KSME

Springer

Journal of Mechanical Science and Technology 37 (6) 2023

Original Article
DOI 10.1007/s12206-023-0504-8

Keywords:
· Rolling bearing
· Fault evolution
· Status indicator
· Deep learning
· Wavelet analysis

Correspondence to:
Guo Chen
cgnuaacca@163.com

Citation:
Liu, X., Chen, G., Wei, X., Liu, Y., Wang, H. (2023). A rolling bearing fault evolution state indicator based on deep learning and its application. Journal of Mechanical Science and Technology 37 (6) (2023) ?~?.
http://doi.org/10.1007/s12206-023-0504-8

Received November 1st, 2022
Revised February 7th, 2023
Accepted February 24th, 2023

† Recommended by Editor
No-cheol Park

A rolling bearing fault evolution state indicator based on deep learning and its application

Xiyang Liu[1], Guo Chen[2], Xunkai Wei[3], Yaobin Liu[2] and Hao Wang[3]

[1]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China, [2]College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Liyang 213300, China, [3]Beijing Aeronautical Technology Research Center, Beijing 100076, China

Abstract Aiming at the limitation of early fault warning and the diagnosis of aero-engine main bearing when there are only normal operation data, a rolling bearing fault evolution state indicator based on deep convolutional neural network (CNN) and wavelet analysis was proposed. To be specific, firstly, the wavelet band envelope method was adopted to identify the early fault evolution process, and the feature distance between the degraded data and the normal ones was extracted by using deep CNN to develop the evolution state indicator. Then, the evolution stages were divided by using unsupervised clustering method. Finally, the remaining useful life (RUL) was predicted based on particle filter (PF). Three different groups of whole life cycle data of rolling bearings under various working conditions were used to prove the feasibility of the indicator. The results show that the wavelet-CNN features of completely different fault data show similar evolution trends, and the normalization of warning threshold can be realized based on the train labels. In conclusion, the results are of great significance for the early fault evolution monitoring, condition evaluation and remaining useful life prediction of rolling bearings without the absence of fault samples under actual aeroengine operation.

1. Introduction

Data driven is one of the most widely used methods for the condition monitoring and diagnosis of rolling bearings of aero-engine. It is popular to use the monitoring data during bearing operation to develop the condition indicator, determine the failure threshold, establish the evolution monitoring model, and realize the condition evaluation of rolling bearing [1-3]. However, there are currently some key points and difficulties in the researches based on this method, such as the inaccessible real fault samples in actual operation, the poor generalization ability of diagnostic models, the lack of scientific standards in defining failure thresholds, and the limitation in identifying early weak faults.

A series of related researches have been conducted to solve the main problems, such as the fault detection based only on normal samples, the definition of failure threshold, the prediction of remaining useful life, etc. Lin et al. [4] proposed a rolling bearing fault detection method based on hypersphere optimization support vector data description, improving the spatial distribution of feature vectors through hypersphere optimization, and realizing bearing fault detection based only on normal data with high accuracy. Islam et al. [5] defined the health index based on the defect degree of rolling bearings, and applied a class of least squares support vector machine models assisted by Bayesian reasoning (Bayesian-OCLSSVM) for anomaly detection to determine the prediction starting point and fault threshold. Luo et al. [6] proposed a method based on continuous wavelet transform (CWT) and convolutional automatic encoder (CAE) to extract deep features related to bearing degradation, and used self-organizing mapping (SOM) method to obtain degradation indicators. Zhou et al. [7] developed unsupervised health indicators (HI) by using Gaussian mixture model (GMM) and Kullback Leibler divergence (KLD), and achieved continuous prediction of unknown HI based on a new reinforced memory gated re-

cursive unit (RMGRU) network. In the optimization of feature extraction model based on rotary machines, Iqbal and Madan [8] proposed a convolution neural network model with vibration and acoustic signature, which realized the early real-time fault diagnosis of computer numerical control machine. The proposed method has obtained very high classification accuracy, and has prominent advantages over traditional machine learning models. Choudhary et al. [9] presented a non-invasive thermal image-based method for bearing fault diagnosis based on CNN model, which automatically identified the faults and attained 99.80 % classification accuracy and outperformed ANN.

However, in the existing studies, it is still difficult to define the alarm threshold of rolling bearing, and apply the model or threshold based on a specific data set well to other bearing data. Though there are sufficient methods for rolling bearing fault diagnosis, the researches of early warning technology of rolling bearing are relatively rare. Therefore, it is crucial in early warning, status evaluation and life prediction of rolling bearing to build a suitable health status indicator which can be used to timely detect the spalling fault and evolution status of rolling bearing in different stages, that is, to extract fault features submerged by a large amount of noise from vibration signals.

Deep learning method, receiving extensive attention at present, exhibits great advantages in the adaptive extraction of deep features, which can be used to unify failure thresholds of different bearings based on training labels. In this case, a rolling bearing fault evolution state indicator based on deep learning is proposed, which is obtained by training a deep convolution neural network with normal stage data of rolling bearing. For the vibration signal of rolling bearing in the healthy state, the status indicator can be normalized to the value of about 0.5. As bearing fault progresses, the value of the status indicator increases monotonically. When the index reaches certain value, similar between different bearing data, the degradation or failure of bearing can be determined. Finally, through the features reflected by the status indicator, the remaining useful life of rolling bearing is predicted by using the particle filter method.

2. Basic principle and process of fault evolution state indication and RUL prediction of rolling bearings

Firstly, the detail envelope signal of rolling bearing is obtained based on wavelet transform and autocorrelation analysis. The high-frequency detail signals in the normal stage are selected and input into the convolutional neural network for training, and the corresponding evolution feature extraction model is obtained. Then, the whole life cycle data are input into the trained network. The evolution characteristic curve is extracted adaptively based on the feature distance between the degradation sample and the normal one. After that, the fault evolution stages are divided, and the thresholds are determined by using unsupervised clustering method. Rolling bearings of different

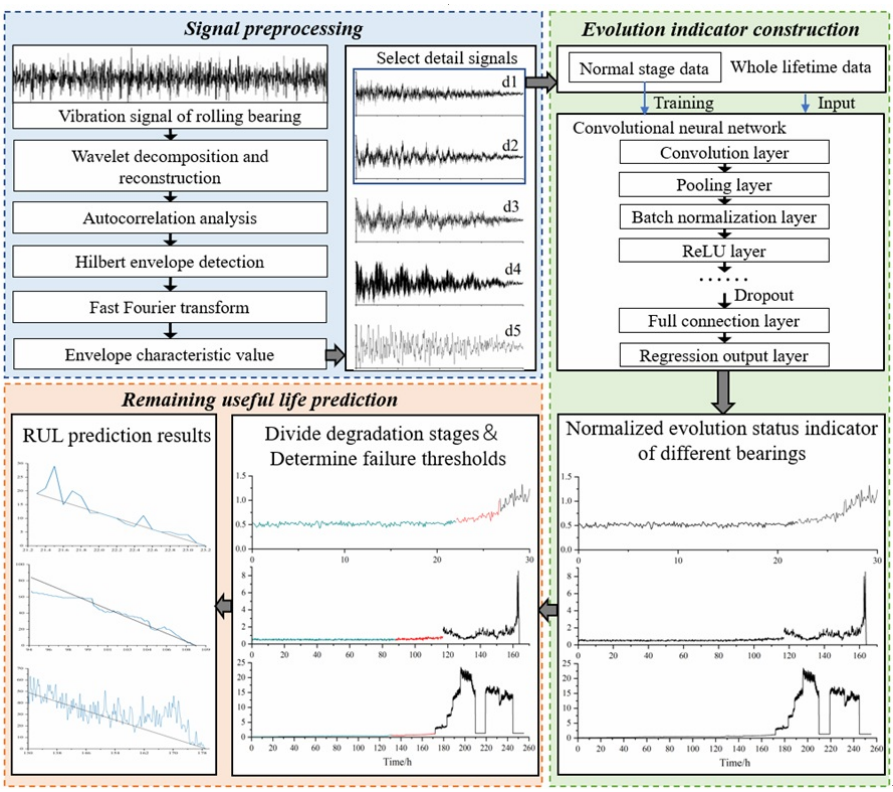


Fig. 1. Fault evolution state indication and RUL prediction of rolling bearings.

types and under diversified working conditions show similar evolution characteristic curves and highly consistent degradation and failure thresholds. Using particle filter jointly, the RUL of rolling bearings are gradually tracked and predicted. At last, the feasibility of the method is verified based on three groups of experimental data. The process of fault evolution state indication and RUL prediction of rolling bearings is shown in Fig. 1. The main innovation points involved in this paper are as follows:

1) The deep CNN model is trained based on the normal operation data of rolling bearing only, which better suits the actual situation. With the gradual input of the fault evolution data into the model, the feature value of the model output rises monotonically, which is regarded as the bearing fault evolution state indicator.

