RUL prediction of rolling bearings based on improved empirical wavelet transform and convolutional neural network

Wenbo Wang, Jiaojiao Zhao and Guorong Ding

Abstract
Accurate prediction the remaining useful life (RUL) of rolling bearings under complex environmental conditions is crucial for prognostics and health management (PHM). In this paper, a new method for rolling bearing RUL prediction based on improved empirical wavelet transform (IEWT) and one-dimensional convolutional neural network (1D-CNN) is proposed to overcome the interference of noise and other disturbance signals. Firstly, in view of the problem of too many spectrum divisions in the traditional empirical wavelet transform (EWT) process, the mutual information value is used to re-determine the frequency band demarcation point in the EWT. The IEWT method is introduced to adaptively divide the original vibration signal to obtain a series of empirical mode functions (EMFs). Secondly, the effective components after IEWT decomposition are extracted by mutual information and kurtosis criteria and used to extract multidimensional time-frequency domain features. Finally, the 1D-CNN is constructed with the percentage of remaining life as the tracking metric to predict the RUL of the bearings. Based on two publicly available rolling bearing datasets, the model proposed in this paper have high prediction accuracy, which is better than other prediction models. Compared to other methods, its mean absolute error (MAE) and root mean square error (RMSE) are reduced.

Keywords
Rolling bearing, remaining useful life prediction, improved EWT, 1D-CNN

Date received: 10 March 2022; accepted: 13 May 2022

Handling Editor: Chenhui Liang

Introduction
Rolling bearings are key components of rotating electromechanical equipment whose reliable operation increases the safety and efficiency of modern production equipment. Generally speaking, bearings often experience different types of failures in different environments. If effective protection measures are not taken in time, the whole machine may fail and cause huge economic losses. Therefore, accurate estimation of the running state of the bearing can provide early warning reports for equipment maintenance personnel and improve the safety of equipment operation.1

At present, the mainstream forecasting methods of remaining useful life (RUL) mainly include model-driven and data-driven methods. Model-driven prediction methods combine physical models with measured data to predict the future degradation behavior of the system as well as the RUL. However, it is difficult to describe the complex process of bearing degradation comprehensively and clearly under different environments. Therefore, it is...
hard to construct a mathematical analysis model to predict bearing RUL.\textsuperscript{2,3}

Data-driven RUL prediction can automatically infer causal relationships hidden in the data and directly extract degenerate features of complex systems. It can better deal with massive monitoring data and provide accurate RUL prediction results.\textsuperscript{4} In some existing studies, the relevant degradation features are mainly extracted from the time-frequency domain. Wavelet packet decomposition (WPD),\textsuperscript{5} empirical mode decomposition (EMD),\textsuperscript{6} and Hilbert-Huang transform (HHT)\textsuperscript{7} are used to capture the degradation trend of bearings by constructing a health index (HI).

However, a large number of degenerate features not only require expert experience, but also easily lead to feature redundancy. Neural networks can capture small changes in the bearing degradation process and are used by many scholars. Refs.\textsuperscript{8-10} based on convolutional neural network (CNN) to extract the degradation features of mechanical equipment for RUL prediction. Yoo and Baek\textsuperscript{11} build HI based on continuous wavelet transform (CWT) to synthesize time-frequency image features, and CNN is used to build a model. Recurrent neural networks (RNN)\textsuperscript{12} and its variants long short-term memory networks (LSTM)\textsuperscript{13,14} are also used for RUL prediction.

