	16	0.002

To verify the excellence of the HI construction using SSAE fusion and the accuracy of the severe degradation time chosen as the prediction time. In this paper, HI points before 540h, 640h, 740h and the severe degradation time of 840 hours established in this paper were chosen as the training set for prediction. The fault threshold is set to 1 and then the LSTM network is used to predict the unknown HI points that follow. This is shown in Figure 13.

The predictions have been partially enlarged in order to show the results more clearly. It can be concluded from Figure 13 that the prediction is poorer starting at 540h and 640h. This is because 540h and 640h belong to the early and middle stages of initial gearbox degradation, where gearbox wear is low and the degradation trend is relatively smooth, hence the poorer prediction results. As the prediction starts and the time point continues to pass backward, the LSTM model will constantly revise the prediction results, and its output HI curve is nearer to the actual health status of the Planetary gearboxes. At 740h the predictions are already close to the true HI curve. The best results are achieved at 840h, when the prediction curve and the true curve coincide almost perfectly, resulting in a more accurate prediction. Although the smaller the training set, the worse its predictions and the longer it takes to reach the failure threshold. However, the predicted HI curves starting at different moments have similar degradation trends to the true HI curves. Furthermore, the prediction times of 540h and 640h fall within the early and middle stages of initial gearbox degradation, so the errors present are acceptable in the early predictions. The excellence of the HI method constructed in this paper is verified.
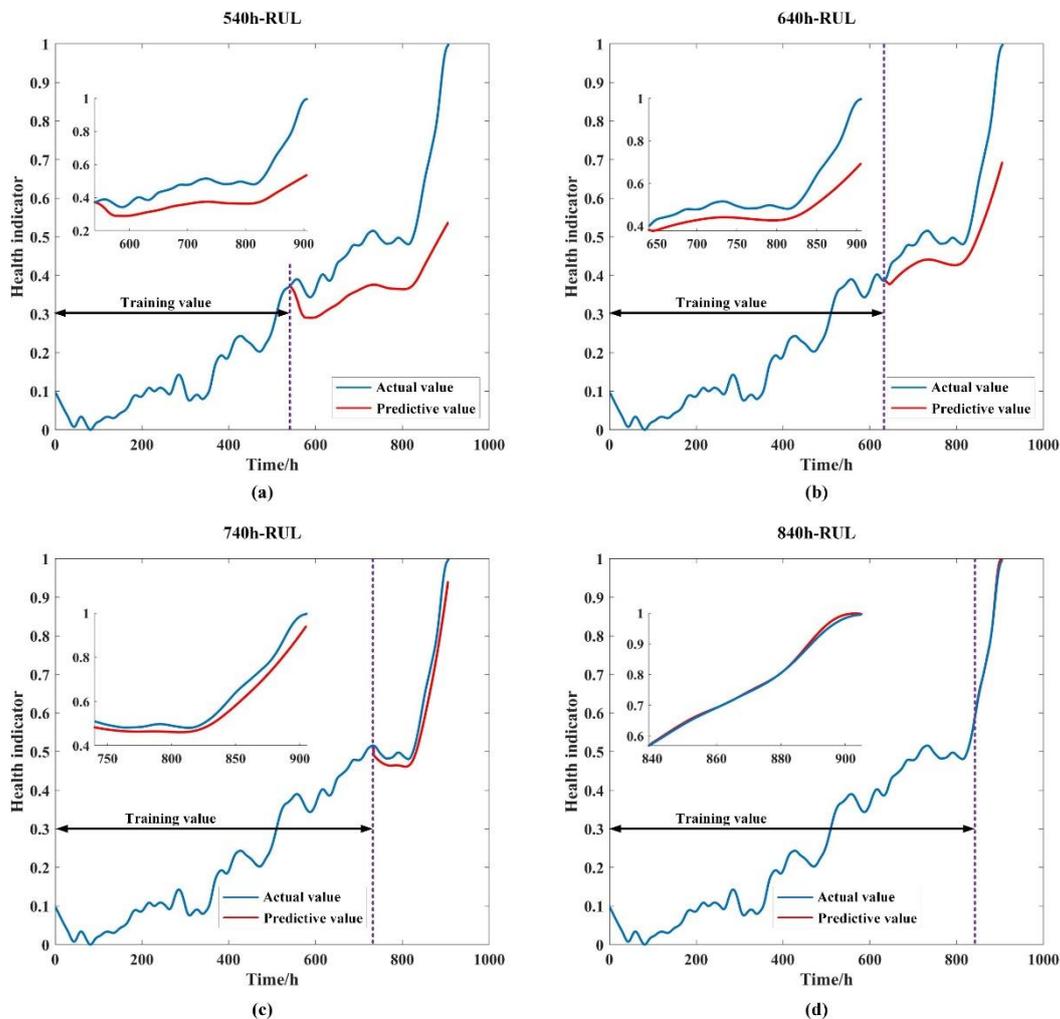