2) The features extracted from CNN model of bearing data of different types and under diversified working conditions show similar evolution trends, and the high consistency of alarm threshold can be achieved based on training labels, thereby providing a reliable basis for general diagnosis between different rolling bearings.

3) The high-frequency detail signal based on wavelet envelope analysis is more sensitive to the vibration of early faults. Therefore, the combination of wavelet-CNN and particle filter for rolling bearing RUL prediction is of great significance for the early evolution monitoring and condition evaluation of rolling bearing under the new working condition with insufficient fault samples.

3. Introduction of datasets

Since it is difficult to obtain the real rolling bearing fault data of aero engine, the data from different test rigs similar to aero-engine working principles are collected and analyzed to develop a unified threshold state indicator and realize early fault diagnosis and RUL prediction. Three groups of data from rolling bearing test rigs are used in this paper. And the bearing fault data used in Test 1 is from the intelligent maintenance systems (IMS), University of Cincinnati, USA, while the data of tests 2 and 3 are from the Intelligent Diagnosis and Expert System Research Office (IDES), Nanjing University of Aeronautics and Astronautics, China.

3.1 Intelligent maintenance systems (IMS) dataset

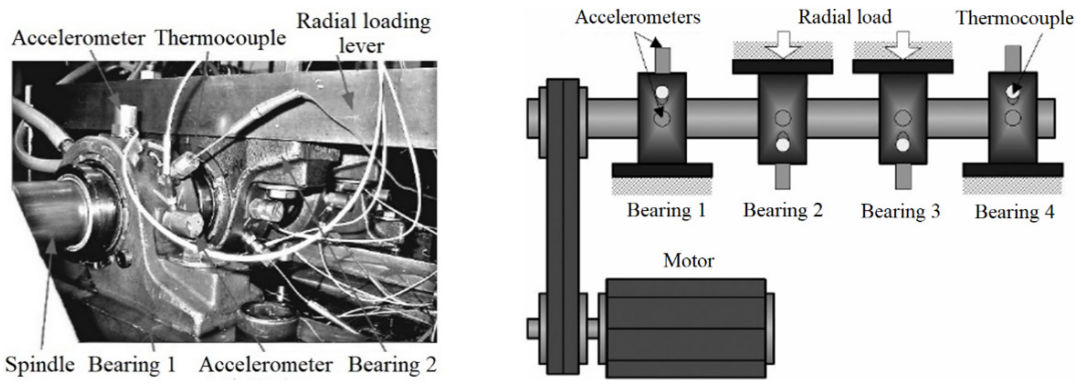
The rolling bearing fatigue test rig is shown in Fig. 2. The rolling bearing used in the test is Rexnord ZA 2115 (made by Regal Rexnord Ltd, USA), of which the model parameters are shown in Table 1. Four identical bearings were used during the test with the test speed of 2000 rpm. Each bearing was under the load of 26.67 kN and sufficient lubrication. The sampling frequency was 20480 Hz, the data were stored every 10 minutes, and the length of each sample was 20480. The effective duration of the test was 163 hours, and finally the no. 3 bearing received a single-point fault in its outer ring. Fig. 3. shows the appearance of the fault bearing.

3.2 Intelligent diagnosis and expert system (IDES) dataset

Tests 2 and 3 were carried out on the HRB 6026 and the BMD 6009 single-row deep-groove ball rolling bearings by using the AB-LT1A bearing life enhancement testing machine. The main parameters of the two bearings are shown in Table 2, and the test rig and the loading diagram are shown in Fig. 4. For the test on the HRB 6206 bearings, the constant speed was 11500 rpm. Each of the four bearings was under the load of 6.25 kN and sufficient lubrication. The test sampling frequency was 32k Hz, and the samples were stored every 10 minutes. After the operation of 30 hours, the test was terminated due to excessive vibration. The spalling fault occurred on

Table 1. Main parameters of ZA-2115 rolling bearing.

Model	Pitch diameter	Contact angle	Ball diameter	Number of balls
ZA-2115	71.5 mm	15.17°	8.4 mm	16



(a) Test rig (b) Loading diagram

Fig. 2. Rolling bearing fatigue test rig and loading diagram.

Table 2. Parameters of HRB 6206 and BMD 6009 rolling bearings.

Model	Bore diameter	Outer diameter	Width	Diameter of balls	Pitch diameter	Number of balls	Contact angle
HRB 6206	30 mm	62 mm	16 mm	9.5 mm	46 mm	9	0°
BMD 6009	45 mm	75 mm	16 mm	8.7 mm	60 mm	12	0°

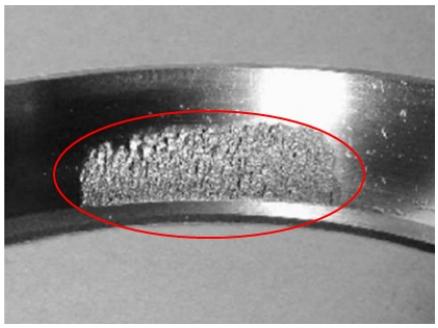
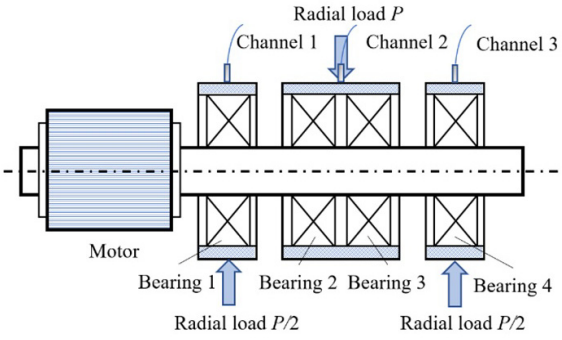


Fig. 3. Outer ring damage of ZA-2115 bearing.

(a) Test rig



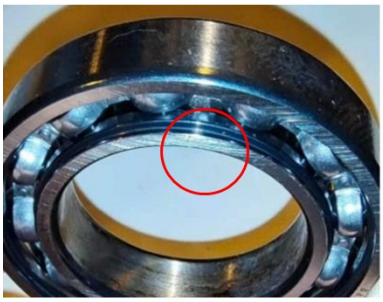
(b) Loading diagram

Fig. 4. ABLT-1A test rig and loading diagram.

the inner ring, as shown in Fig. 5(a).

For the test on the BMD 6009 bearings, the constant speed was 12000 rpm. Each of the four bearings bears a radial load of 5.1 kN and an axial load of 1.2 kN at the same time. The test sampling frequency was 32 kHz, and the samples were stored every 2.5 minutes. After the operation of about 240 hours, the test was terminated due to excessive vibration. The spalling

(a) HRB 6026



(b) BMD 6009

Fig. 5. Inner ring damages of HRB 6206 and BMD 6009 deep groove ball bearings.

fault occurred on the inner ring, as shown in Fig. 5(b).

4. Development of deep learning condition indicator (DLCI) of rolling bearing

4.1 Extraction of high-frequency envelope signal of rolling bearing based on wavelet analysis

The fault evolution process of rolling bearing shows certain regularity in high-frequency band (above 20 kHz), medium-frequency band (1 kHz-20 kHz) and low-frequency band (below 1 kHz). Taking the life cycle data of ZA-2115 bearing as an example, Fig. 6 shows the four typical stages of rolling bearing fault evolution in frequency domain.