With the development of deep learning, more related algorithms are applied to estimate RUL. The use of deep networks improves the predictive performance of HI. Existing studies have shown that deep learning combined with different neural networks can achieve good results in RUL prediction. However, there is still room for improvement in bearing life prediction methods based on deep networks. Moreover, most of the existing researches directly process the original vibration signal and construct the HI through the original signal (or original Fourier spectrum, edge spectrum, etc.). In practical engineering applications, due to environmental noise and signal attenuation, the fault signal of rolling bearing is often very weak compared with the strong background noise. Especially in the early stage of bearing operation, the fault signal is often annihilated by the background noise.\textsuperscript{15} Therefore, if the weak fault features of the bearing are extracted from the raw vibration signal, the prediction accuracy of RUL can be further improved, and the adaptability of the algorithm under different working conditions can also be improved.\textsuperscript{10}

Aiming at the above problems, a new bearing RUL prediction method based on improved empirical wavelet transform (IEWT) weak fault feature extraction is proposed in this paper. Firstly, the collected bearing vibration signal is divided adaptively by EWT. The spectrum is re-divided and combined according to the mutual information (MI) value to reduce the number of frequency bands and overcome the problem of too much spectrum division of EWT. Secondly, minimum entropy deconvolution (MED) is used to reduce the noise of the reconstructed signal. Six features of wavelet packet entropy, root mean square, variance, frequency kurtosis, frequency skewness, and energy are extracted from the denoised empirical mode functions (EMFs) to characterize the bearing degradation. Finally, a 1D-CNN is used to predict the RUL.

The organization and layout of the rest of this paper are as follows: Section 2 introduces the basic structure knowledge of IEWT and CNN; Section 3 explains the construction of the model in this paper and the evaluation indicators used; Section 4 presents the specific experimental process and experiments results; the last section is the conclusion of this paper and a prospect for future work.

**Preliminaries**

**Improved empirical wavelet transform**

EWT is an adaptive analysis method based on the wavelet theoretical framework. The method firstly divides the spectrum of the original vibration signal, constructs a set of adaptive wavelet filter bank, and then analyzes the different frequency components to extract the signal with tight support characteristics. After the original signal is processed by EWT, the signal-to-noise ratio can be effectively reduced and the signal quality can be improved.

Firstly, the fast Fourier transform (FFT) is performed on the original vibration signal to obtain the frequency spectrum. The frequency range is defined as $\omega \in [0, \pi]$. The signal interval $[0, \pi]$ is divided into N intervals, and the nth interval $\Lambda_n$ can be expressed as:

$$\Lambda_n = [\omega_{n-1}, \omega_n], n = 1, 2, ..., N$$

$$\bigcup_{n-1}^{N} \Lambda_n = [0, \pi]$$

(1)

Where, $\tau_n$ is the boundary width of each frequency band, $\omega_n$ is the center frequency and the region with a frequency bandwidth of $2\tau_n$ as the transition section. The empirical wavelet is the narrowband filter defined in each frequency band $\Lambda_n$. Based on wavelet theory, the scaling function $\psi_n(\omega)$ and wavelet function $\phi_n(\omega)$ of empirical wavelet are defined in the frequency domain as follows:

$$\phi_n(\omega) = \begin{cases} 1, & |\omega| \leq (1 - \gamma)\omega_n \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_n} \left( |\omega| - (1 - \gamma)\omega_n \right) \right) \right], & (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_n \\ 0, & \text{else} \end{cases}$$

(2)
In this paper, the frequency band is re-partitioned by which brings inconvenience to the subsequent analysis. The traditional EWT adopts a scale-space method according to the initial demarcation point. Then the demarcation point is re-determined according to the relative magnitude of the mutual information value.

\[
\tilde{\psi}_n(\omega) = \begin{cases} 
1, & (1 + \gamma)\omega_n \leq |\omega| \leq (1 - \gamma)\omega_n \\
\cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\gamma n + 1} (|\omega| - (1 - \gamma)\omega_n) \right) \right], & (1 - \gamma)\omega_{n+1} \leq |\omega| \leq (1 + \gamma)\omega_{n+1} \\
\sin \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\gamma n} (|\omega| - (1 - \gamma)\omega_n) \right) \right], & (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_{n+1} \\
0, & \text{else}
\end{cases}
\]  

(3)

