



# Remaining useful life prediction for rolling bearings based on similarity feature fusion and convolutional neural network

Lei Nie<sup>1</sup> · Lvfan Zhang<sup>1</sup> · Shiyi Xu<sup>1</sup> · Wentao Cai<sup>1</sup> · Haoming Yang<sup>1</sup>

Received: 3 March 2022 / Accepted: 17 June 2022 / Published online: 11 July 2022  
© The Author(s) 2022

## Abstract

As a critical content of condition-based maintenance (CBM) for mechanical systems, remaining useful life (RUL) prediction of rolling bearing attracts extensive attention to this day. Through mining the bearing degradation rule from operating data, the deep learning method is often used to perform RUL prediction. However, due to the complexity of operating data, it is usually difficult to establish a satisfactory deep learning model for accurate RUL prediction. Thus, a novel convolutional neural network (CNN) prediction method based on similarity feature fusion is proposed. In this paper, the similarity features are extracted based on the correlation between statistical features and time series. After sensitive feature screening, eligible features are applied to develop a health indicator (HI), which can be used to define the bearing failure stages and reduces the complexity of the CNN model. Subsequently, a one-dimensional CNN is established to predict the RUL of bearing, and the HI is utilized to train the prediction model. The proposed approach is verified by FEMTO bearing datasets and IMS bearing datasets. And the experimental results reveal the superiority and effectiveness of the feature fusion-based CNN method in constructing HI and accurate RUL prediction.

**Keywords** Rolling bearing · Health indicator · RUL prediction · CNN

## 1 Introduction

As one of the most essential components, rolling bearing is widely used in various mechanical systems, and its running state determines the machinery's RUL. The failure of bearings can result in serious accidents and major economic loss. Therefore, to ensure the safe and reliable operation of a mechanical system, RUL prediction of bearings is indispensable.

With the increasing complexity of sampling data, supervised data-driven methods have been extensively researched [1–3]. The core idea is to train neural network models using HI and RUL labels, which can then be used to predict the RUL of bearings [4–6]. Therefore, HI performance significantly influences RUL prediction, and the construction of HI is generally based on monitoring signals. Different

HI construction strategies can be divided into physical HI (PHI) and fusion HI (FHI) [7]. PHI can be obtained through a simple signal processing method, such as root mean square (RMS) and kurtosis of a bearing vibration signal, which contains some physical meanings. In [8], the RMS is employed as PHI to indicate the degradation trend of bearings. Meanwhile, FHI is constructed by fusing or reducing the dimensionality of physical features, which can reflect the degradation process but has no real physical significance. For example, Gao et al. [9] extracted time-domain features and fused the data to build an HI; they then predicted the bearing RUL using the Metropolis–Hastings algorithm. In addition, Tian et al. [10] combined four selected features into one HI to characterize the life state. The above HI performed well and achieved good prediction. Based on these studies, this paper proposes a new HI construction method. First, the approach extracts similarity features based on the correlation between statistical features and time-series data. Then, it combines the time–frequency domain features to build an HI, which can effectively mine useful degradation information and represent the health states of bearings.

HI quality is an important factor affecting the RUL prediction accuracy; another essential factor is the learning

---

Technical Editor: Jarir Mahfoud.

✉ Lvfan Zhang  
zhanglvfan@163.com

<sup>1</sup> College of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China

capability of intelligent computational algorithms. With the improvement in computing power, deep learning (DL) [11], including long short-term memory (LSTM) [12, 13], deep belief network (DBN) [14], and CNN [15–17], has become a hot research topic. Ding et al. [18] considered the transition of different degradation stages and proposed one- and multi-stage iteration prediction models based on the LSTM neural network for RUL prediction. The experimental results showed the competitive performance of the LSTM neural network. DBN, one of the critical approaches for big data processing and analysis, can also be used to predict RUL. For example, Jason et al. [19] used DBN to predict the RUL of rotating equipment, which explore the self-taught feature-learning capability of DBN. Among the various DL models, a CNN attracts special attention because of its feature extraction ability, which is widely used in recognition and detection problems in the field of computer vision. [20–22] In addition, 1-D CNNs have recently been successfully applied to speech recognition and document reading [23–25]; moreover, the utilization of CNN-based models shows considerable potential in RUL prediction [25–28]. Compared to a deep neural network, a CNN uses fewer parameters to achieve the same functionality or precision [29]. Cheng et al. [30] exploited the adoption of CNNs in predicting the bearing RUL, and the experimental results demonstrated a superior performance of the proposed CNN model. Hence, the current paper focuses on constructing significant bearing HI which can be utilized to define the bearing failure stages. And proposes a 1-D CNN model to mine the underlying information of the HI, learning the degradation of bearings according to the HI and RUL labels. The model's main function is to reveal the hidden dependence between an HI and training bearing RUL by using the feature extraction ability of CNNs for RUL prediction.

This study tests the bearing RUL by using two DL methods: a CNN and an LSTM network model, which are also applied to validate the effectiveness of the proposed approach.

The viability of the proposed approach is proved using the FEMTO bearing datasets [31] and the Intelligent Maintenance System (IMS) bearing dataset [32]. The following contributions can be summarized:

1. A related similarity (RS) feature construction approach, used for both time and frequency domains, is presented based on the similarity between the current data and original operating-point data. The RS features are built to mine valid degradation information, ranging from 0 to 1, so that normalization can be omitted.
2. A special HI based on the RS feature and principal component analysis (PCA) dimension reduction is constructed. This can effectively illustrate the degradation in bearing operation.

3. A CNN model that inputs the HI of rolling bearings is presented, which improves the accuracy of RUL estimations.
4. The potential of the proposed method is demonstrated using the experimental datasets. The results show that CNNs perform better than the LSTM prognosis approach.

The novelty of the proposed method includes (1) the HI based on the RS feature and PCA can effectively reflect the health status of bearings; and (2) the special HI significantly reduces the complexity of the CNN model and improves the accuracy of predicting the RUL of bearings.

The remaining paper is organized as follows. Section 2 describes the proposed approach; Sect. 3 presents the experimental verification of the proposed methodology and provides the results. Finally, Sect. 4 concludes the study.

## 2 Proposed approach

The framework of the proposed method is based on the run-to-failure data of rolling bearings: the time, frequency, and time–frequency domain features are first extracted to construct similarity features, and then sensitive features are screened from them. The HI using the data fusion method to construct is to characterize bearing degradation status. Finally, the bearing RUL is predicted using a data-driven model.

Figure 1 shows the flowchart of the proposed method. First, 23 features in the time, frequency, and time–frequency domains are extracted to form the characteristic time series of rolling bearing vibration signals. Then, the similar characteristics of the time and frequency domains are constructed based on the Pearson correlation coefficient—which, together with the time–frequency domain features, constitute 14 bearing degradation features. Next, the features are sorted by the Cori index according to monotonicity and tendency, and the features with scores exceeding the preset threshold are selected. Using PCA, these features are combined to construct the HI. Finally, a CNN degradation model is constructed, and the HI is used as its input to predict the RUL of rolling bearings. The implementation process is detailed in Sects. 2 and 3. The RUL prediction method is verified using the whole life data of bearing vibration acceleration signals in FEMTO bearing datasets.

### 2.1 Data acquisition

In this paper, FEMTO bearing datasets [31] and IMS bearing datasets [32] have been used to analyze the degradation of the bearings and estimate their RUL. Details of the dataset are provided in Sect. 3.