It can be seen that the fault evolution of rolling bearing is an energy migration process from high frequency to low frequency, covering the whole frequency band at last. Early faults are usually reflected more in the high frequency band. Wavelet analysis [10] can be used to obtain the rolling bearing fault signal energy in each energy band through wavelet band decomposition, envelope analysis and autocorrelation analysis. It is also available to reflect the evolution process of the fault, especially the detail signal in the high frequency band which is more sensitive to the evolution of early faults. In this paper, to

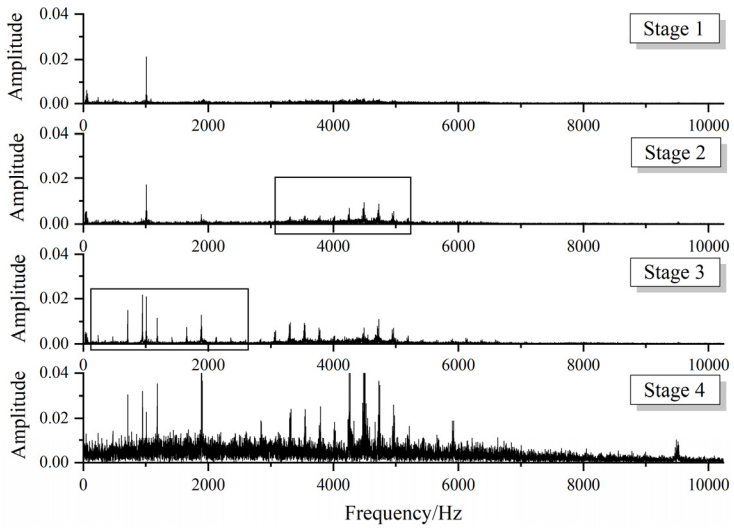


Fig. 6. Four stages of rolling bearing fault evolution in frequency domain.

obtain high-frequency detail signals available to reflect the evolution of early faults, the db8 wavelet function and 5-layer decomposition are selected through comparative tests and based on historical experience. The steps of fault evolution monitoring based on wavelet analysis are as follows:

Step 1: Discrete dyadic wavelet transform is carried out on the collected original signal. 5-layer wavelet decomposition is conducted with db8 wavelet as the base, and the detail signal d_i (i = 1, 2, …, 5) is obtained.

Step 2: To eliminate the interference of random signals, the autocorrelation noise reduction method is used to reduce the noise of the band decomposition signal, and suppress the aperiodic components in the detail signal. Then, the Hilbert transform is adopted to obtain the time domain waveform W_i (i = 1, 2, …, 5) of the envelope of detail signal d_i, which contains the information from high frequency to low frequency of the rolling bearing vibration signal, respectively.

Step 3: For comparison, based on different time point t_j (j = 1, 2, 3, …, N), the time series W_{ij} of band envelope energy is calculated according to step 1 and step 2, of which the RMS value is calculated respectively, and the RMS evolution trend of different detail signals of the whole life time is obtained.

Step 4: For different time point t_j (j = 1, 2, 3, …, N), the time series of band envelope energy is calculated according to step 1 and step 2, which is input into the CNN model to obtain the CNN feature evolution trend of different detail signals of the whole life time.

4.2 Feature extractor based on CNN

4.2.1 Vibration time series and its pretreatment

For the rolling bearing evolution data based on vibration time series signals, one-dimensional network, such as GRU or LSTM is generally used to extract time memory correlation characteristics [11]. However, relevant studies [12, 13] suggested that in the field of fault diagnosis, one-dimensional network training is both difficult and imprecise, and the existing pre-training network is insufficient, which is not as universal as the two-dimensional convolutional network. The low-dimension feature extractor suffers certain limitation in fully mining the characteristics of original information. The conversion of one-dimensional time series signals into two-dimensional maps can be used to extract high-dimension features of rolling bearing signals, and fully leverage the advantages of CNN network [14].

To retain the fault evolution trend contained in the original data to the greatest extent, and avoid the weakening or loss of fault impact components due to the utilization of the normalization processing method, the amplitude of the original time sequence signal is directly converted into pixel information to improve the feature dimension, and retain the time memory characteristics in the rolling bearing vibration sequence. The specific conversion expression is as follows:

$$P(i,j) = A(i*N - N + j)$$
$$i \in [1,M], j \in [1,N] \tag{1}$$

where P refers to the transformed two dimension matrix, M represents the length of the matrix, N stands for the width of the matrix, A is the amplitude of one dimension time domain signal, and $P(i,j)$ denotes the intensity of pixels at point (i,j) in an $N \times M$ image.

When constructing the matrix graph, the signals are sampled at equal pulse intervals, indicating that the attention area of each sample is relatively fixed. During the convolution feature extraction, the model will pay more attention to and learn the peak value and nearby parts that can best reflect the bearing vibration characteristics, as shown in Fig. 7.

4.2.2 CNN structure and parameters

As a multilayer perceptron neural network, CNN can be used to obtain abundant data information by using shared weights

and increasing the depth of the network [15]. The collocation of convolution-pooling-activation layer is a basic nonlinear transformation module, which is the most commonly used, simple and efficient CNN network construction method. Experiments show that simple CNN model can be used for the feature extraction required in this paper, therefore, the design of complex network structures is not necessary. By stacking different numbers of nonlinear transformation modules and comparing their performance, the deep CNN model is determined to be composed of the following parts: the input layer, 3 convolution-pooling-ReLU layers, the batch normalization (BN) layer, the dropout layer, the full connection (FC) layer and the regression layer. Other relevant parameters of CNN are determined through the adjustment and comparison of the accuracy and speed of the model. The parameters of the CNN model are shown in Table 3.

The specific working principles of each layer of the network are as follows:

Table 3. Parameters of CNN model.

Structure	Parameters	Output
Input layer	64×32×1	64×32×1
Pooling layer	2×2	32×16×1
Convolutional layer	3×3, 16	32×16×16
Pooling layer	2×2	16×8×16
Convolutional layer	3×3, 32	16×8×32
Pooling layer	2×2	8×4×32
Convolutional layer	3×3, 32	8×4×32
Full connection layer	1	1×1×1
Regression layer	1	1

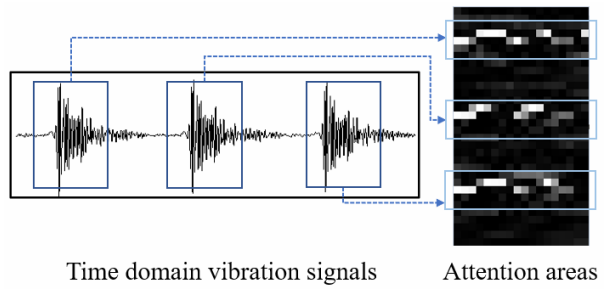


Fig. 7. Attention areas corresponding to vibration signals.

1) Convolution layer and pooling layer. The random parameter convolution kernel ω glides through the whole picture on the image, then a series of feature maps $f_i = \omega_i * X + b_i$ are obtained on a convolution layer. After that, the condensed feature maps are formed by further extracting the dimension reduction of features through the maximum pooling layer.

2) Dropout layer. A dropout layer is connected before the full connection layer which discards the neuron node with a given probability p (p = 0.2) to reduce the complex coadaptation relationship between neurons, and improve the robustness of the network to the loss of specific neuron connections [8].

3) Batch normalization (BN) layer. The batch normalization layers are added to both sides of the core feature extractor (convolutional-pooling-ReLU layer) to prevent the gradient from disappearing or exploding, and speed up the training.

4) Full connection layer. The full connection layer is set after the second BN layer to synthesize the feature extraction results, and map them to the sample tag space. The output size is set as 1 in this paper, which means that when a sample is input into the CNN network, the network will output a value, representing the evolution index of the rolling bearing at that time.

5) Regression layer. When the full connection layer changes the two-dimensional feature into one dimensional one, the feature is output directly through regression function without other mapping for reducing the loss of features. At this time, the regression layer can be regarded as a special case of fully-connected layer whose coefficient is 1, and the offset is 0. The loss function in this paper is set to root mean square error.

The acquisition process of rolling bearing fault evolution status indicator is shown in Fig. 8.

4.2.3 Extraction results of DLCI

A new method for extracting the evolution index of rolling bearings based on deep CNN is proposed. The normal bearing signals are supervised and trained by using the designed CNN network. After that, the whole life cycle data are input into the network in turn to obtain the corresponding bearing deterioration characteristics at each time point. Generally, the difference of the feature value between the fault sample and the normal one increases with the gradual degradation of bearing. In this paper, the increasing feature distance between normal and degeneration samples is used to develop the rolling bearing fault evolution state indicator.