Where,

\[
\tau_n = \gamma \omega_n \quad 0 < \gamma < 1, \quad \gamma = \min \left( \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right)
\] 

(4)

\[
\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3), x \in [0, 1]
\] 

(5)

From the above analysis, it can be seen that the core of EWT is to reasonably divide the Fourier spectrum, that is, to accurately find \( N - 1 \) boundaries within the interval from 0 to \( \pi \). According to the idea of traditional wavelet change technology, the detail coefficient and approximate coefficient of EWT can be defined as:

\[
W_j^f(n, t) = [f(t), \psi_j(t)]
\]

\[
= \left[ f(\tau)\psi_j(t - \tau) d\tau = F^{-1}[f(\omega), \hat{\psi}_j(\omega)] \right]
\]

(6)

Among them: \( \psi_j(t) \) represents the empirical wavelet function, \( \phi_j(t) \) represents the scale function, \( \hat{\psi}_j(t) \) and \( \hat{\phi}_j(t) \) represent the Fourier transform of \( \psi_j(t) \) and \( \phi_j(t) \) in turn. \( \hat{\psi}_j(t) \) and \( \hat{\phi}_j(t) \) complex conjugate in turn, \( F(\cdot), F^{-1}(\cdot) \) represent Fourier transform and Fourier transform Leaf inverse transformation.

Based on the above formula, the reconstructed signal of the original vibration signal can be expressed as:

\[
f(t) = W_j^f(0, t) \star \phi_j(t) + \sum_{n=1}^{N} W_j^f(n, t) \star \psi_n(t)
\]

\[
= F^{-1} \left[ W_j^0(0, t) \star \phi_1(t) \right] + \sum_{n=1}^{N} \left[ W_j^f(n, t) \star \psi_n(t) \right]
\]  

(7)

In the above formula: \( \ast \) represents the convolution operation, \( W_j^0(0, t) \) and \( W_j^f(n, t) \) represent the Fourier transform of \( W_j^0(0, t) \) and \( W_j^f(n, t) \) in turn.

The mutual information value of each component obtained by spectrum division is calculated according to the initial demarcation point. Then the demarcation point is re-determined according to the relative magnitude of the mutual information value.

**Step 1:** Perform FFT on the original signal to obtain the spectrum of the vibration signal.

**Step 2:** According to the scale space method, the initial frequency band boundary point is determined, and the initial divided spectrum boundary point is obtained.

**Step 3:** The adjacent frequency bands whose component mutual information value is greater than the average value are merged into one frequency band, and the adjacent frequency bands whose component mutual information value is smaller than the average value are merged into one frequency band.

![Figure 1. IEWT Reconstruction Signal Flow Chart](image-url)
Step 4: According to the newly determined demarcation point, re-divide the frequency spectrum to obtain new decomposed signal components.

Step 5: The optimal component is selected according to the kurtosis index.

**Minimum entropy deconvolution**

The vibration signal of rolling bearing is decomposed by IEWT to obtain discrete modal components. In order to extract the more obvious shock signal in the signal, the component with larger kurtosis value is selected to reconstruct the signal. Applying MED to extract shock signals from mixed multi-source interference signals can effectively reduce the impact of acquisition paths on signal attenuation, and further highlight the shock characteristics of vibration signals. It has achieved good analysis results in the extraction of rolling bearing fault features. The specific content of the algorithm can be found in Ref. 16. The bearing vibration signal after denoising by MED can better reflect the degradation state of the rolling bearing in this life cycle, which is conducive to better evaluation of its RUL.

**1D-CNN**

CNN is a very popular deep learning framework model with powerful feature extraction capabilities and has achieved good applications in image recognition, natural language processing, and other fields. CNN mainly consists of three main parts: convolutional layers, pooling layers, and fully connected layers.