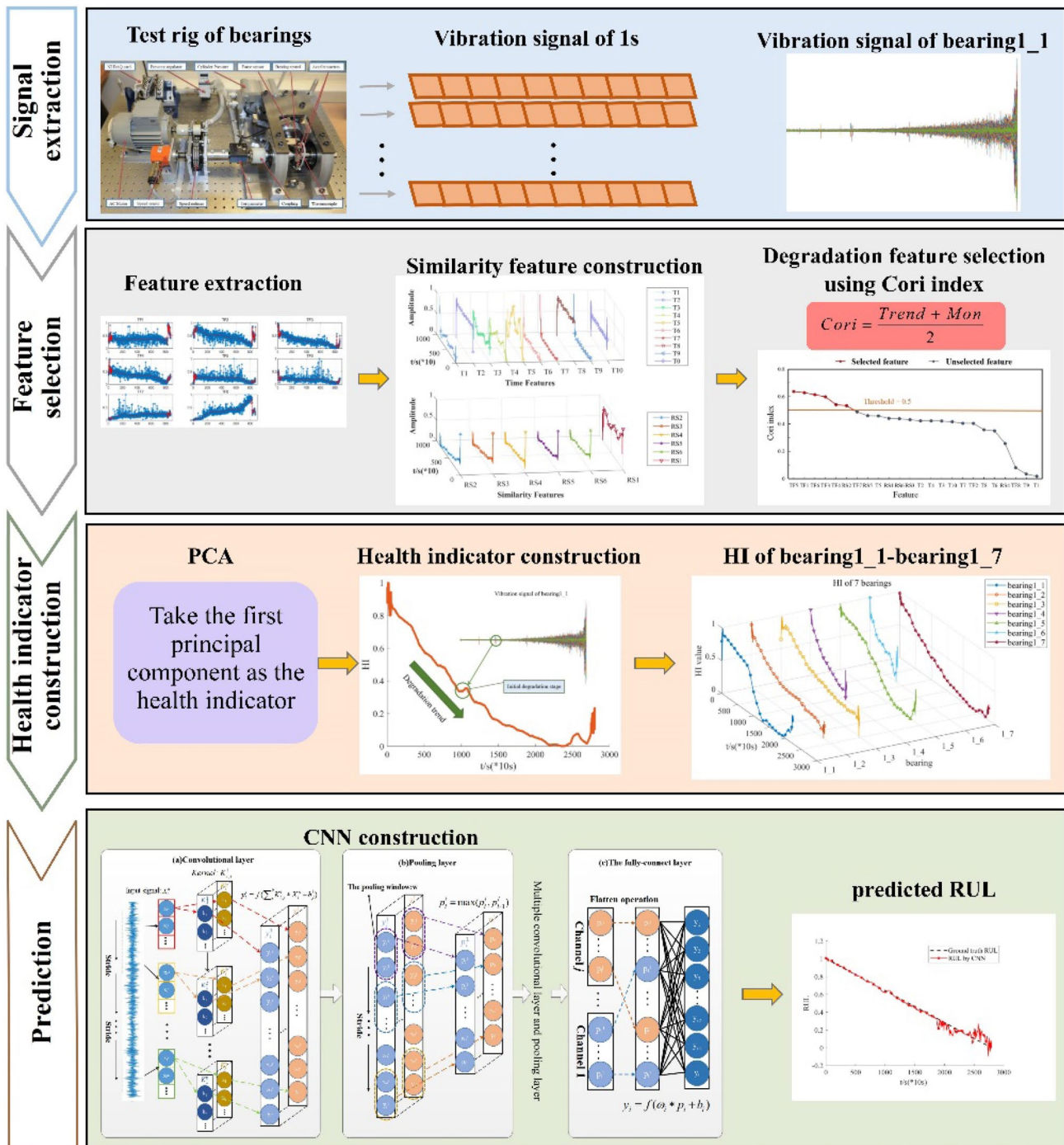


Fig.1 Process of remaining life CNN prediction method based on similarity feature fusion

### 2.2 Feature extraction

For the vibration signal  $x = [x_1, x_2, \dots, x_n]$  collected in each sampling period, 10 traditional time-domain statistical characteristics are extracted (Appendix).

In the frequency domain, FFT transforms the time-domain signal  $x = [x_1, x_2, \dots, x_n]$  into the frequency-domain

signal  $s = [s_1, s_2, \dots, s_{nfft}]$ . FFT is used to extract the complete frequency spectrum of the time-domain signal, which is divided into four sub-band spectra, and the amplitudes of this spectrum and the four sub-band spectra are applied as five frequency-domain features. In the time–frequency domain, to obtain the nonlinear characteristics of the vibration signal more comprehensively, multilevel wavelet packet

decomposition is used to extract the time–frequency features. In this study, the Haar wavelet is used to decompose the vibration signal into a group of wavelet packet nodes through three-level wavelet packet decomposition, which divides the frequency axis into eight frequency bands, and the energy ratio of each band is exploited as eight time–frequency features.

### 2.3 Similarity feature construction based on pearson’s correlation coefficient

Pearson’s correlation coefficient can better reflect the correlation between feature time-series data, which is widely used in similarity measurement [33]. Hence, it is used as the basis for calculating time- and frequency-domain RS features, the calculation method is according to formula (1):

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\left[ \sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2 \right]^{1/2}} \tag{1}$$

Here, R lies in the interval [-1,1]. When the Pearson correlation coefficient is close to 1, the correlation between the x and y sequences is greater than the coefficient close to 0. To describe the correlation between the features more directly, the absolute value of R is adopted to construct the RS features of the time and frequency domains. Let  $f_t$  indicate the feature series sampled at time t and  $f_0$  indicate the time series of the initial observation time; then, the RS features can be calculated according to formula (2).

$$RS = \frac{\left| \sum_{i=1}^N (f_0^i - \bar{f}_0)(f_t^i - \bar{f}_t) \right|}{\left[ \sum_{i=1}^N (f_0^i - \bar{f}_0)^2 \sum_{i=1}^N (f_t^i - \bar{f}_t)^2 \right]^{1/2}} \tag{2}$$

Type:

N – length of each feature time series;

$\bar{f}_0$ – the average of  $\{f_0^i\}_{i=1:N}$ ;

$\bar{f}_t$ – the average of  $\{f_t^i\}_{i=1:N}$ .

If  $f_t$  is equal to  $f_0$ , RS is 1, and the bearing is in the initial health state. When t increases gradually, the time series and initial observation time series show a different trend. At this moment, RS decreases continuously until it reaches 0, and the bearing is in a state of degradation.

After the similarity feature calculation method is applied in 10 traditional time-domain features, one time-domain RS feature, named RS1, is acquired. Using the same method, we obtain five frequency-domain RS features, which are marked as RS2–RS6. We consider them and the eight time–frequency domain features as the original feature set. All features are shown in Table 1.

### 2.4 Sensitive features selection

In the original feature set, there are redundant features that cannot completely describe the life state of the bearings; therefore, the sensitive features is screened from the original feature set to select the features that can accurately reflect the degradation process of the bearing. To construct a suitable HI, the trend and monotonicity of the original feature are quantitatively evaluated. The trend can assess the linear correlation between the feature and observation time. The greater the trend, the more the feature tends to degrade over time, and the greater the correlation with the true value. Monotonicity can evaluate the nature of a feature’s increasing or decreasing trend on the time axis. The greater the monotonicity, the better is the monotonicity of the feature; otherwise, the feature is oscillating.

For the feature sequence  $f = [f_1, f_2, \dots, f_k]$  and time sequence  $t = [t_1, t_2, \dots, t_k]$ ,  $f_t$  is the corresponding degradation feature value at time t, where K is the largest sample number value. The specific calculation method of Trend is shown in Formula (3):

$$Trend(F, T) = \frac{\left| \sum_{k=1}^K (f_t - \bar{F})(t_k - \bar{T}) \right|}{\sqrt{\sum_{k=1}^K (f_t - \bar{F})^2 \sum_{k=1}^K (t_k - \bar{T})^2}} \tag{3}$$

**Table 1** Original feature set

RS features		Energy ratio features
Time-domain	Frequency-domain	Time–Frequency domain
RS1	RS of 10 classical time-domain features	TF1 Energy ratio of (3,0)
	RS2 RS of [0,12800] Hz	TF2 Energy ratio of (3,1)
	RS3 RS of [0,3200] Hz	TF3 Energy ratio of (3,2)
	RS4 RS of [3200,6400] Hz	TF4 Energy ratio of (3,3)
	RS5 RS of [6400,9600] Hz	TF5 Energy ratio of (3,4)
	RS6 RS of [9600,12800] Hz	TF6 Energy ratio of (3,5)
		TF7 Energy ratio of (3,6)
		TF8 Energy ratio of (3,7)

Type:

$N$  – length of each feature time series;

$\bar{F}$  – the average of  $\{f_i\}_{i=1:K}$ ;

$\bar{T}$  – the average of  $\{t_i\}_{i=1:K}$ .

The specific calculation method of monotonicity is shown in formula (4):

$$Mon(F) = \left| \frac{dF > 0}{K - 1} - \frac{dF < 0}{K - 1} \right| \tag{4}$$

Type:

dF–Differentiation of characteristic sequences.

Both trend and monotonicity are in the range [0, 1]. To measure the original features, this paper uses the Cori index to screen the sensitive feature. The specific calculation method is shown in formula (5):

$$Cori = \frac{Trend + Mon}{2} \tag{5}$$

The Cori index for each original feature is calculated according to the ranking of the Cori values, and the corresponding features above the set threshold are screened out to form a sensitive feature set. Hence, features that have good representation capabilities of bearings are selected.

Feature extraction is carried out for time-series data within  $T$  time units, and  $N$  feature time series on a sensitive feature set constitute a feature matrix  $X_{T \times N}$ ; then,  $F_{\{1:T\}}^i$  is used to represent the feature time series of the  $i$ th column of  $X_{T \times N}$ .