Fig. 13. Plots of predicted effects at different times: (a) RUL at 540h; (b) RUL at 640h; (c) RUL at 740h; (d) RUL at 840h.

In order to quantify the life predictions at different times, RMSE and MAE were chosen for the quantitative analysis. The results of the quantitative analysis of the experiment are shown in Table 8. 840h had lower values than the rest of the time predictions in all respects, with values of only 0.027 and 0.022 for RMSE and MAE, which validates the accuracy of the choice

of time of severe degradation as the prediction time. In addition, the 540h error is the largest because it is in the early stage of the initial recession of the gearbox, and the degradation trend is relatively smooth, so the prediction result is poor.

Table 8 Quantitative analysis of comparative experiments

Time	RMSE	MAE
840h	0.027	0.022
740h	0.491	0.349
640h	1.154	0.968
540h	1.379	1.162

4.5.2 RUL prediction for different models

To demonstrate the superiority of prediction using LSTM, 840h was chosen as the starting prediction point and the lifetime predictions of SVM, CNN and BP methods were compared. The comparison results are shown in Fig. 13. In order to see the predictions more clearly, the prediction results have been partially enlarged.

As can be seen in Figure 14(a), the predicted HI values are close to the actual HI values and the degradation trend of the predicted HI curve is consistent with the degradation trend of the actual HI curve. The rest of the methods have relatively poor prediction results compared to the prediction methods proposed in this paper. However, the SVM life prediction algorithm can still achieve good results, and its prediction trend is consistent with the real HI curve, which can reach the failure threshold. Indirectly, it shows that the HI constructed in this paper using SSAE model fusion can help RUL prediction and can improve the accuracy of prediction results. The lifetime prediction methods of CNN and BP have a good prediction trend in the early stages. After 880 hours, the results started to deteriorate and did not reach the failure threshold. This may be due to the poor performance of the CNN in processing long sequences of one-dimensional signals. The BP neural network is less effective in prediction due to its simple structure and poor generalization performance. In addition, using the LSTM algorithm, the prediction results using the HI points up to 740 hours ago as the training set are similar to the results of the BP neural network that chose the HI points up to 840 hours ago as the training set. This further validates the superiority of the LSTM algorithm in predicting long time series.
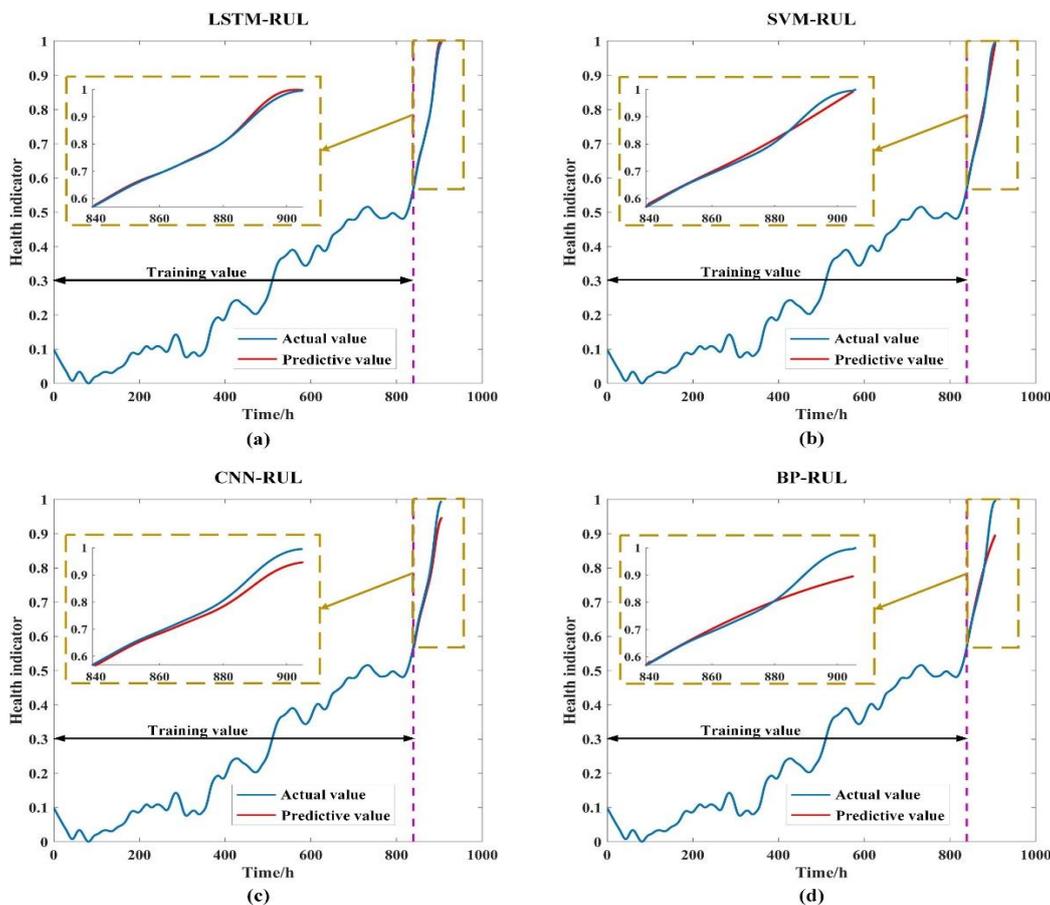