The function of the convolution layer is to perform a convolution operation on the input data and the local area of the convolution kernel, and make the local receptive field traverse the entire input data by sliding the convolution kernel window. The convolution formula is defined as follows:

\[
X_i^{l+1} = f(W_i^{l+1} * X^{(l)} + B_i^{l+1})
\]  

(9)

In the above formula: \(X_i^{l+1}\) means the \(i\)-th feature of the output value of the \((l+1)\) layer. \(W_i^{l+1}\) represents the weight matrix of the \(i\)-th convolution kernel of the \((l+1)\) layer, “\(*\)” is the convolution operation, \(X^{(l)}\) represents the output of the \((l+1)\) layer, and \(B_i^{l+1}\) represents the bias term. The function \(f\) represents the output activation function. CNN solve real-life nonlinear problems through nonlinear activation functions. Figure 2 shows several of the more common activation functions.

The role of the pooling layer is down sampling, which reduces the dimensionality of the feature map while manipulating the most important signals. The max pooling expression is as follows:

\[
y_i^{(l+1)}(j) = \max(x_i^{(k)}), k \in D_j
\]  

(10)

Among them: \(y_i^{(l+1)}(j)\) represents the element in the \(i\)-th feature map of the \((l+1)\) layer after pooling; \(D_j\) represents the \(j\)-th pooling area, and \(x_i^{(k)}\) represents the element of the \(i\)-th feature map of the \((l+1)\) layer in the range of the pooling core.

In the CNN structure, one or more fully connected layers are connected after multiple convolutional layers and pooling layers. Each neuron in the fully connected layer is fully connected with all neurons in the previous layer. Connection layers can integrate local information from convolutional or pooling layers.

**Experiment framework**

In this section, the details of all the steps of the introduced RUL prediction will be discussed. The specific experimental framework is shown in Figure 3 below.

Firstly, IEWT is used to adaptively divide the bearing vibration signals and the appropriate EMF is selected based on the kurtosis value for signal reconstruction. Secondly, MED is applied to the reconstructed signal to reduce noise. Six characteristic indexes of wavelet packet entropy, root mean square (RMS), variance, frequency kurtosis, frequency skewness, and energy are extracted from the optimal EMF. Finally, the 1D-CNN is introduced to predict the RUL.

In the 1D-CNN structure, the size of the first convolution kernel is \(6 \times 1\) and the stride is 6. Each kernel computes and operates on 6 features simultaneously, and the convolutional layer is followed by a corresponding max-pooling layer to reduce computation. The ReLU function is used as the activation function.

**Bearing degradation feature extraction**

The degenerate features are extracted from the reconstructed signal, including root mean square (RMS),
energy, variance, frequency kurtosis and frequency skewness, and wavelet packet entropy. The extracted feature expressions are shown in Table 1.

Where, $X_i$ is the amplitude of the original vibration signal, $\bar{X}$ is the average of the amplitude, $\bar{F}_i$ is the amplitude after the FFT, $\bar{F}$ is the average amplitude after the FFT. $N$ is the sampling data length. $S(\cdot)$ is the standard deviation function.

**Percentage of remaining life**

Set the life label of the $i$–th row of data to $RUL_i$, which represents the ratio between the time corresponding to the $i$–th row and the time when the bearing fails, and the ratio of the time between the starting time of the bearing and the time when the bearing fails.

$$RUL_i = \frac{n - i}{n - 1}$$  \hspace{1cm} (11)

In formula (11), $i$ is the current number of rows, and $n$ is the total number of rows. The normalization of the life label can reduce the difference between different working conditions and different life values of the bearing, which is beneficial to improve the prediction accuracy of the remaining service life. Apparently, the first time the signal is acquired, the remaining life percentage is 1. When the signal is finally obtained, the RUL is 0.