### 2.5 HI construction based on PCA

To quantitatively evaluate the degradation state of the bearings, the HI is constructed based on the whole bearing life cycle. In the HI construction process, PCA is a widely used method in data fusion and feature extraction [34], which can screen out the interaction information of features on sensitive feature sets. Through orthogonal transformation, the original features are projected to a few comprehensive features; therefore, a small amount of data can be used to represent the main information features implied by the input data to achieve dimensionality reduction. The HI construction approach based on PCA can be described as follows:

- (1) To address the problem of inconsistent data weights caused by the excessive difference in sensitive feature dimensions, the maximum and minimum normalization method is used to normalize  $X_{T \times N}$ , which is indicated as  $X_{norm}^{T \times N}$ .

$$X_{norm}^{T \times N} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{6}$$

- (2) The covariance matrix  $C$  of  $X_{norm}^{T \times N}$  is identified.

$$C = \sum_{i=1}^n ((X_i - \mu)(X_i - \mu)^T) \tag{7}$$

where  $X_i$  are the variables of the data set  $X_{norm}^{T \times N}$ , and  $\mu$  is the mean vector.

- (3) The eigenvalue  $\lambda_i$  of  $C$  and the corresponding eigenvector  $v_i$  are calculated, and the relationship is as follows.

$$Av_i = \lambda_i v_i, i = 1, 2, 3, \dots, n \tag{8}$$

The eigenvalue is the variance of the first principal component (PC1), and the eigenvector is the column vector of the transformation matrix.

- (4) The eigenvalues are sorted, the largest  $K$  eigenvalues are selected, and the corresponding  $K$  eigenvectors are used as row vectors to form the eigenvector matrix  $P$ .
- (5) The data are converted to a new space constructed by  $K$  feature vectors, and  $Y\{1:T\} = [y1, y2, \dots, yT]$  is obtained through  $Y = PX$ , which is the principal component acquired by PCA fusion.
- (6) The maximum and minimum normalization processing is performed on PC1, which is applied as a dimensionless index that maps the result to 0–1. As the HI in this paper, PC1 does not have any physical meaning but has good monotonicity and trend, which can reflect the performance degradation trend of the rolling bearings more intuitively and significantly.

### 2.6 One-dimensional CNN for RUL prediction

A CNN is a representative DL algorithm, which is a feed-forward neural network inspired by the local receptive field in biological neurology. Among an increasing number of variants of the CNN model, a 1-D CNN is a suitable model for processing 1-D signals in the industrial field [35].

In general, the one-dimensional CNN structure is mainly composed of the convolutional layer, the pooling layer, and the full connection layer, which exhibit the characteristics of local connection, weight-sharing, pooling, and multilayer structure [36].

#### (1) Convolutional layer

Frist is the input layer, which mainly includes the convolutional layer shown in Fig. 2(a). Each convolutional layer consists of a set of learnable kernels  $K$  (also known as the filter). The convolved feature is calculated by the dot product between the weight of the filter and the perceived region of input data, which can be expressed as:

$$y_i^l = f \left( \sum_i^k K_{ij}^l * X_i^n + b_j^l \right) \tag{9}$$

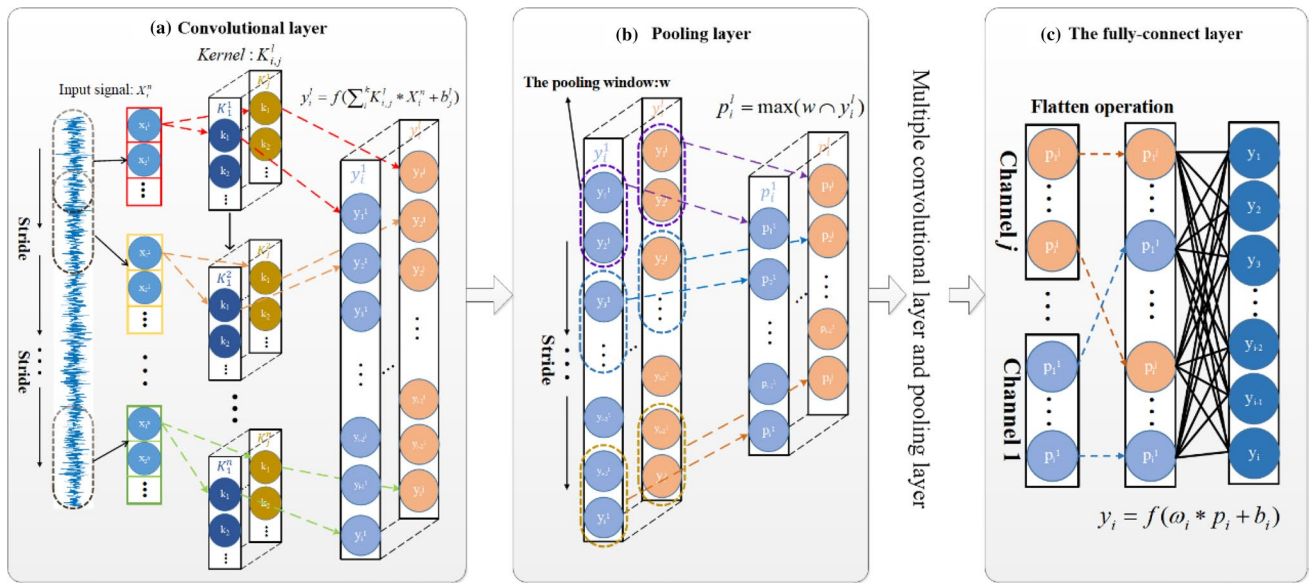


Fig. 2 One-dimensional CNN structure diagram

where  $X_i^n$  is the input data in the first layer  $n$ .  $K_{i,j}^l$  and  $b_j^l$  denote the weight and bias vectors respectively.  $y_j^l$  is the output of the  $j$ th feature map channel in the layer 1.  $f(\cdot)$  is the nonlinear activation function that determines if the neuron is awake. The Rectified Linear Unit (ReLU) is used for neurons in the convolutional layer and is widely employed in pattern recognition. Its ability to accelerate the convergence and mitigate the vanishing gradient problem has been proven [35]. The ReLU can be represented as follows:

$$f(x) = \max(0, x) \tag{10}$$

**(2) Pooling layer**

In the second part, the pooling layer can compress the data volume and improve the computing efficiency after the convolutional layer, which is shown in Fig. 2(b). The type of pooling layer includes maximum pooling, average pooling, random pooling, and sum area pooling. Since the maximum pooling can get faster convergence and improve generalization, the maximum pooling is utilized in this paper, and its mathematical model can be represented as:

$$p_i^l = \max(w \cap y_j^l) \tag{11}$$

where  $y_j^l$  and  $p_i^l$  represent the channel input and output of the pooling layer.  $W$  represents the pooling window, which can slide with a certain step.  $\cap$  represents the overlap between the pooling window and the channel output of relu.  $\max(\cdot)$  represents the operation that takes the max of input data.

**(3) Fully connected layer**

The third part is composed of a fully connected layer and is used to further extract the features and connect with the output layer after the convolutional layer and pooling layer. Figure 2 shows the above-described topology structure of CNN. The neurons in the fully connected layer are interconnected according to the following calculation:

$$y_i = f(w_i * p_i + b_i) \tag{12}$$

where  $y_i$  and  $p_i$  represent the output and input of the fully connected layer respectively.  $w_i$  represents the weight matrix, and  $b_i$  is the bias.  $f(\cdot)$  is the activation function in the fully connected layer.

The prediction process of the bearing is based on 1-D time-series signals. The training model is affected by an accurate addition of labels to the data [37]. Therefore, after the segmented processing of the HI, RULs are added to the original data as labels by using the constructed HI value.

The run-to-failure data sample of the bearing is considered as the training set  $\{x_t, y_t\}$ , and the HI values are taken as the  $x_t$  of the training set; the percentage of the bearing RUL at  $t$  is applied as the corresponding label  $y_t$ . Assuming that RUL1 is the whole life of the bearing, the RUL at time  $t$  is represented as  $RUL_t$ . The calculation of the RUL percentage is as shown in formula (13):

$$y_t = \frac{RUL_t}{RUL_1} \tag{13}$$

When  $t$  is 0,  $y_t$  is 1, which indicates that the bearing has just begun to degenerate. When  $y_t$  is 0, it means that the bearing has completely failed, and the RUL is 0 at the moment.