Fig. 14. RUL prediction comparison plots, (a) LSTM lifetime prediction; (b) SVM lifetime prediction; (c) CNN lifetime prediction; (d) BP lifetime prediction.

For a more intuitive comparative analysis of the above models, the RMSE and MAE were chosen for quantitative analysis. The results of the quantitative analysis of the experiments are shown in Table 9. Among them, LSTM has lower values than SVM, CNN and BP in all aspects. this also confirms that LSTM is able to cope with time series prediction problems better than other deep network models, often looking for potential correlations and important features between data, thus achieving more accurate predictions. However, Table 9 also shows that the RMSE and MAE values of other algorithms are still less than 0.5. It is verified that the SSAE-HI curve constructed in this paper still has a high comprehensive quality, which is helpful for lifetime prediction.

Table 9. Quantitative analysis of comparative experiments.

Models	RMSE	MAE
LSTM	0.027	0.022
SVM	0.139	0.114
CNN	0.267	0.228
BP	0.493	0.354

5. Conclusion

This paper takes an HI construction perspective. The signal processing techniques are combined with mathematical properties such as monotonicity, correlation and robustness. The excellent features in the multi-domain of the four vibration signals are extracted and selected respectively. They are then fed into a deep learning network model to fuse and construct HI.

It gives a more comprehensive reflection of the state of degradation of the equipment and helps with subsequent RUL predictions. The main conclusions of this paper are as follows.

(1) Using signal processing techniques such as SK and VMD, degradation features are extracted from the multi-domain respectively as feature sets, which can reflect the degradation state of the mechanical equipment in a more comprehensive manner.

(2) Excellent features were selected with the help of monotonicity, correlation and robustness and input to the SSAE network model for fusion to construct HI, which has the best comprehensive quality and gets rid of the shortcomings of traditional construction of HI.

(3) The HI constructed by the SSAE model with simple structure and strong feature extraction ability can better characterize the degradation state of the Planetary gearboxes. The RUL predictions was performed using the LSTM and validated on a full-life experimental dataset. The prediction results have minimal error.

The method of RUL predictions presented in this paper is mainly applied to the same working conditions, whereas in real industrial production, machinery and equipment are often in different working conditions. Therefore, the next step of the work will be to investigate the prediction of RUL across different working conditions.

References

1. Anowar F, Sadaoui S, Selim B. Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). Computer Science Review 2021; 40: 100378. doi: 10.1016/j.cosrev.2021.100378

2. Antoni J. Fast computation of the kurtogram for the detection of transient faults. Mechanical Systems and Signal Processing 2007; 21(1): 108-124. https://doi.org/10.1016/j.ymssp.2005.12.002

3. Ali J B, Saidi L, Harrath S, Bechhoefer E, Benbouzid M. Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning. Applied Acoustics 2018; 132: 167-181. https://doi.org/10.1016/j.apacoust.2017.11.021

4. Bai H, Zhan X, Yan H, Wu Z, Wen L, Jia X. Combination of Optimized Variational Mode Decomposition and Deep Transfer Learning: A Better Fault Diagnosis Approach for Diesel Engines. Electronics, 2022, 11(13): 1969. doi: 10.3390/electronics11131969.