**Evaluation index**

In this paper, the following five metrics are used to measure the predictive performance of the proposed predictive model: mean absolute error (MAE), root mean squared error (RMSE), correlation index ($R^2$), adjusted correlation coefficient (Adjusted_ $R^2$), and relative accuracy (RA). The calculation formula of each indicator is as follows.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$  \hspace{1cm} (12)
\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \]  

\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \]  

\[ Adjusted R^2 = 1 - (1 - R^2) \times \frac{n - 1}{n - p - 1} \]  

\[ RA = 1 - \frac{[y_i - \hat{y}_i]}{y_i} \]

In formulas above, \( \hat{y}_i \) and \( y_i \) represents the predicted and original data, respectively. \( \bar{y} \) is average of the original data. \( n \) is the length of the sampled data and \( p = 1 \).

**Experiment verification**

**Dataset description**

In order to verify the effectiveness and superiority of this method in dealing with the rolling bearing RUL prediction problem, two experimental datasets are used to verify the experiments.

**FEMTO-ST dataset (IEEE PHM 2012).** The experiments collected different working conditions on the PRONOSTIA\textsuperscript{18} platform as shown in Figure 4. The horizontal vibration signal frequency is 25.6 kHZ and is recorded every 0.1 s. The sampling interval is 10 s. The specific sampling description is shown in Figure 5 and Table 2.

The original vibration signal of the bearing 1_1 during its entire service life is shown in Figure 6. The horizontal and vertical coordinates represent time and vibration amplitude. As time goes by, the amplitude of the bearing vibration signal gradually increases, indicating that the signal has a rich diagnosis useful information.

**XJTU-SY Dataset.** The XJTU-SY dataset\textsuperscript{19} contains the full life cycle vibration signals of 15 rolling bearings under 3 working conditions. The experimental platform is shown in Figure 7. The sensor sampling frequency is
25.6 kHz, and the sampling interval is 1 min. Each sampling time is 1.28 s, and each sampling point is 32,768. Table 3 gives a detailed description of the dataset. Figures 8 and 9 are diagram of the sampling process and original vibration signal, respectively.

**Experiment procedure**

Take the PHM bearing 2_7 dataset as an example to illustrate the effect of IEWT. Figure 10 shows the initial 16 frequency bands obtained according to the scale-space method. Based on the initial demarcation point, the frequency bands are re-divided according to mutual information, as shown in Figure 11 below.

The spectrum is re-divided according to the component mutual information in the above Figure 12, and the adjacent frequency bands whose mutual information value is greater or less than the average value are combined to obtain the re-divided spectrum from left to right. The number of bands has been reduced from 16 to 6. Figure 13 is a time domain plot of the repartitioned components.

Figures 14 and 15 are the original bearing vibration signal and the reconstructed signal after IEWT processing. Figures 16 and 17 are the Fourier spectrum of original vibration signal and reconstructed signal. After filtering out the optimal component of IEWT, the interference of noise and other interference is suppressed. The extracted component signals preserve the main fault information of the bearing. Therefore, the constructed factor can more effectively reflect the bearing failure characteristics and thus better predict the remaining life of the bearing. The corresponding six feature indicators are calculated for each collected data, and the percentage of remaining lifespan is used as a tracking indicator for 1D-CNN training.

A 1D-CNN model is constructed to estimate the RUL as shown in Figure 18. The hidden hyperparameters of CNN are robust. The CNN structure used in this article has seven layers, including three convolutional layers, two maximum pooling layers, and two fully connected layers. The experiment was performed using Windows 10 (Microsoft, USA) system, the central processing unit (Central Processing Unit, CPU) used a 1.80 GHz i5 processor, the memory was 8GB, and the experiment software used MATLAB 2019a (MathWorks, USA) version.

**Experiment results**

Each time the data passed into the 1D-CNN structure is normalized 6-dimensional feature data. During the training of the 1D-CNN, the optimizer uses the “adam” optimizer, which runs for 200 iterations. Through the leave-one-out method test, the results of the five evaluation indicators of each bearing are as follows Tables 4 and 5.