The HI divided into 10 time steps and the RUL label as the 11th time step are directly input into the 1D-CNN model, and the output of the model is the predicted RUL value, which reflects the input vibration signal moment of the bearing. The step size is set as 1 to train the CNN model. Assuming that the HI includes 2803 values, the dimensions of the input sequence it will split are  $2793 \times 10$ . In this paper, the convolution kernel of the CNN model is set to 64, the kernel size is 2, and the pooling layer size is 2.

### 3 Experimental study and result discussion

Two datasets are used to validate the proposed approach: FEMTO bearing datasets [31] and IMS dataset [32]. First, the sensitive features are extracted from the raw data, and then an HI is constructed to train the CNN model for RUL estimation. The details of the datasets and the experiments are revealed in the following subsection.

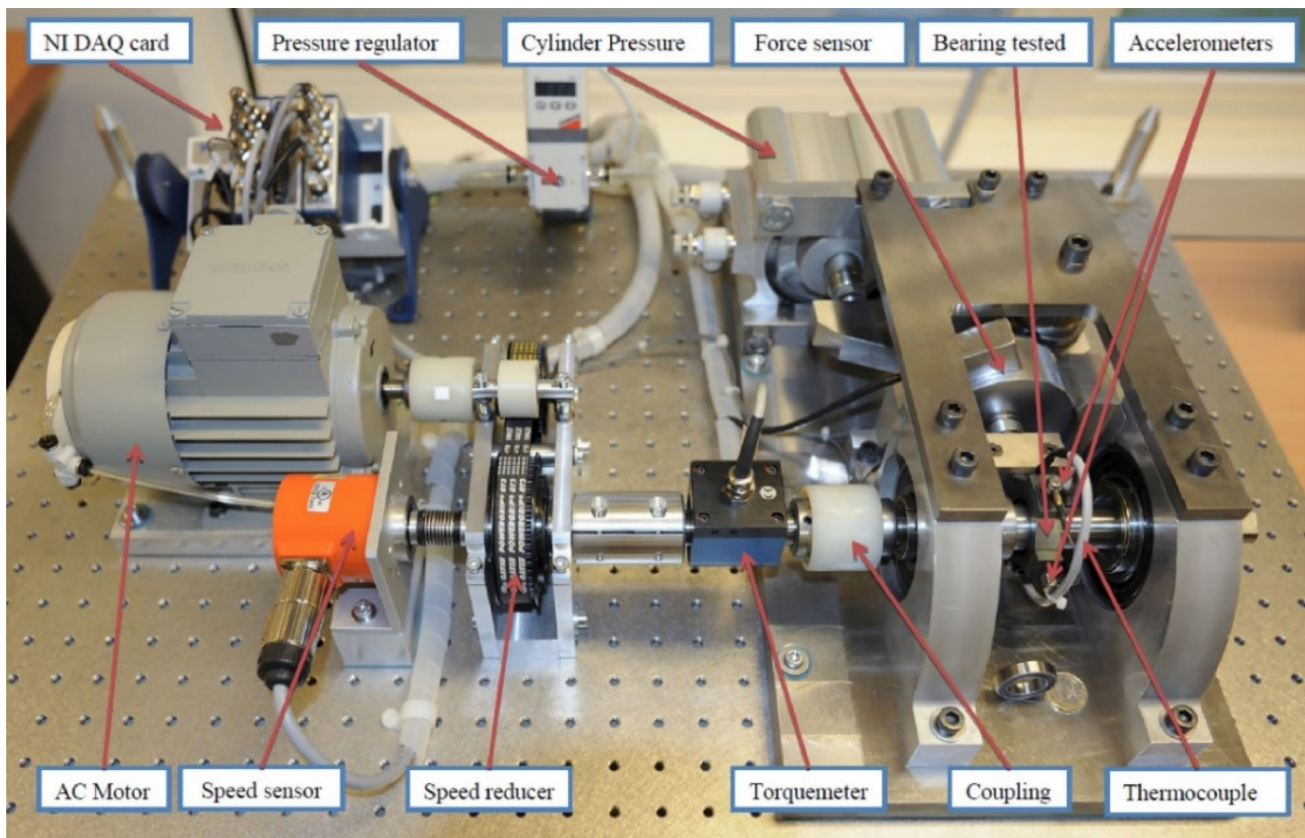
### 3.1 Case study on the FEMTO bearing datasets

#### 3.1.1 Data description

The bearing testbed is shown in Fig. 3. The proposed method is tested using the public FEMTO-ST bearing dataset from the PHM Data Challenge held by the Institute of Electrical and Electronics Engineers (IEEE) in 2012, which covers the vibration acceleration data under three load conditions (Table 2): (1)1800 rpm and 4000 N, (2)1650 rpm and

**Table 2** FEMTO bearing datasets

Dataset	Condition1	Condition2	Condition3
Learning set	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Testing set	Bearing1_3	Bearing2_3	Bearing3_3
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	



**Fig. 3** Data acquisition test platform of FEMTO bearing datasets

4200 N, (3)1500 rpm, and 5000 N. Under each working condition, there are two run-to-failure bearings, which are considered as training bearings, and the remaining bearings are considered as test bearings. The data sampling frequency is 25.6 kHz; sampling is performed every 10 s, with each sampling lasting for 0.1 s; and 2560 vibration data points are recorded at a time. The test bench and test description are detailed in reference [31].

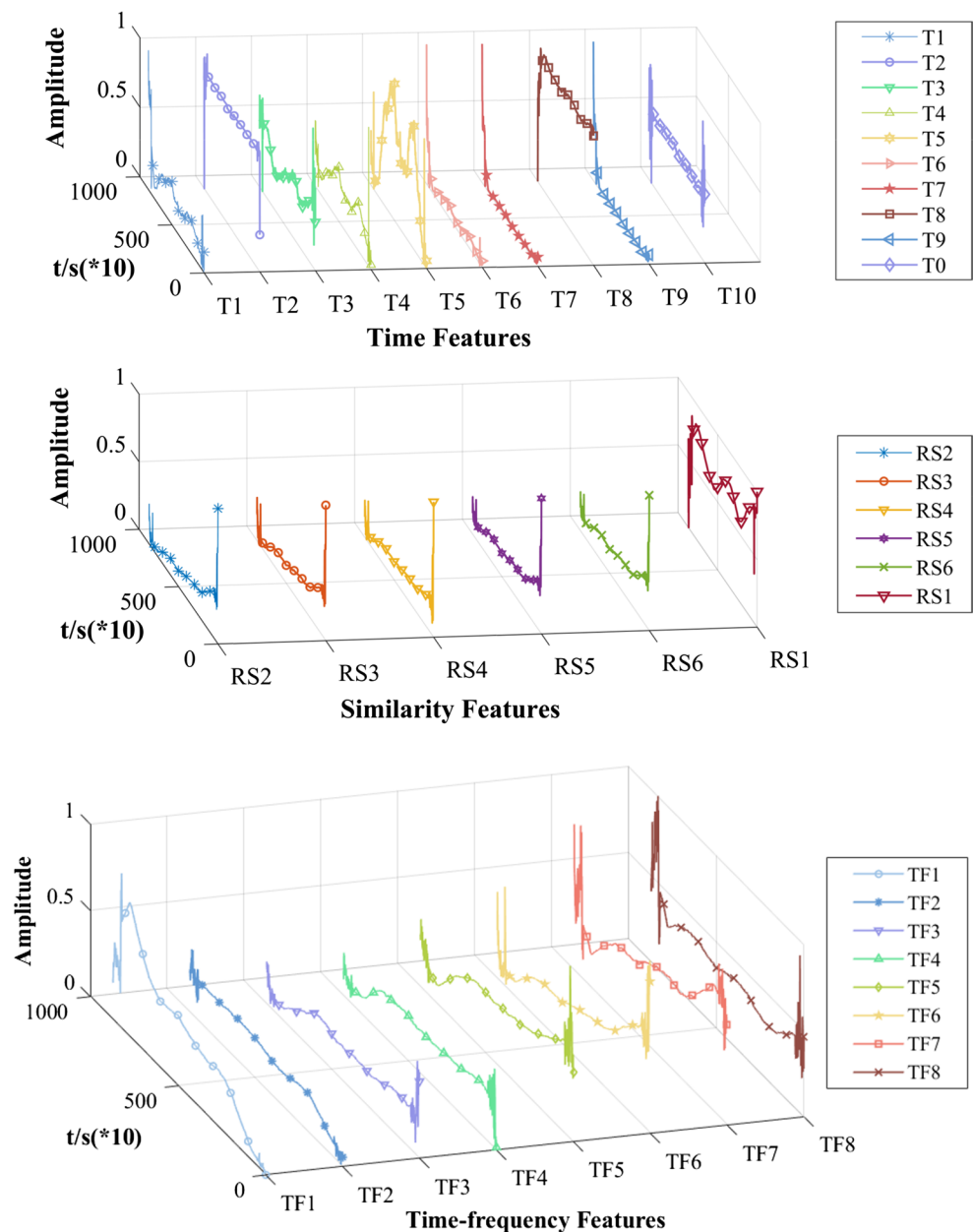
To validate the proposed approach, Bearings 1\_1, 2\_1, and 3\_1 are selected as the training sets, and 11 are randomly

selected as testing sets, including 1\_2 to 1\_7, 2\_2 to 2\_5, and 3\_3.