5. Chen Y, Rao M, Feng K, Niu G. Modified Varying Index Coefficient Autoregression Model for Representation of the Nonstationary Vibration From a Planetary Gearbox. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-12. doi: 10.1109/tim.2023.3259048.

6. Chen D, Qin Y, Qian Q, Wang Y, Liu F. Transfer life prediction of gears by cross-domain health indicator construction and multi-hierarchical long-term memory augmented network. Reliability Engineering & System Safety, 2022: 108916. doi:10.1016/j.ress.2022.108916.

7. Chen X, Shen Z, He Z, Sun C, Liu Z. Remaining life prognostics of rolling bearing based on relative features and multivariable support vector machine. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 2013; 227(12): 2849-2860. https://doi.org/10.1177/0954406212474395

8. Chen J, Jing H, Chang Y, Qian L. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. Reliab Eng Syst Saf 2019; 185: 372-382. https://doi.org/10.1016/j.ress.2019.01.006

9. Cao Y, Ding Y, Jia M, Tian R. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. Reliab Eng Syst Saf 2021; 215: 107813.

10. Cheng C, Ma G, Zhang Y, Sun M, Teng F, Ding H, et al. A deep learning-based remaining useful life prediction approach for bearings. IEEE/ASME transactions on mechatronics 2020; 25(3): 1243-1254. https://doi.org/10.1109/TMECH.2020.2971503

11. Chen W, Wang Z, Xie H, Yu W. Characterization of surface EMG signal based on fuzzy entropy.IEEE Transactions on neural systems and rehabilitation engineering 2007; 15(2): 266-272. https://doi.org/10.1109/TNSRE.2007.897025

12. Chen D, Qin Y, Wang Y, Zhou J. Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction. ISA transactions 2021; 114: 44-56. https://doi.org/10.1016/j.isatra.2020.12.052

13. Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE transactions on signal processing 2013; 62(3): 531-544. https://doi.org/10.1109/TSP.2013.2288675

14. Fei Z, Yang F, Tsui K L, Li L, Zhang Z. Early prediction of battery lifetime via a machine learning based framework. Energy 2021; 225: 120205. https://doi.org/10.1016/j.energy.2021.120205

15. He Z, Shi T, Xuan J. Milling tool wear prediction using multi-sensor feature fusion based on stacked sparse autoencoders. Measurement 2022; 190: 110719. https://doi.org/10.1016/j.measurement.2022.110719

16. Igba J, Alemzadeh K, Durugbo C, Eiriksson E. Analyzing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes. Renewable Energy 2016; 91: 90-106. DOI: 10.1016/j.renene.2016.01.006

17. Javed K, Gouriveau R, Zerhouni N, Nectoux P. Enabling health monitoring approach based on vibration data for accurate prognostics. IEEE Transactions on industrial electronics 2014; 62(1): 647-656. DOI: 10.1109/tie.2014.2327917

18. Liu Y, Hu X, Zhang W. Remaining useful life prediction based on health index similarity[J]. Reliab Eng Syst Saf 2019; 185: 502-510. DOI: 10.1016/j.ress.2019.02.002

19. Luo J, Zhang X. Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction. Applied Intelligence 2022; 52(1): 1076-1091. DOI: 10.1007/s10489-021-02503-2

20. Lin H, Lei Z, Wen G, Tian X, Huang X, Liu J, Zhou H, et al. A Novel Approach of Label Construction for Predicting Remaining Useful Life of Machinery. Shock and Vibration 2021; 2021. DOI: 10.1155/2021/6806319

21. Liu J, Pan C, Lei F, Hu D, Zuo H. Fault prediction of bearings based on LSTM and statistical process analysis. Reliab Eng Syst Saf 2021; 214: 107646. doi:10.1016/j.ress.2021.107646.

22. Li J, Wang Z, Liu X, Feng Z. Remaining Useful Life Prediction of Rolling Bearings Using GRU-DeepAR with Adaptive Failure Threshold. Sensors, 2023, 23(3): 1144. doi: 10.3390/s23031144.