The extraction process of rolling bearing fault evolution state

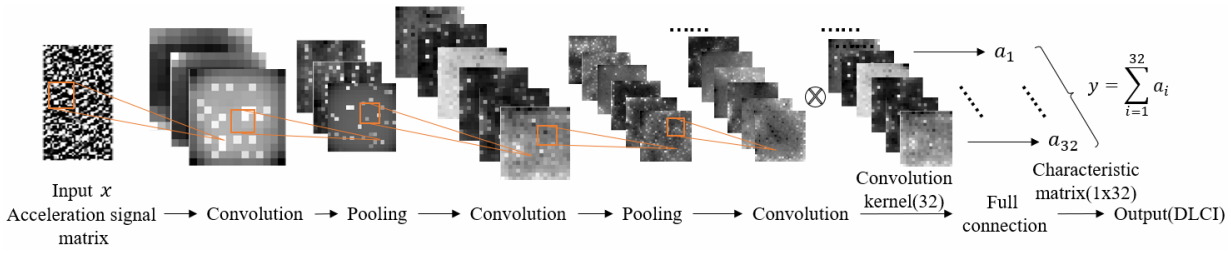


Fig. 8. Acquisition process of rolling bearing fault evolution status indicator based on CNN network.

index includes the following steps:

1) The vibration signal of the whole lifetime of rolling bearing is collected, which includes N time points, and each time point contains X vibration points.

2) The first N' time points of normal data are selected, and divided into training and test set. Considering both the accuracy and the convergence speed of the network, the training labels are finally set to 0.5 through comparative testing. The vibration acceleration signal at each time point is divided into X /2048 matrix diagrams for training.

3) The training set is input into the deep CNN model for training, and the training rounds are set to 300; the initial learning rate is set to 0.01, which is decreased by 10 times every 100 rounds; and the training batch size is set to 128. The principle of parameters selection is to converge the network and minimize the loss value. After parameters selection, the model is trained to minimize the loss on the test set.

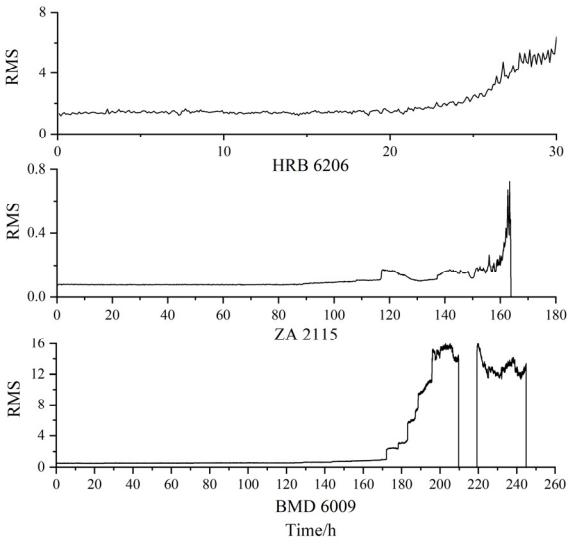
4) All samples in the whole life cycle are imported into the trained network, and each sample x outputs a feature value y, which corresponds to the deterioration index of rolling bearing. N feature values are obtained at last, which represent the degradation states of rolling bearings at N time points.

Going through the above steps, the whole lifetime evolution characteristic trends based on CNN of three kinds of rolling bearings are finally obtained. For comparison, the root mean square (RMS) index widely applied in the research of rolling bearing failure early warning and life prediction is selected at the same time, and its change trend in the whole life cycle is also calculated. The description of the RMS value is expressed as follows:

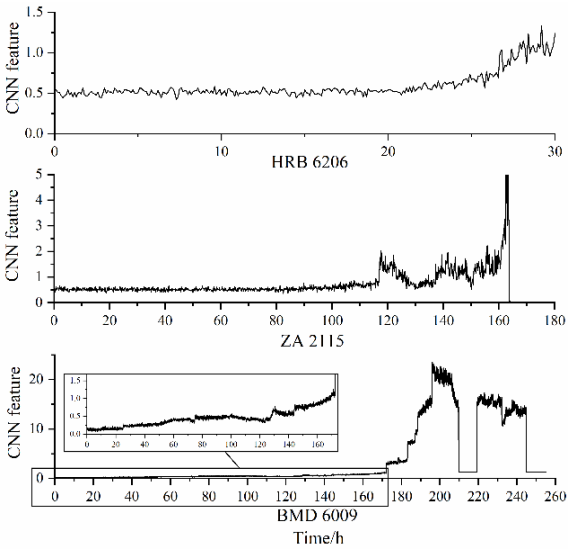
$$\mathrm{RMS}=\sqrt{\sum_{i=1}^{n}x_i^2/n}\ . \tag{2}$$

The evolution trend comparison of the extracted CNN characteristic values and the RMS values during the whole life cycle are shown in Fig. 9.

It can be seen from Fig. 9(a) that, for three rolling bearings of different types and under diversified working conditions, the RMS values of them in the normal stage are about 1.5, 0.1 and 0.5, respectively, which means it is impossible to determine the unified alarm threshold between different data according to RMS value. However, for the CNN values shown in Fig. 8(b), although the values increase to varying degrees in later failure stages based on different working conditions of bearings, the value in the normal stage is always normalized by CNN model to around 0.5, and there is a clear degradation trend when the value reaches about 1. This rule works simultaneously between the three groups of life cycle data of different types of bearings, under diversified working conditions and with various test rigs, which proves that the method can be used to realize fault alarm with consistent threshold only based on normal data, and provides a new idea for one-class classification and general diagnosis under variable working conditions.



(a) RMS



(b) CNN feature

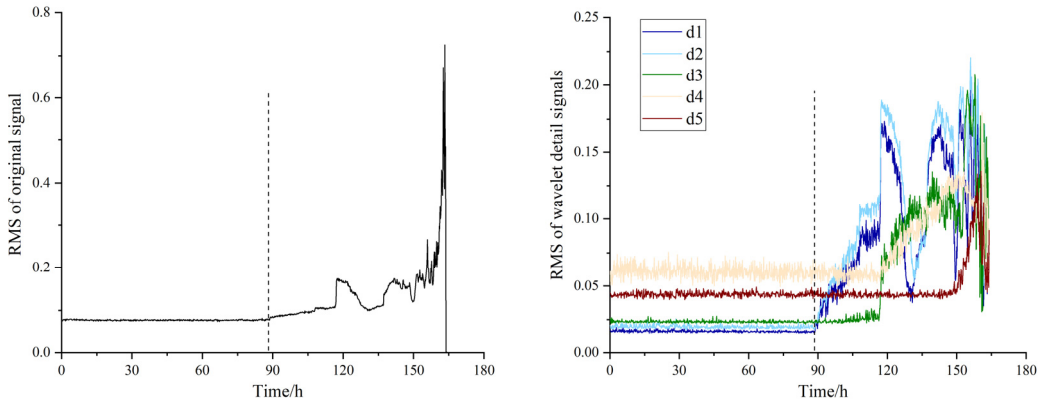
Fig. 9. Comparison of feature value evolution based on three groups of whole lifetime data.

4.3 DLCI construction based on wavelet-CNN

Going through the above steps, the RMS values and CNN feature values of three groups of rolling bearing data in whole frequency band and each detail frequency band are obtained, as shown from Figs. 10-15.

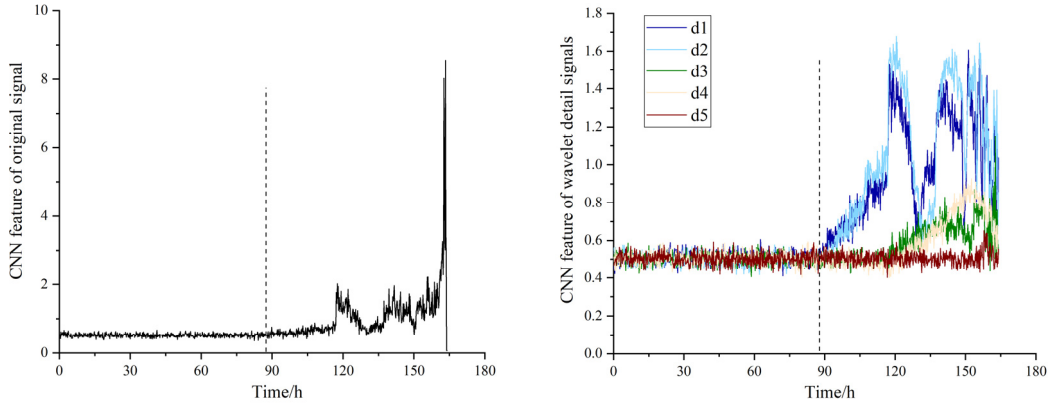
As can be seen from the figures:

1) The RMS values of different bearings in whole frequency band or in certain frequency range are distributed in different numerical dimensions in normal stage. For example, in Figs. 10(a), 12(a) and 14(a), the RMS values in whole frequency band are distributed somewhere around 0.08, 1.4 and 0.7, respectively in the normal stage. In the meantime, the RMS values of each detail signal, as shown in Figs. 10(b), 12(b) and