### 3.1.2 Feature extraction and selection

The horizontal vibration signals of 14 bearings are pre-processed and the original signals are filtered using the wavelet denoising method. To eliminate the influence of periodic changes and random fluctuations, the original features are processed with a moving average (MA) of a window 100; then, normalized and eliminated outliers are processed. Taking Bearing 1\_2 as an example, the

**Fig. 4** Visualized image of initial features of bearing 1\_2





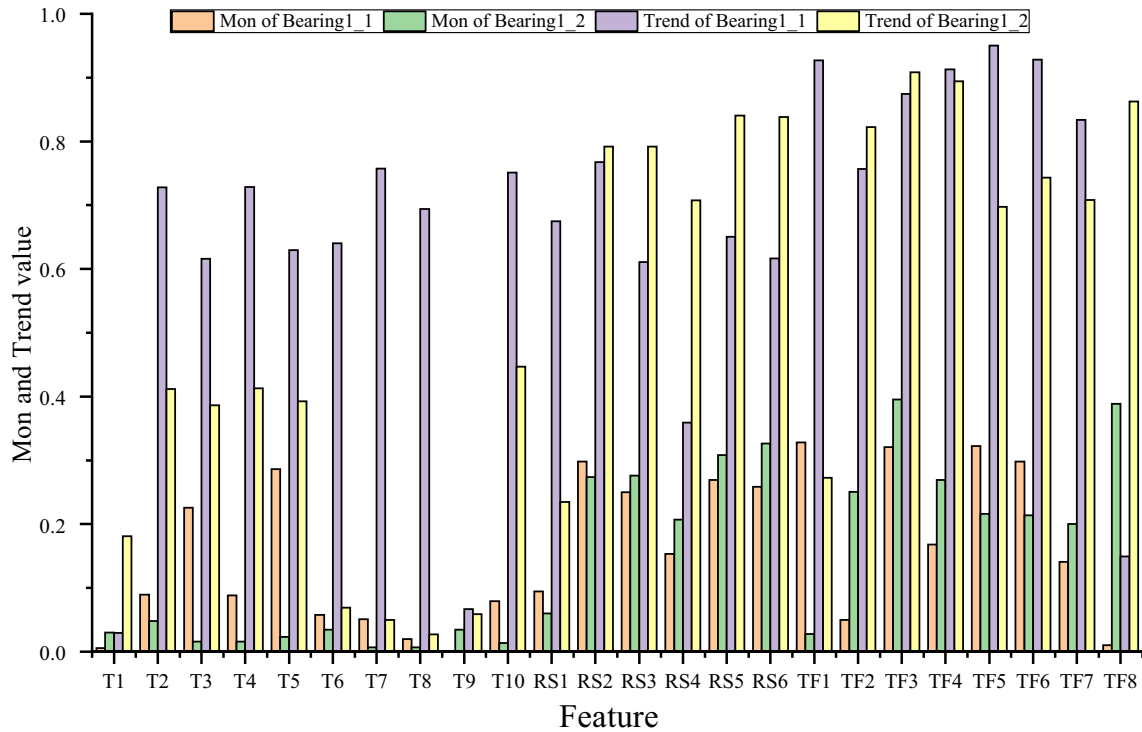


Fig. 5 Monotonicity and trend index of bearings 1\_1and1\_2

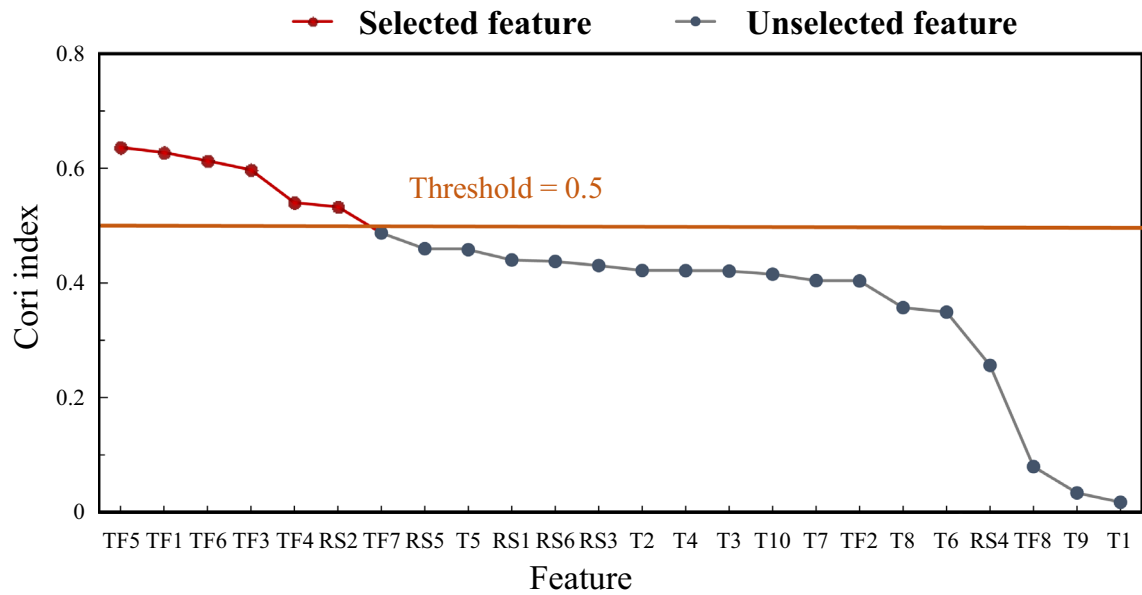


Fig. 6 Cori index screening of bearing 1\_1

visualized image of the extracted and fused features is shown in Fig. 4.

The two indicators of the 23 features in Bearings 1\_1 and 1\_2 are shown in Fig. 5.

The Cori index of the 14 features of Bearing 1\_1 is calculated and sorted; the result is shown in Fig. 6. The threshold is set as 0.5. The features higher than the threshold are TF5, TF1, TF6, TF3, TF4, and RS2, which form the sensitive feature set.

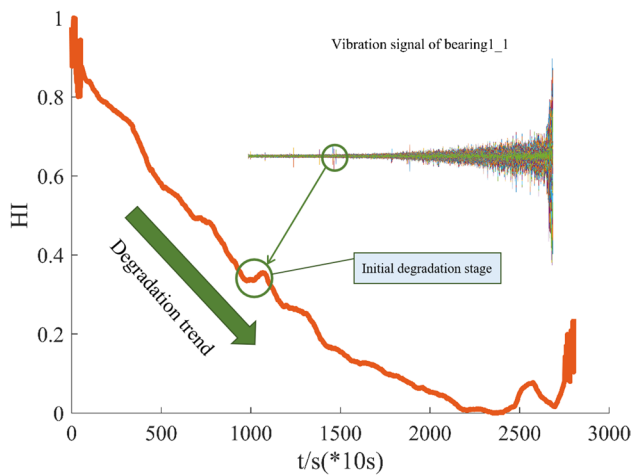


Fig. 7 HI of bearing 1\_1

### 3.1.3 HI construction

The result of the fusion of the features in the sensitive feature set based on PCA is shown in Fig. 7, which is the HI of Bearing 1\_1.

The RMS features used to calculate the stop threshold of the rolling bearing test are compared to verify the effectiveness of the HI. As mentioned in the International ISO 2372 Mechanical Vibration Industry Standard, for medium-sized machinery, when the RMS of the vibration signal reaches 2.0–2.2 g, it indicates that the equipment is in a dangerous state [38]. Figure 8 shows the vibration signal RMS and

HI of Bearings 1\_1 and 1\_2 in the whole life cycle. Figure 8(a) shows that RMS cannot effectively characterize the performance degradation state of the bearing. For the HI constructed using this method, as shown in Fig. 8(b), the HI value scale range of the two bearings is roughly the same throughout the life cycle. Relative to Bearing 2, whose RMS cannot characterize the performance degradation state, HI can sense its early degradation and characterize its degradation state.