23. Malhi A, Yan R, Gao R X. Prognosis of defect propagation based on recurrent neural networks. IEEE Transactions on Instrumentation and Measurement 2011; 60(3): 703-711. DOI: 10.1109/tim.2010.2078296

24. Qin Y, Yang J, Zhou J, Pu H, Mao Y. A new supervised multi-head self-attention autoencoder for health indicator construction and similarity-based machinery RUL prediction. Advanced Engineering Informatics, 2023, 56: 101973. doi: 10.1016/j.aei.2023.101973.

25. Qin Y, Zhou J, Chen D. Unsupervised health indicator construction by a novel degradation-trend-constrained variational autoencoder and its applications. IEEE/ASME Transactions on Mechatronics, 2021, 27(3): 1447-1456. doi: 10.1109/tmech.2021.3098737.

26. Skrimpas G A, Ursin T, Sweeney C, Marhadi K, Mijatovic N, Holboell J. Residual signal feature extraction for gearbox planetary stage fault detection. Wind Energy 2017; 20(8): 1389-1404. doi: 10.1002/we.2099.

27. Saidi L, Ali J B, Bechhoefer E, Benbouzid M. Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived

indices and SVR. Applied Acoustics 2017; 120: 1-8. doi: 10.1016/j.apacoust.2017.01.005.

28. Shi H, Guo J, Bai X, Guo L, Liu Z, Sun J. Gearbox incipient fault detection based on deep recursive dynamic principal component analysis. IEEE Access 2020; 8: 57646-57660. doi: 10.1109/access.2020.2982213.

29. She D, Jia M, Pecht M G. Sparse auto-encoder with regularization method for health indicator construction and remaining useful life prediction of rolling bearing. Measurement Science and Technology 2020; 31(10): 105005. https://doi.org/10.1088/1361-6501/ab8c0f

30. Wang D. Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients. Mechanical Systems and Signal Processing 2018; 108: 360-368. DOI: 10.1016/j.ymssp.2018.02.034

31. Widodo A, Yang B S. Application of relevance vector machine and survival probability to machine degradation assessment. Expert Systems with Applications 2011; 38(3): 2592-2599. DOI: 10.1016/j.eswa.2010.08.049

32. Wang H, Liao H, Ma X, Bao R. Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model. Reliability Engineering & System Safety, 2021, 210: 107504. doi:10.1016/j.ress.2021.107504.

33. Yoo Y, Baek J G. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. Applied Sciences 2018; 8(7): 1102. DOI: 10.3390/app8071102

34. Yu W, Kim I I Y, Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. Reliab Eng Syst Saf 2020; 199: 106926. DOI: 10.1016/j.ress.2020.106926

35. Yin J, Yan X. Mutual information–dynamic stacked sparse autoencoders for fault detection. Industrial & Engineering Chemistry Research 2019; 58(47): 21614-21624. DOI: 10.1021/acs.iecr.9b04389

36. Zhou H, Huang X, Wen G, Lei Z, Dong S, Zhang P, et al. Construction of Health Indicators for Condition Monitoring of Rotating Machinery: A Review of the Research. Expert Systems with Applications 2022; 117297. DOI: 10.1016/j.eswa.2022.117297.

37. Zhou J, Qin Y, Luo J, Zhu T. Remaining useful life prediction by distribution contact ratio health indicator and consolidated memory GRU. IEEE Transactions on Industrial Informatics, 2022. doi: 10.1109/TII.2022.3218665.

38. Zhang Y, Xie M, He Y, Han X. Capability-based remaining useful life prediction of machining tools considering non-geometry and tolerancing features with a hybrid model. International Journal of Production Research, 2022: 1-17. doi: 10.1080/00207543.2022.2152126.

39. Zhou Y, Kumar A, Parkash C, Vashishtha G, Tang H, Xiang J. A novel entropy-based sparsity measure for prognosis of bearing defects and development of a sparsogram to select sensitive filtering band of an axial piston pump. Measurement, 2022, 203: 111997. doi: 10.1016/j.measurement.2022.111997.

40. Zhou Y, Kumar A, Gandhi C P, Vashishtha G, Tang H, Kundu P, Singh M, Xiang J. Discrete entropy-based health indicator and LSTM for the forecasting of bearing health. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2023, 45(2): 120. doi: 10.1007/s40430-023-04042-y.

41. Zhan X, Bai H, Yan H, Wang R, Guo C, Jia X. Diesel engine fault diagnosis method based on optimized VMD and improved CNN. Processes, 2022, 10(11): 2162. doi: 10.3390/pr10112162.