For the PHM 2012 dataset, it can be seen that the prediction effect under the three working conditions is still relatively consistent, and the trend of the remaining life of the bearing can be better predicted. The load of the first working condition is 4 kN, which is relatively low in the comparative experimental load, so the bearing running time is relatively long. The collected training data is longer, so the training model is more

| Condition  | Rotating speed (r/min) | Load (kN) | Bearing information |
|------------|------------------------|-----------|---------------------|
| Condition 1 | 2100                   | 12        | 1_1,1_2,1_3,1_4,1_5 |
| Condition 2 | 2250                   | 11        | 2_1,2_2,2_3,2_4,2_5 |
| Condition 3 | 2400                   | 10        | 3_1,3_2,3_3,4_3,5 |
Figure 7. XJTU-SY dataset experimental platform.

Figure 8. Sensor data collection process.

Figure 9. Raw vibration signal of XJTU-SY Bearing 1_1.

Figure 10. Initial frequency band.

Figure 11. Mutual information value of each frequency band.
**Figure 12.** Re-divided frequency band.

**Figure 13.** Time series plot of the new component.

**Figure 14.** Original vibration signal.

**Figure 15.** Reconstructed signal.

**Figure 16.** Fourier spectrum of original vibration signal.

**Figure 17.** Fourier spectrum of reconstructed signal.
effective. In the same experiment, the third condition had the largest load. The corresponding bearing has a shorter operating time and less data is collected. The samples trained by the 1D-CNN model are greatly reduced, and the error of the corresponding model is relatively large, but it is still within an acceptable range.

For the XJTU-SY dataset, the radial load of working condition 1 is 12 kN, which is the largest in the same experiment and the corresponding speed is the lowest. The radial load of working condition 3 is 10 kN, the maximum speed is 2400 r/min, and the collected data is also the most. Figures 19 and 20 are renderings of rolling bearing RUL predictions under different operating conditions in the two datasets.

As a comparative experiment, EMD, stationary wavelet transform (SWT), and variational mode decomposition (VMD) are selected to compare with IEWT. And in order to show the robustness of 1D-CNN, LSTM is selected for comparative experiments. In the experiments, the number of EMD layers is adaptive. The number of SWT layers is 3, and the wavelet basis is “morlet.” The number of VMD layers is 5. The number of nerve cells in each hidden layer of LSTM is 150, and the number of hidden layers is 5.

In the selected PHM dataset, the corresponding average value of the evaluation index calculated for each bearing is calculated, and the results are shown in the Table 6. The MAE for IEWT-CNN, EMD-CNN, SWT-CNN, VMD-CNN, and IEWT-LSTM are 0.0685, 0.0814, 0.1196, 0.0719 and 0.0726, respectively. Comparing the other four methods, the MAE proposed in this paper decreased by 15.85%, 42.73%, 4.73%, and 5.64%, respectively.

In the selected XJTU-SY dataset, the evaluation index of each bearing and the corresponding mean value are calculated. The results are shown in the Table 7. The MAE of the method proposed in this paper is 0.0451 and the RMSE is 0.0553. Among them, MAE, RMSE, and MAPE are all smaller than the prediction results obtained by EMD-CNN, SWT-CNN, VMD-CNN and IEWT-LSTM. Moreover, $R^2$, adjusted $R^2$, and RA are all higher than the prediction accuracy of the above four models, indicating that the model proposed in this paper is superior to the above four models. Compared with several other methods, the complexity of the model is also relatively low, and the running time is relatively fast.