Figure 9 shows the HI constructed by the sensitive feature selection method for seven bearings. When the HI value exceeds 0.5, it means that the test bearing is in the normal state. When the HI value ranges from 0.2 to 0.5, the test bearing is in the initial degradation stage. When it is less than 0.2, the test bearing is in the rapid degradation stage until complete failure.

### 3.1.4 RUL prediction and comparison with different model

The proposed 1-D CNN model is used to predict the bearing RUL, and the vibration signal acquired at each sampling time of Bearing 1\_1 is utilized to construct the HI. After the HI is constructed, it is applied to train the CNN model and predict RUL. As shown in Fig. 10, the abscissa is the current time and the ordinate is the RUL percentage. The figure shows that the RUL curve of the bearing predicted by the CNN model is close to the actual remaining life curve. When the observation time is approximately 10000S, the curve has evident fluctuations. Compared to the bearing HI, the results predicted by the feature fusion-based CNN method can perceive the initial degradation of the bearing. The prediction

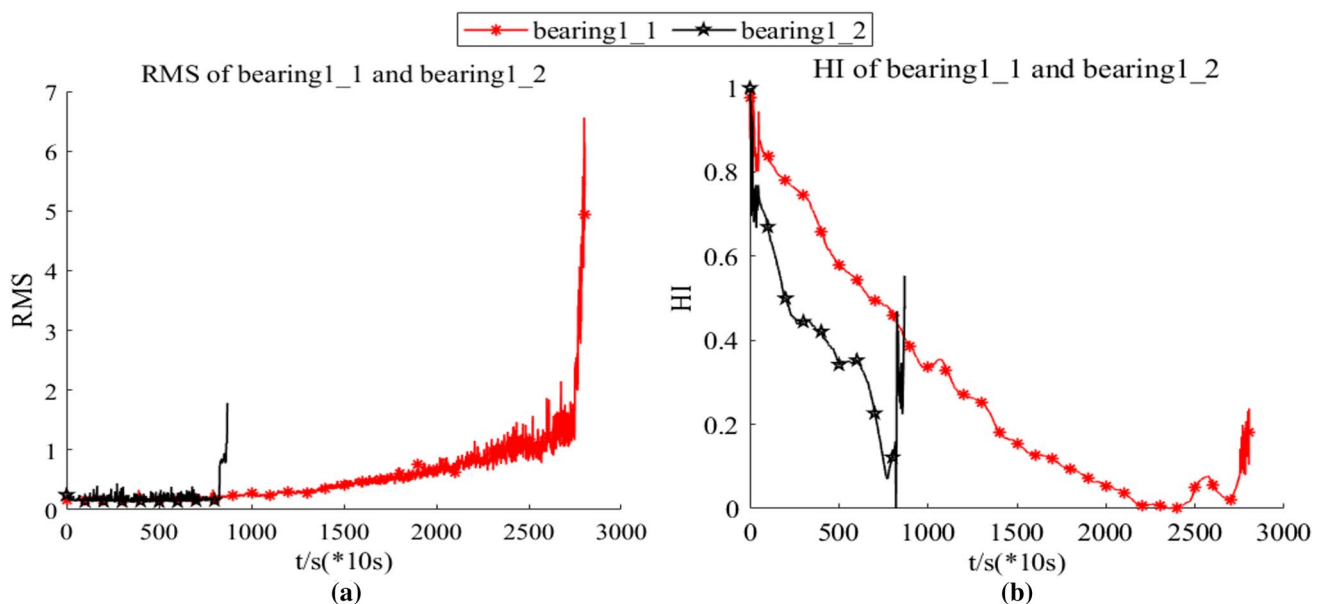
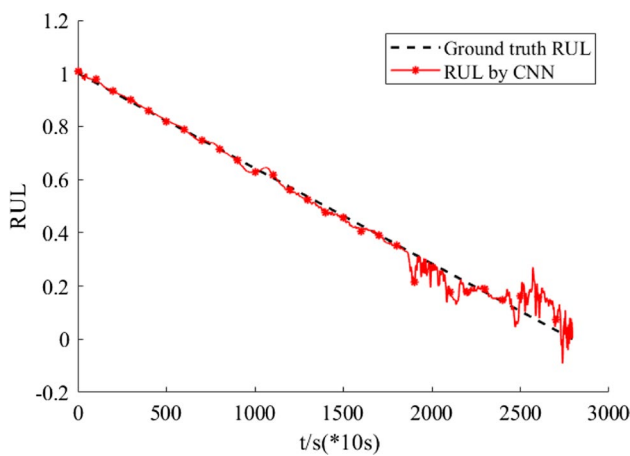
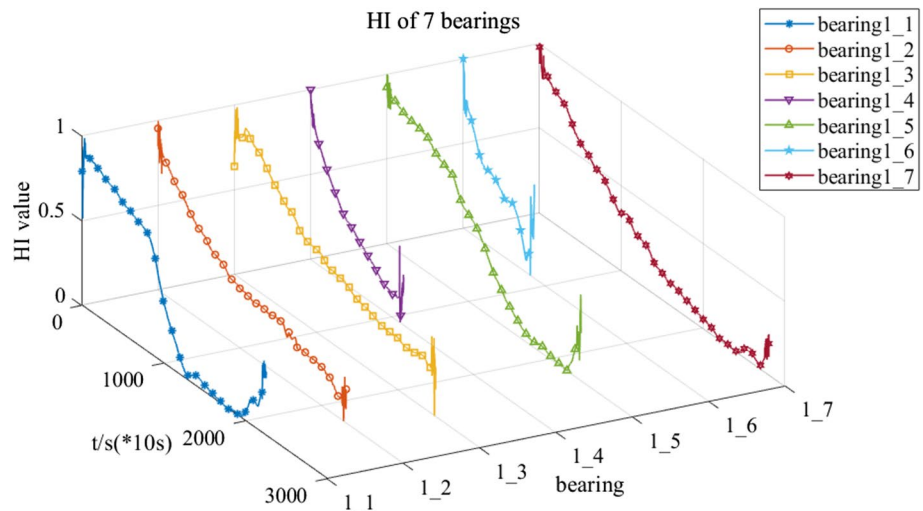


Fig. 8 RMS and HI of bearings 1\_1 and 1\_2

**Fig. 9** Bearing HI under similar operating conditions



**Fig. 10** Prediction effect map of bearing 1\_1

result fluctuates up and down in the true RUL at the end of the lifetime, which is close to the truth RUL as a whole, showing a downward trend.

To verify the effectiveness and superiority of RUL prediction acquired by the CNN model, the model is compared with the LSTM model and the Bi-LSTM model. LSTM is a special memory neural network that can consider the memory data of current and historical periods to achieve the prediction, which can extract time-series features after training and solve long-term dependence problems [39]. Similar to LSTM, Bi-LSTM is a corresponding variant of the new recursive network architecture; both of them function as solutions to long-term dependency problems [40].

Taking Bearings 1\_3, 1\_5, 1\_7, 2\_2, 2\_4, and 3\_3 as examples, the HIs are input into the trained CNN, LSTM, and Bi-LSTM to obtain the truth RUL fit curves and the predicted RUL values for each time point. This paper applies MA to smoothen the predicted curves to visually display the curve of the predicted RUL values acquired by the CNN

model. The actual and predicted percentage of RUL curves are shown in Fig. 11. The result shows that the results predicted by the CNN model show a better fit to the real RUL compared to other models, which signifies that the proposed approach can appropriately predict the bearing RUL. The method is still effective for bearings' RUL prediction under different working conditions.