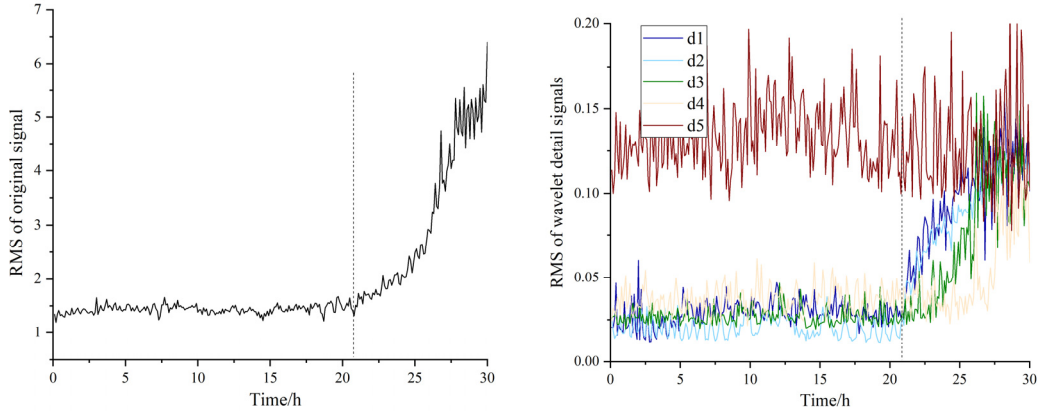
(a) RMS in whole frequency band (b) RMS in different frequency ranges

Fig. 10. RMS evolution trend of ZA 2115 bearing during its whole lifetime.



(a) CNN feature in whole frequency band (b) CNN feature in different frequency ranges

Fig. 11. CNN feature evolution trend of ZA 2115 bearing during its whole lifetime.



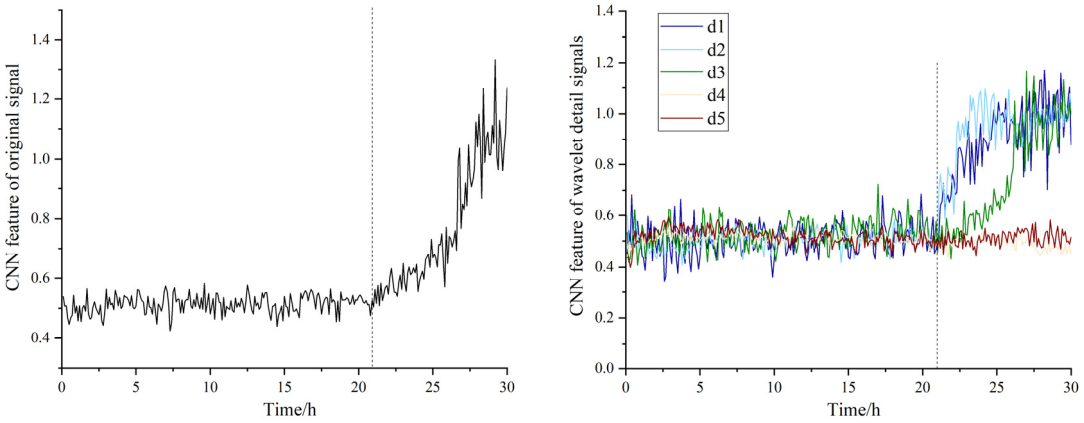
(a) RMS in whole frequency band (b) RMS in different frequency ranges

Fig. 12. RMS evolution trend of HRB 6206 bearing during its whole lifetime.

14(b) are also distributed discretely without a unified threshold.

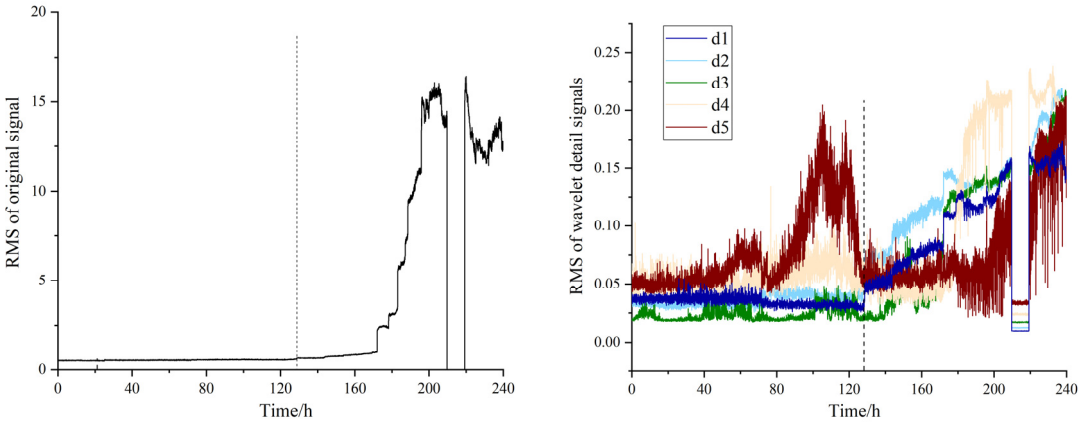
2) Differently from the description of paragraph 1, as can be seen from Figs. 11, 13 and 15, the values of CNN features in the normal stage are normalized to about 0.5 in both whole frequency band and the high to low frequency bands, and the evolution trends of wavelet detail signals of different bearings are mapped to the same dimension, which is of great significance for the development of general models and variable condition diagnosis.

3) From the comparison of (a) and (b) from Figs. 10-15, it can be seen that in the early stage of the faults, the RMS values and the CNN feature values in whole frequency band show



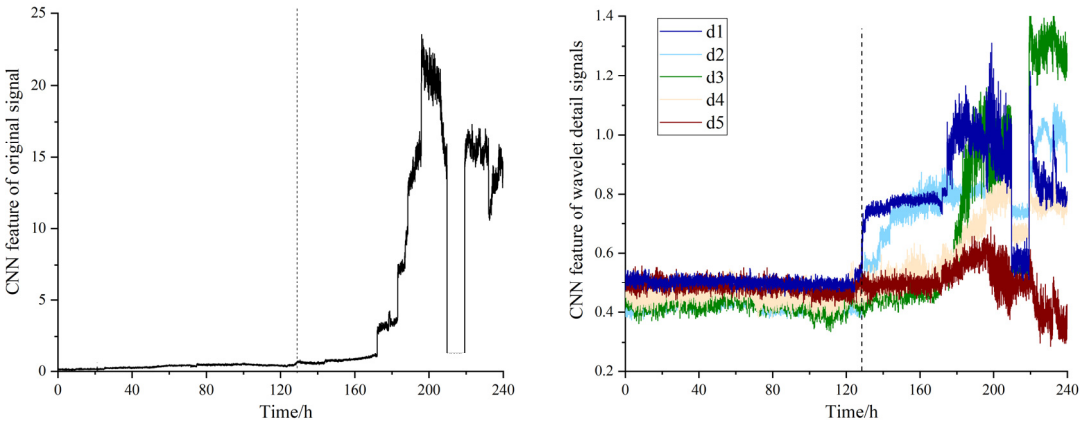
(a) CNN feature in whole frequency band (b) CNN feature in different frequency ranges

Fig. 13. CNN feature evolution trend of HRB 6206 bearing during its whole lifetime.



(a) RMS in whole frequency band (b) RMS in different frequency ranges

Fig. 14. RMS evolution trend of BMD 6009 bearing during its whole lifetime.



(a) CNN feature in whole frequency band (b) CNN feature in different frequency ranges

Fig. 15. CNN feature evolution trend of BMD 6009 bearing during its whole lifetime.

very weak upward trends only, which are difficult to monitor by the sensors. However, the upward trends are quite obvious for the detail signals in the high-frequency bands (d1 and d2), which facilitates the division of evolution stages. Therefore, in this paper, d1 and d2 high-frequency detail signals of bearings are selected as the feature indicators extracted by CNN, and the alarm and failure threshold are determined, thereby predicting the remaining useful life based on this.

5. Remaining useful life prediction of rolling bearings based on DLCI and PF

5.1 Determination of degradation and failure thresholds

Taking the advantages of the threshold normalization ability of CNN, the unified degradation thresholds and failure thresholds are conducive to the establishment of the model under different working conditions of rolling bearings. Before stage division and threshold determination, it is necessary to filter and denoise the characteristic curve to improve the accuracy of RUL prediction. The moving average filter is optimal to reduce random noise while retaining a sharp step, of which the formula is expressed as follows:

$$y[i]=\frac{1}{M}\sum_{j=0}^{M-1}x[i+j] \tag{3}$$

where $x[\ ]$ and $y[\ ]$ refer to the input and output signals, respectively, and M stands for the sliding window length which is set to 5 in this paper.