**Conclusion**

In this paper, A new method for rolling bearing RUL prediction based on IEWT weak fault feature...
extraction and 1D-CNN is proposed to overcome the interference of noise and other disturbance signals. Firstly, the bearing vibration signal is adaptively divided by EWT, and the frequency bands are re-divided according to the mutual information value to reduce the number of frequency bands. Secondly, some

Table 5. Results of each evaluation index of the XJTU-SY.

|                | MAE    | RMSE   | R²      | Adjusted_R² | RA     |
|----------------|--------|--------|---------|-------------|--------|
| Bearing 1_1    | 0.0252 | 0.0313 | 0.9885  | 0.9884      | 0.9597 |
| Bearing 1_2    | 0.0360 | 0.0431 | 0.9778  | 0.9777      | 0.9423 |
| Bearing 1_3    | 0.0438 | 0.0539 | 0.9685  | 0.9683      | 0.9401 |
| Bearing 1_4    | 0.0507 | 0.0596 | 0.9570  | 0.9567      | 0.9191 |
| Bearing 1_5    | 0.0167 | 0.0215 | 0.9947  | 0.9946      | 0.9751 |
| Bearing 2_1    | 0.0660 | 0.0847 | 0.9138  | 0.9137      | 0.9003 |
| Bearing 2_2    | 0.0189 | 0.0240 | 0.9934  | 0.9934      | 0.9720 |
| Bearing 2_3    | 0.0318 | 0.0416 | 0.9774  | 0.9774      | 0.9562 |
| Bearing 2_4    | 0.0369 | 0.0461 | 0.9763  | 0.9757      | 0.9420 |
| Bearing 2_5    | 0.0325 | 0.0417 | 0.9771  | 0.9771      | 0.9494 |
| Bearing 3_1    | 0.0323 | 0.0414 | 0.9771  | 0.9771      | 0.9498 |
| Bearing 3_2    | 0.0475 | 0.0680 | 0.9426  | 0.9425      | 0.9296 |
| Bearing 3_3    | 0.0609 | 0.0767 | 0.9116  | 0.9113      | 0.8985 |
| Bearing 3_4    | 0.0654 | 0.0833 | 0.9289  | 0.9288      | 0.8994 |
| Bearing 3_5    | 0.0592 | 0.0684 | 0.9468  | 0.9463      | 0.9116 |
| Average        | 0.0415 | 0.0523 | 0.9621  | 0.9619      | 0.9363 |

Figure 19. PHM 2012 dataset predicted result. (a) PHM bearing 1_1 predicted result, (b) PHM bearing 2_2 predicted result, (c) PHM bearing 3_1 predicted result, (d) PHM bearing 3_3 predicted result.
appropriate components are selected according to the kurtosis index, and deconvolution with minimum entropy is used to reduce the noise of the reconstructed signal. Six feature metrics are extracted from the denoised EMF. Based on the constructed feature metrics, a 1D-CNN is applied to predict the RUL of figure 20.

Table 6. Evaluation index comparison for PHM 2012 dataset.

| Methods   | MAE   | RMSE  | $R^2$     | Adjusted $R^2$ | RA     | Time/s |
|-----------|-------|-------|-----------|----------------|--------|--------|
| EMD-CNN   | 0.0814| 0.1139| 0.8002    | 0.8001         | 0.8428 | 37.1853|
| SWT-CNN   | 0.1196| 0.1395| 0.7725    | 0.7723         | 0.8021 | 37.0917|
| VMD-CNN   | 0.0719| 0.090  | 0.8350    | 0.8349         | 0.8594 | 79.7619|
| IEWT-LSTM | 0.0726| 0.1072| 0.8396    | 0.8394         | 0.8506 | 64.3129|
| IEWT-CNN  | 0.0685| 0.0872| 0.8891    | 0.8889         | 0.8954 | 35.0621|

Table 7. Evaluation index comparison for XJTU-SY dataset.

| Methods   | MAE   | RMSE  | $R^2$     | Adjusted $R^2$ | RA     | Time/s |
|-----------|-------|-------|-----------|----------------|--------|--------|
| EMD-CNN   | 0.0683| 0.0767| 0.9221    | 0.9220         | 0.9096 | 39.4902|
| SWT-CNN   | 0.0712| 0.0817| 0.9064    | 0.9063         | 0.8854 | 37.9383|
| VMD-CNN   | 0.0757| 0.0683| 0.9389    | 0.9388         | 0.9127 | 81.2595|
| IEWT-LSTM | 0.0644| 0.0672| 0.9448    | 0.9446         | 0.9284 | 75.5420|
| IEWT-CNN  | 0.0415| 0.0523| 0.9621    | 0.9619         | 0.9363 | 33.9318|
rolling bearings. Finally, based on the validation on two public rolling bearing datasets, the proposed bearing RUL prediction method has higher prediction performance.