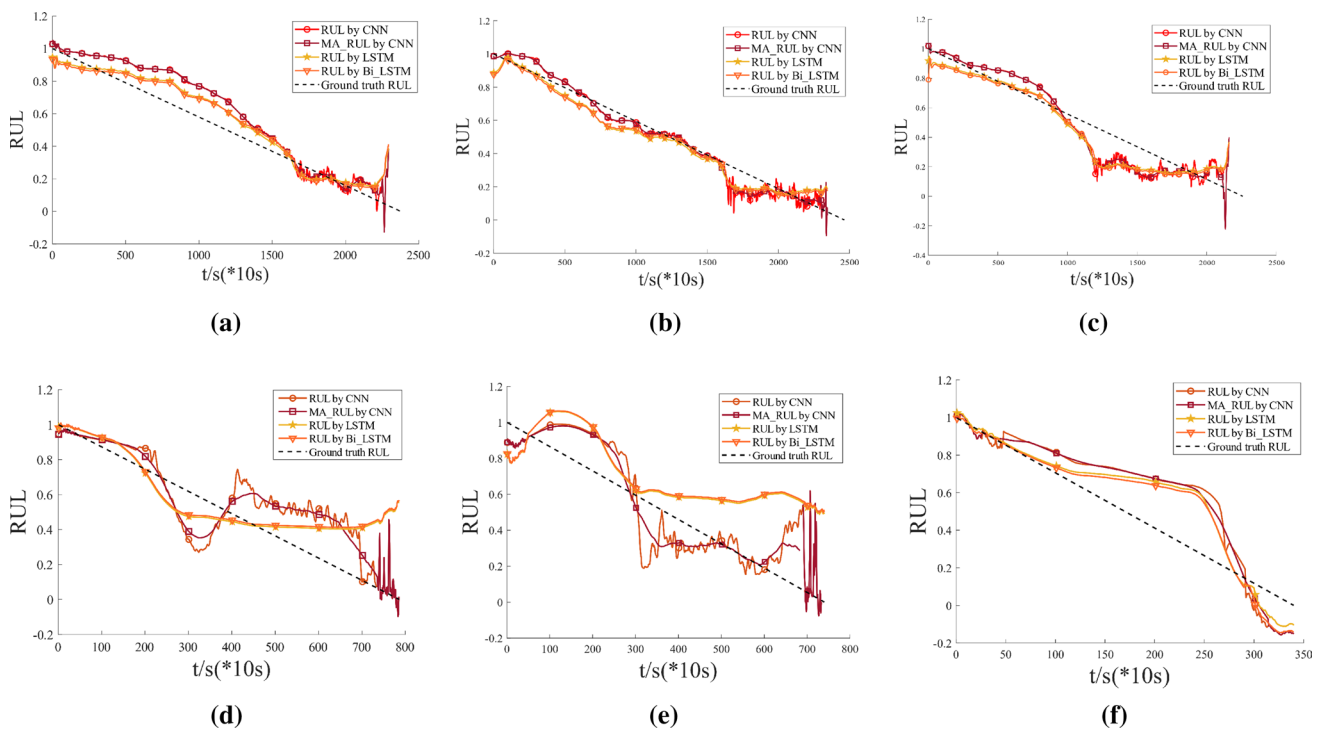
To describe the prediction effect of CNN and LSTM models more accurately, the models are evaluated using mean absolute error (MAE), root mean squared error (RMSE), and maximum error (MAX), which are calculated as follows:

$$MAE = \left( \sum_{i=1}^n |p_i - y_i| \right) / n \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - y_i)^2} \tag{15}$$

$$MAX = \max(|p_i - y_i|) \tag{16}$$

The approach proposed to predict RUL is compared with other methods. To indicate the advantage of the proposed feature-fusion HI at RUL prediction, this paper uses the RMS of bearing vibration signals as the HI to predict RUL by CNN, LSTM, and Bi-LSTM. Taking seven bearings, as illustrated, the RUL prediction error results of LSTM-HI, Bi-LSTM-HI, CNN-RMS, LSTM-RMS, Bi-LSTM-RMS, and our method are shown in Fig. 12. The proposed method outperforms the other methods. Most cases in predicting 11 bearings' RUL indicate that the proposed approach can achieve the lowest MAE, RMSE, and MAX errors. In particular, compared to the method using RMS, the feature fusion-based HI can appropriately predict the bearing RUL, and the prediction error is much lower than that of RUL



**Fig. 11** RUL prediction result: **a** bearing 1\_3, **b** bearing1\_5, **c** bearing1\_7 **d** bearing2\_2, **e** bearing2\_4 and **f** bearing3\_3

prediction based on RMS. The lower the error, the higher is the prediction ability of the method.

**State-of-the-art comparison:** In order to illustrate the advantage of the proposed method, the obtained results are compared with five other recent relevant methods on the FEMTO dataset. In particular, the sparse domain adaptation network (SDAN) in [41], the method based on time empirical mode decomposition (EMD) and temporal convolutional networks (TCN) in [42], the gated dual attention unit (GDAU) in [43], the multiscale convolutional neural network (MSCNN) in [44], and the deep separable convolutional network (DSCN) in [45] are adopted and labeled as SDAN, EMD-TCN, GDAU, MSCNN and DSCN, respectively. The comparison results are indicated in Fig. 13.

As shown in Fig. 13, the RMSE of the proposed approach is the smallest among the six methods, which means that the prediction accuracy of CNN-HI is greatly superior to other methods, and the proposed method achieves the best performance among all these methods.

### 3.2 Case study on the IMS bearing datasets

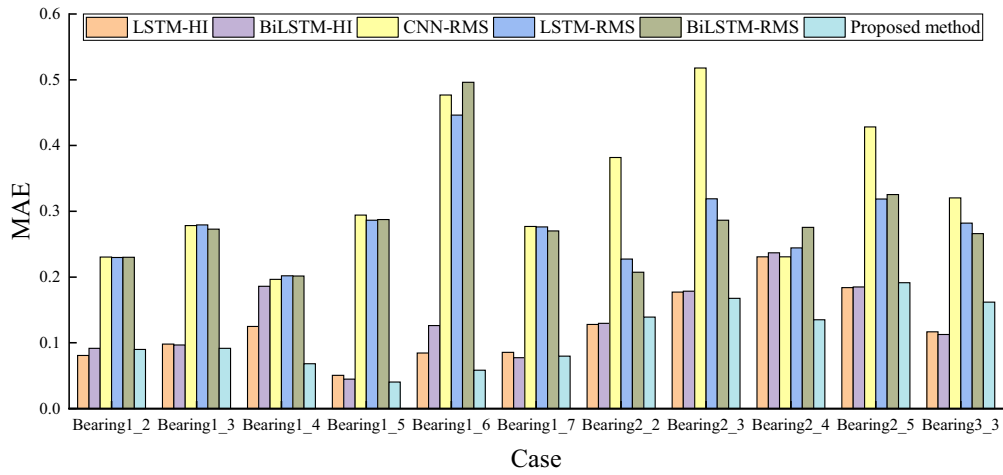
This dataset is provided by IMS, University of Cincinnati [46], and it is fully described on NASA's website. IMS datasets are composed of three data subsets. The second subset is

applied to verify the feature fusion-based CNN method, containing 984 files, each of which comprises 20,480 data samples with a sampling frequency of 20 kHz. And the vibration data was collected every 10 min by a NI 6062E DAQ Card.

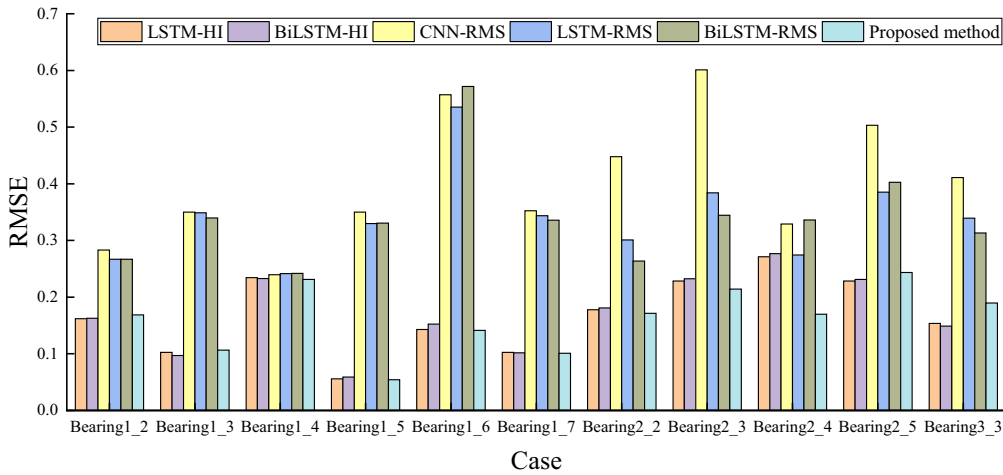
Unlike the FEMTO dataset bearings, the bearings in this dataset experience longer and more complicated degradation processes, which increases the reliability of this dataset and the difficulty of RUL prediction. Hence, the proposed method predicts the RUL of IMS bearings. The subset includes four bearings, named IMS 2\_1 to IMS 2\_4. IMS 2\_1 is selected as a training set, and IMS 2\_2, 2\_3, and 2\_4 are used as testing sets. The prediction results are shown in Fig. 14, and the prediction errors are shown in Fig. 15.

A comparison of the life prediction results of the FEMTO and IMS datasets indicates that.