After smoothing the characteristic curves of the detail signals of d1 and d2, the evolution trends are divided into several stages to determine the unified degradation and failure threshold. Due to the large gap between normal and abnormal samples of rolling bearings, and the relatively low dimension of evolution index, the *k*-means unsupervised clustering algorithm is selected to automatically divide the evolution index during the whole life time into three parts, representing the three stages of the evolution of rolling bearings. The *k*-means clustering method is used to compute the shortest Euclidean distance between each sample point and the cluster center by iteration, which is used as the standard to cluster samples into their categories [16]. The calculation of *k*-means is expressed as follows:

$$\min\sum_{i=1}^{k}\sum_{x\in C_i}\left\|x-\mu_i\right\|_2^2 \tag{4}$$

where μ_i refers to the centroid of cluster partition C_i, and x represents the characteristic curve.

The rolling bearing evolution stage is adaptively divided into normal, degradation and failure stage by clustering. The values corresponding to the starting point of degradation and failure stage are respectively marked as the degradation and the failure threshold. Finally, the characteristic curve after smooth filtering and unsupervised evolution stage division is obtained, as shown in Fig. 16.

The results in Sec. 4.3 prove that the detail signals in the band with relatively high frequency can better reflect the early fault evolution. Therefore, the evolution stages are divided, and the RUL is predicted based on wavelet detail signals of d1 and d2. The following comparison includes the evolution stages divided by *k*-means clustering based on d1 or d2 detail signal and their mean values, and the determined degradation and failure thresholds and their corresponding moments, as shown in Table 4.

It can be concluded from Table 4:

1) The early degradation features are mainly concentrated in d1 and d2 wavelet detail signals, and single-detail signal is prone to produce large errors when determining the threshold. Therefore, in this paper, the mean value of d1 and d2 is taken as the evolution measurement index to predict the residual life of rolling bearing.

2) The degradation and failure thresholds of the features extracted by CNN are highly unified for the three bearings with large differences, and their early fault evolution ranges are mainly from DLCI value of 0.6 to 1.0. It is proved that the wavelet-CNN model developed in this paper can be used to well normalize the threshold. Then, the PF method is adopted to track and predict the RUL of the three groups of bearings.

5.2 Remaining useful life prediction based on particle filter

Based on the empirical degradation model, the statistical filtering method using particle filter (PF) algorithm to predict the RUL of machine element is widely used in practice because it can adapt to non-linear and non-Gaussian state prediction, and provide the uncertainty expression of prediction results [17, 18]. With the particle filter method, a set of discrete random particle sets can be utilized to approximate the probability density func-

Table 4. Thresholds of different detail signals of three bearings based on *k*-means clustering.

		HRB 6206 (whole life cycle with 300 time points)		ZA 2115 (whole life cycle with 984 time points)		BMD 6009 (whole life cycle with 6380 time points)	
		Beginning of degradation	Beginning of failure	Beginning of degradation	Beginning of failure	Beginning of degradation	Beginning of failure
d1	Threshold	**0.6146**	**0.9837**	**0.6094**	**1.0335**	**0.6297**	**0.9714**
	Time point	213	251	552	700	3219	4439
d2	Threshold	**0.5482**	**0.9621**	**0.6141**	**1.2108**	**0.5623**	**0.8827**
	Time point	208	230	573	701	3248	4938
$\frac{(d1+d2)}{2}$	Threshold	**0.6177**	**0.9316**	**0.6219**	**1.1470**	**0.5992**	**0.9408**
	Time point	212	237	572	701	3241	4630

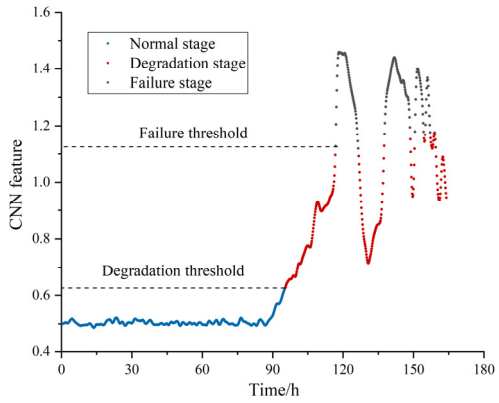


Fig. 16. The smoothed detail signal divided into 3 evolution stages after clustering.

tion of the system, and the integral operation can be replaced with the sample mean to obtain the minimum variance estimate of the state, thereby achieving the prediction with inconsistent data distribution. In addition, particle filter is not affected by system noise and measurement noise, which has good prediction function and high accuracy for non-Gaussian non-linear systems, such as rolling bearing evolution trend. Each RUL prediction based on particle filter algorithm calculated in this paper is within 1 minute, while the prediction results based on PF are in hours, in that case, the real-time prediction is enough to meet the monitoring needs of aeroengines.

Assuming that the feature curve of bearing follows the formula $y = a \times e^{bK} + c \times e^{dK}$ where a, b, c and d contain Gaussian white noise, and K refers to the time step, then the state vector is expressed as follows:

$$x(K) = \left[a(K), b(K), c(K), d(K)\right]. \tag{5}$$

The equation of state and observation are as follows:

$$\begin{cases} a(K+1) = a(K) + w_a(K), w_a \sim N(0, \sigma_a) \\ b(K+1) = b(K) + w_b(K), w_b \sim N(0, \sigma_b) \\ c(K+1) = c(K) + w_c(K), w_c \sim N(0, \sigma_c) \\ d(K+1) = d(K) + w_d(K), w_d \sim N(0, \sigma_d) \end{cases} \tag{6}$$

$$\begin{aligned} y(K) &= a(K) \times e^{b(K) \times K} + c(K) \times e^{d(K) \times K} \\ &+ v(K), v(K) \sim N(0, \sigma_v). \end{aligned} \tag{7}$$

The RUL prediction of rolling bearing based on PF algorithm includes 6 steps:

1) After collecting the vibration signal of bearing, the degradation index representing the condition of rolling bearing is extracted by using CNN model.

2) The degradation threshold and the failure threshold are determined. The time corresponding to the degradation threshold is taken as the beginning moment of the RUL prediction, and the moment corresponding to the failure threshold is taken as the termination of prediction.

3) Considering that the constructed equation should conform to the degradation trend of rolling bearings with high fitting accuracy and without over-fitting, a four-parameter double exponential model is established, and the n feature points before the degradation threshold are fitted to get the initial parameters of a_0, b_0, c_0 and d_0 with the least squares curve.

4) The feature value V_{K+1} at time $K+1$ is predicted in advance by using PF method. In the case that V_{K+1} has not reached the failure threshold, continue to calculate the feature value V_{K+2} until it exceeds the failure threshold, then, finish the prediction and record the present time K_r at which RUL = $K_r - K$.

5) Continue to collect the bearing data, extract the feature index at time $K+1$ by using CNN, and take the n features before the time; repeat step 3 and step 4 to get the remaining useful life at time $K+1$.

6) When the degradation index reaches the failure threshold, finish the prediction, and the RUL curve in degradation stage is obtained. Compared it with the actual RUL curve to get the relative error of prediction.

The prediction error can be described as root mean square error, expressed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{m}\left(\mathrm{RUL}_p(i) - \mathrm{RUL}_r(i)\right)^2} \tag{8}$$

where RUL_p refers to the predicted RUL value of rolling bearing, RUL_r represents the real RUL value, and m stands for the number of predicted time points.

5.3 Comparison of features of wavelet detail and whole frequency band of CNN

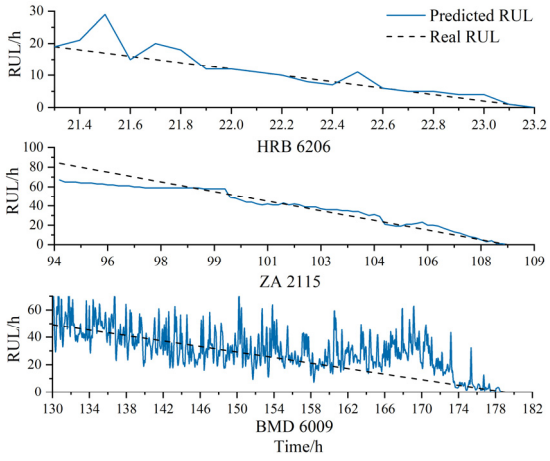
5.3.1 Comparison between CNN detail features and CNN whole frequency band features

According to the above process, the progressive RUL tracking prediction is carried out for the bearing degradation stage based on wavelet-CNN detail feature, and the RUL prediction results of the three groups of data are obtained, as shown in Fig. 17(a). To compare the difference of RUL prediction results based on the evolution indexes of high-frequency detail feature and whole frequency feature, in this paper, the RUL is tracked and predicted according to the evolution trend of CNN whole frequency band features, as shown in Figs. 11(a), 13(a) and 15(a). And the degradation and failure threshold is determined by smoothing and clustering. Besides, the RUL prediction results of three groups of data are finally obtained, as shown in Fig. 17(b).