In the future, this method will be applied to more industrial scenarios, including gearboxes and aero-engines, etc. In addition, some other methods of adaptively extracting features to construct HI should be applicable. Other potential degradation metrics will try to combine with the CNN model for higher health prediction accuracy. Future work also includes applying the proposed framework to a wider range of case studies on experimental data in other applications, as well as investigating other potential degenerate labels to achieve higher RUL estimation accuracy.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Natural Science Foundation of China under Grant 61671338 and Grant 51877161 and Open fund of Hubei Key Laboratory of Metallurgical Industry Process System Science (No. Y202007).

ORCID iD
Guorong Ding https://orcid.org/0000-0001-9320-0084

References
1. Zeng D, Yang J, et al. Bearing life prediction method based on parallel multi-channel convolutional long and short-term memory network. China Mech Eng 2020; 31: 2454–2462 + 2471.
2. Pan Y, Hong R, Chen J, et al. Performance degradation assessment of a wind turbine gearbox based on multi-sensor data fusion. Mech Mach Theory 2019; 137: 509–526.
3. Lei Y, Li N, Gontarz S, et al. A model-based method for remaining useful life prediction of machinery. IEEE Trans Reliab 2016; 65: 1314–1326.
4. Xia T, Song Y, Zheng Y, et al. An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. Comput Ind 2020; 115: 103182.
5. Wang Y, Peng Y, Zi Y, et al. A two-stage data-driven-based prognostic approach for bearing degradation problem. IEEE Trans Ind Inform 2016; 12: 924–932.
6. Hong S, Zhou Z, Zio E, et al. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method. Digit Signal Process 2014; 27: 159–166.
7. Soualhi A, Medjaher K and Zerhouni N. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Trans Instrum 2015; 64: 52–62.
8. Li X, Ding Q and Sun JQ. Remaining useful life estimation in prognostics using deep convolution neural networks. Reliab Eng Syst Saf 2018; 172: 1–11.
9. Guo L, Lei Y, Li N, et al. Machinery health indicator construction based on convolutional neural networks considering trend burr. Neurocomputing 2018; 292: 142–150.
10. Cheng Y, Hu K, Wu J, et al. A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings. Adv Eng Inform 2021; 48: 101247.
11. Yoo Y and Baek JG. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. Appl Sci 2018; 8: 1102.
12. Guo L, Li N, Jia F, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 2017; 240: 98–109.
13. Zhang J, Wang P, Yan R, et al. Long short-term memory for machine remaining life prediction. Int J Ind Manuf Syst Eng 2018; 48: 78–86.
14. Wu Y, Yuan M, Dong S, et al. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing 2018; 275: 167–179.
15. Huang G, Zhang Y and Ou J. Transfer remaining useful life estimation of bearing using depth-wise separable convolution recurrent network. Measurement 2021; 176: 109090.
16. Qiao Z, Liu Y and Liao Y. Application of improved wavelet transform and minimum entropy deconvolution in railway bearing fault diagnosis. J Vib Shock 2021; 40: 81–90 + 118.
17. Wei T and Xu L. Research on the method of bearing fault feature extraction of mine drainage pump based on MED and wavelet packet entropy. Coal Mine Machinery 2021; 42: 170–173.
18. Nectoux P, Gouriveau R, Medjaher K, et al. PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In: Proceedings of the IEEE International Conference on Prognostics and Health Management. IEEE, 2012; 1–8.
19. Wang B, Lei Y, Li N, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Trans Reliab 2020; 69: 401–412.