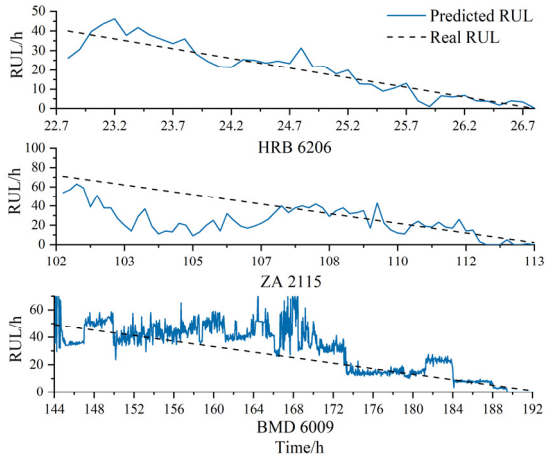
From the comparison between Figs. 17(a) and (b), it can be seen that due to the indistinctive change in early fault stage, the degradation and failure thresholds of CNN whole frequency band features determined based on unsupervised clustering lags behind the wavelet-CNN detailed features, resulting in the overall delay in the prediction of degradation stage. To quantify the difference between the prediction accuracy of details and whole frequency features, the error values of the two features

Table 5. RUL accuracies predicted by PF method based on CNN and wavelet-CNN.

	HRB 6206 (300 time points)		ZA 2115 (984 time points)		BMD 6009 (6380 time points)	
Feature extraction method	CNN	**Wavelet-CNN**	CNN	**Wavelet-CNN**	CNN	**Wavelet-CNN**
Degradation start point	227	**212**	602	**572**	3594	**3241**
Failure start point	267	**231**	655	**650**	4782	**4470**
RMSE of fitting curve	0.1053	**0.0683**	0.0177	**0.0163**	0.1081	**0.0925**
RMSE of prediction results	7.45	**3.31**	20.49	**6.92**	290.48	**327.59**
Normalized RMSE of prediction results	2.48 %	**1.10 %**	2.08 %	**0.73 %**	4.55 %	**5.13 %**



(a) Wavelet-CNN detail features



(b) CNN whole frequency band features

Fig. 17. RUL prediction results of three groups of data degradation stages based on CNN features.

in the particle filter curve fitting stage, the error values of RUL predicted results, and the error values of RUL predicted results normalized according to the whole life points are calculated, respectively, as shown in Table 5.

As can be seen from Table 5:

1) The CNN features based on d1 and d2 wavelet detail signals can be used to predict the degradation and failure earlier than the whole frequency band CNN features, which proves that the high-frequency detail signals are more sensitive to early degradation, and more conducive to early fault warning.

2) The prediction accuracy of wavelet-CNN for HRB 6206 and ZA 2115 bearings are significantly higher than that of whole frequency CNN. For BMD 6009, the prediction accuracy of whole frequency CNN is basically equivalent to that of wavelet-CNN.

3) Compared with the other two groups of data, the normalized error of BMD 6009 bearing prediction results is significantly higher, which is due to longer test time and more prediction steps, resulting in more iterations of PF prediction, and relatively larger error accumulation.

5.3.2 Comparison of CNN wavelet detail features and RMS wavelet detail features

It is proved with the above comparison that compared with the whole frequency feature, the high-frequency detail feature can be used to perceive the emergence and development of early damage in a more timely manner. The prediction accuracy of CNN features and traditional RMS features based on wavelet are compared and analyzed below. According to the evolution trend of high-frequency detail features of RMS, as shown in Figs. 10(b), 12(b) and 14(b), the particle filter RUL tracking prediction is carried out, and the degradation and failure thresholds are also determined by smoothing and clustering. Finally, the RUL prediction results of three groups of data are obtained, and compared with the CNN detail features obtained previously, as shown in Fig. 18. The specific comparison between the prediction results of CNN details and RMS details is shown in Table 6.

From Fig. 18 and Table 6, it can be concluded that:

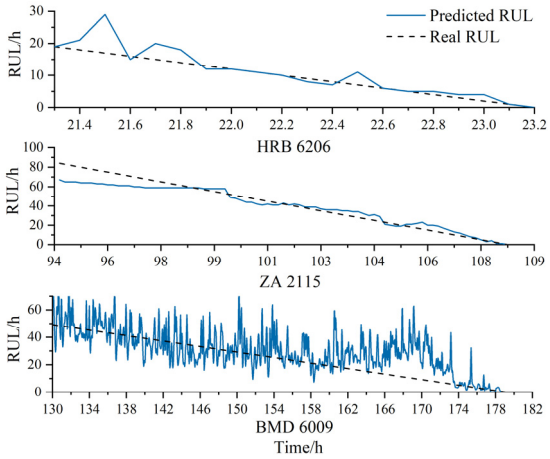
1) Compared with the wavelet-RMS detail features, the wavelet-CNN detail features can better realize the threshold normalization under different working conditions.

2) Seen from the comparison results of the prediction accuracy of three rolling bearing data, the RUL prediction errors based on wavelet-CNN detail features are significantly lower, of which the prediction accuracies are noticeably higher than the wavelet-RMS features.

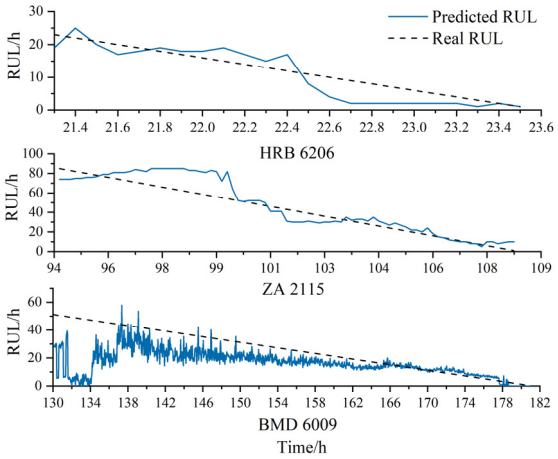
3) The prediction errors of wavelet-RMS values of BMD 6009 data are relatively close to that of wavelet-CNN feature values. However, from the prediction results of BMD 6009, as shown in Fig. 18(b), it can be seen that the wavelet-RMS values are not accurate in predicting early faults, whose overall accuracy is lower than that of wavelet-CNN feature values, as shown in Fig. 18(a).

Table 6. Accuracies of RUL predicted by PF method based on wavelet-RMS and wavelet-CNN.

	HRB 6206 (300 time points)		ZA 2115 (984 time points)		BMD 6009 (6380 time points)	
Feature extraction method	Wavelet-RMS	Wavelet-CNN	Wavelet-RMS	Wavelet-CNN	Wavelet-RMS	Wavelet-CNN
The value corresponding to the degradation start point	0.0410	0.6177	0.0393	0.6219	0.0603	0.5992
The value corresponding to the failure start point	0.0804	0.9316	0.0911	1.1470	0.1301	0.9408
RMSE of fitting curve	0.0101	0.0683	0.0272	0.0163	0.0312	0.0925
RMSE of prediction results	3.52	3.31	10.87	6.92	374.94	327.59
Normalized RMSE of prediction results	1.17 %	1.10 %	1.10 %	0.73 %	5.88 %	5.13 %



(a) Wavelet-CNN detail features



(b) Wavelet-RMS detail features

Fig. 18. RUL prediction results of three groups of data degradation stages based on detail features.

5.4 Comparison with other RUL prediction models

For the combined model of CNN and PF, CNN plays the role of degradation feature extraction and dimension reduction, while PF model is responsible for RUL prediction based on degradation features. However, other deep-learning models, such as LSTM or GRU network, can be used to extract the time memory correlation characteristics from the rolling bearing evolution data based on one-dimensional vibration time series signal, and independently complete the RUL prediction task. To compare with the model developed in this paper, taking the evolution trend of high-frequency wavelet details of HRB 6206 bearing as an example, the one-dimensional time series signals are trained in the normal stage (0-20 h) through LSTM and GRU networks, respectively, and the evolution trend in the degradation and failure stages (20-30 h) is predicted. Considering that the prediction results of the whole life stage may not reach high accuracy by normal data training only, a small amount of data of the early degradation stage (0-21 h) is introduced besides normal data, and the training is conducted based on two models to predict the remaining stages (21-30 h). The prediction results of the evolution trends are shown in Fig. 19.

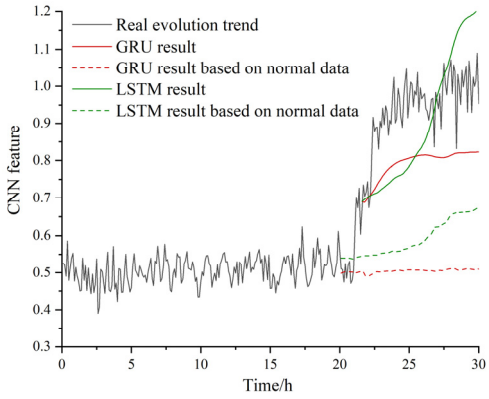


Fig. 19. Prediction results of evolution trend of rolling bearing based on GRU and LSTM.

It can be seen from Fig. 19 that due to the huge difference in data distribution between normal stage and degeneration stage, the LSTM or GRU regression model fails to predict the evolution trend of abnormal data after the training based on normal data only. After introducing a small amount of early degradation fault data in the training stage, the two models exhibit certain prediction ability for the degradation trend, but the LSTM model only reached the expected failure threshold in about the 27th hour, which is later than the 21 hour 12 minute of the CNN-PF model. In addition, the trend predicted by using GRU model can never reach the expected failure threshold. The

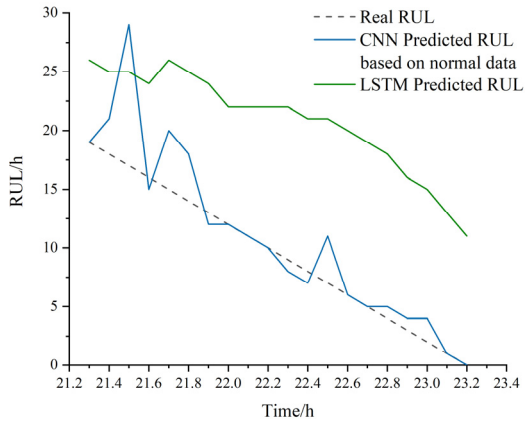


Fig. 20. RULs predicted by CNN-PF and LSTM.

comparison of RUL prediction results based on LSTM model after introducing degradation data and CNN-PF model is shown in Fig. 20.

It can be seen that despite the introduction of a small amount of fault data, the prediction ability of LSTM is still lower than that of CNN-PF. Therefore, it can be concluded that the CNN-PF model has unique advantages in the construction of evolution trend and prediction of RUL based on normal samples only.

6. Conclusions

In this paper, a general early warning model of rolling bearing under variable working conditions was proposed based on wavelet envelope spectrum analysis and deep convolution neural network. The model can be trained based on normal data of rolling bearing only, and the feature indicator was constructed according to the feature distance of normal and degraded samples, which has similar warning threshold for rolling bearings under different working conditions. Through the RUL prediction results of multiple groups of whole lifetime data based on particle filter, this method has been proved to have the following advantages:

1) The deep convolution neural network can be trained based on normal data only, whose threshold can be normalized under different working conditions of rolling bearings based on labels. For the evolution features of the three different bearings under various working conditions in this paper, the degradation and failure thresholds based on CNN values are all distributed somewhere around 0.6 and 1.0, which is of great significance for rolling bearing evolution monitoring and RUL prediction in the absence of fault samples with new test rigs.

2) Compared with the traditional RMS value, CNN features can realize the threshold normalization under different working conditions with lower RUL prediction error.

3) The high-frequency detail features obtained based on wavelet envelope spectrum analysis can be used to detect the occurrence of faults earlier than the original whole frequency features. It is experimentally proved that the wavelet-CNN model developed in this paper is more conducive to the early fault warning of rolling bearings.

Acknowledgments

This research is sponsored by National Science and Technology Major Project (J2019-IV-004-0071), National Natural Science Foundation of China (52272436).

References

[1] B. A. Tama, M. Vania and S. Lee, Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals, *Artificial Intelligence Reviews*, 56 (2023) 4667-4709.

[2] C. Y. Yang, J. Ma and X. D. Wang, A novel based-performance degradation indicator RUL prediction model and its application in rolling bearing, *ISA Transactions*, 121 (2022) 349-364.

[3] A. R. Bastami and S. Vahid, A comprehensive evaluation of the effect of defect size in rolling element bearings on the statistical features of the vibration signal, *Mechanical Systems and Signal Processing*, 151 (2021) 107334.

[4] T. Lin, G. Chen and W. L. Ouyang, Hyper-spherical distance discrimination: a novel data description method for aero-engine rolling bearing fault detection, *Mechanical Systems and Signal Processing*, 109 (2018) 330-351.

[5] M. M. M. Islam, A. E. Prosvirin and J. M. Kim, Data-driven prognostic scheme for rolling-element bearings using a new health index and variants of least-square support vector machines, *Mechanical Systems and Signal Processing*, 160 (2021) 107853.

[6] H. L. Luo, L. Bo and X. F. Liu, A novel method for remaining useful life prediction of roller bearings involving the discrepancy and similarity of degradation trajectories, *Computational Intelligence and Neuroscience*, 2021 (2022) 2500997.

[7] J. H. Zhou, Y. Qin and D. L. Chen, Remaining useful life prediction of bearings by a new reinforced memory GRU network, *Advanced Engineering Informatics*, 53 (2022) 101682.

[8] M. Iqbal and A. K. Madan, CNC machine-bearing fault detection based on convolutional neural network using vibration and acoustic signal, *Journal of Vibration Engineering and Technologies*, 10 (5) (2022) 1613-1621.

[9] A. Choudhary, T. Mian and S. Fatima, Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images, *Measurement*, 176 (2021) 109196.

[10] R. K. Mishra, A. Choudhary and S. Fatima, A self-adaptive multiple-fault diagnosis system for rolling element bearings, *Measurement Science and Technology*, 33 (12) (2022) 125018.

[11] J. S. Yan, Y. Z. Peng and J. S. Xie, Remaining useful life prediction method for bearings based on LSTM with uncertainty quantification, *Sensors*, 22 (12) (2022) 4549.

[12] A. Jastrzebska, Lagged encoding for image-based time series classification using convolutional neural, *Statistical Analysis and Data Mining*, 13 (3) (2020) 245-260.

[13] X. Y. Zhang, G. Chen and T. F. Hao, Rolling bearing fault convolutional neural network diagnosis method based on casing signal, *Journal of Mechanical Science and Technology*, 34

(6) (2020) 2307-2316.
[14] C. Lu, Z. Y. Wang and B. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification, *Advanced Engineering Informatics*, 32 (2017) 139-151.
[15] X. Y. Liu, G. Chen and T. F. Hao, A combined deep learning model for damage size estimation of rolling bearing, *International Journal of Engine Research*, 24 (4) (2023) 1362-1373, DOI: 10.1177/14680 874221086601.
[16] H. M. Chen, Z. H. Lei and F. Y. Tian, A novel complex network community clustering method for fault diagnosis, *Measurement Science and Technology*, 34 (1) (2022) 014010.
[17] A. Neha, S. Subhamoy and M. Laurent, Estimation of local failure in tensegrity using interacting particle-ensemble Kalman filter, *Mechanical System and Signal Processing*, 160 (2021) 107824.
[18] L. Li, A. F. F. Saldivar and Y. Bai, Battery remaining useful life prediction with inheritance particle filtering, *Energies*, 12 (14) (2019) 2784.

Xiyang Liu is a Ph.D. candidate in the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. Her current research interests include deep learning and pattern recognition, and their applications in bearing fault diagnosis of aero engine.

Guo Chen received a Ph.D. degree in the School of Mechanical Engineering from the Southwest Jiaotong University, Chengdu, P. R. China, in 2000. Now he works at the College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. His current research interests include the whole aero-engine vibration, rotor-bearing dynamics, rotating machine fault diagnosis, pattern recognition and machine learning, signal analysis and processing.