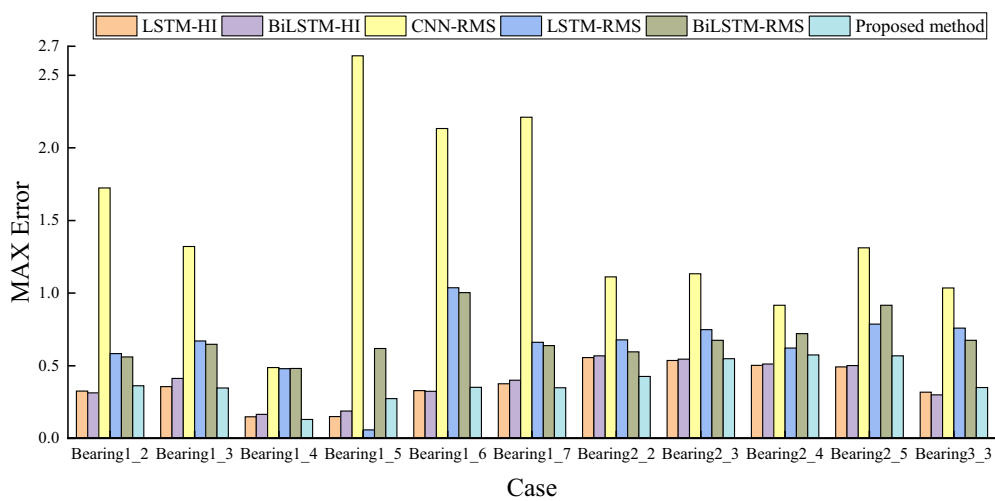
- (1) The time- and frequency-domain similarity features constructed by this method consider the influence of time on the features and lay the foundation for constructing a suitable HI.
- (2) Based on the HI constructed in this paper using the CNN model to predict the early degradation of perceptible bearings, the HI constructed by this method can characterize the degradation state of bearings under similar operating conditions, which has a good generalization capability verified by test bearings with different failure modes and failure degrees.



(a)



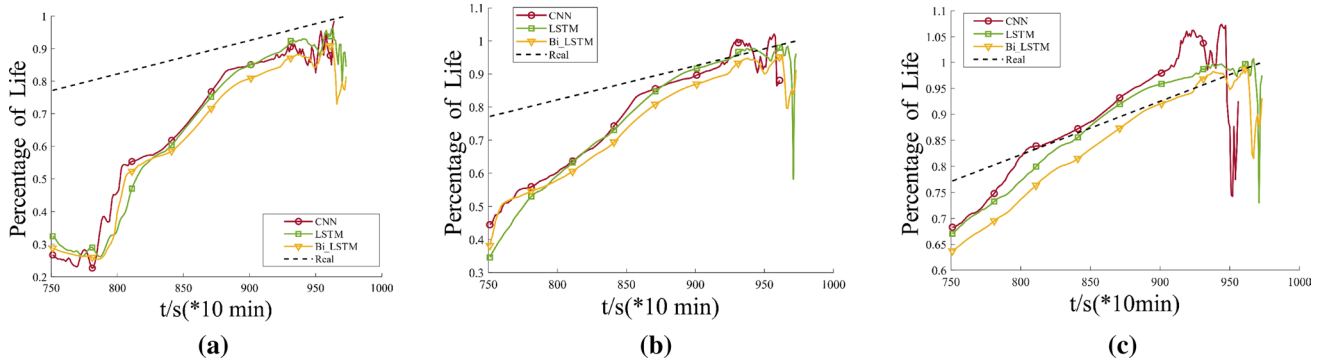
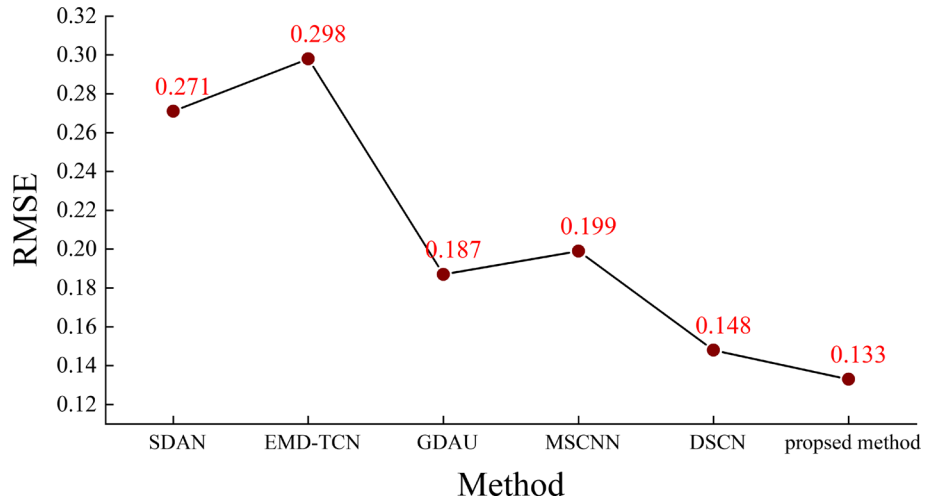
(b)



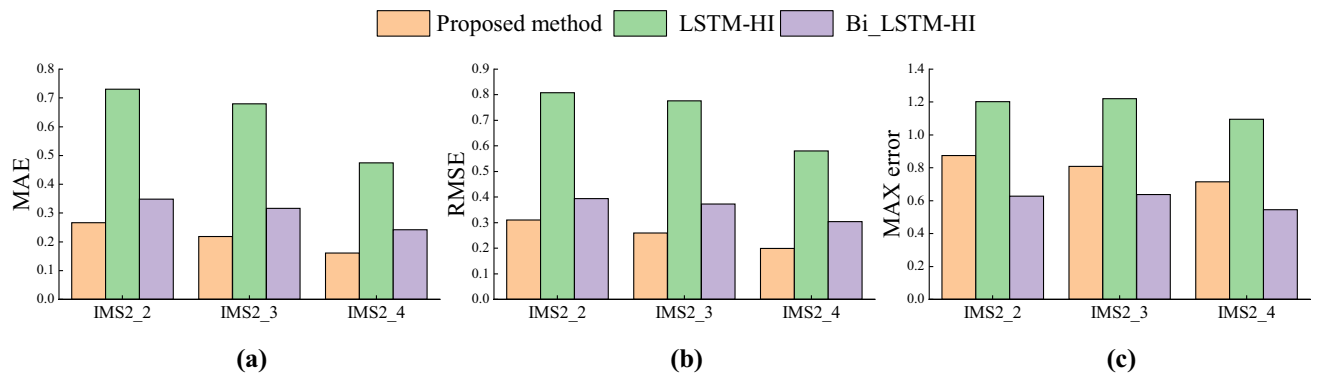
(c)

Fig.12 prediction errors of 11 bearings: a MAE, b RMSE, c MAX error

**Fig. 13** Comparison results with recent methods



**Fig. 14** Percentage of life prediction of 3 bearings: **a** IMS 2\_2, **b** IMS 2\_3, **c** IMS 2\_4



**Fig. 15** prediction errors of 3 bearings: **a** MAE, **b** RMSE, **c** MAX error

(3) The remaining life prediction method based on the CNN model has a better overall effect on predicting the bearing RUL than other neural network models and has a better prediction capability for 1-D time-series signals and relatively better prediction accuracy.

### 4 Conclusion

The construction of health factors directly impacts the accuracy of the remaining life prediction of bearings. In this study, an HI construction method based on the fusion

of similarity features, combined with 1-D CNN, is proposed to predict the RUL of rolling bearings. First, the time- and frequency-domain similarity features are constructed. Second, the Cori index is used to filter out the features with high monotonicity and trend from the three-domain features to construct an HI that can effectively characterize the degradation trend of rolling bearings with universal applicability. Finally, the CNN prediction model is established to predict the RUL of rolling bearings under similar operating conditions, making maximum use of the effective information in the collected signals.

## Appendix

The calculation method of 10 traditional time-domain statistical characteristics is as follows:

Serial number	Time-domain feature	Formula
1	Mean	$F_1 = \frac{1}{n} \sum_{i=1}^n x_i$
2	RMS	$F_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
3	Minimum	$F_3 = \min(x)$
4	RMSE	$F_4 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - F_1)^2}$
5	Peak-to-peak	$F_5 = \max(x) - \min(x)$
6	Kurtosis factor	$F_6 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{F_2^4}$
7	Crest factor	$F_7 = \frac{\max(x)}{F_2}$
8	Skewness	$F_8 = \frac{\sum_{i=1}^n (x_i - F_1)^3}{(n-1)F_2^3}$
9	Impulse factor	$F_9 = \frac{\max(x)}{\frac{1}{n} \sum_{i=1}^n  x_i }$
10	Wave factor	$F_{10} = \frac{F_2}{\frac{1}{n} \sum_{i=1}^n  x_i }$

**Acknowledgements** This research was partially supported by the National Natural Science Foundation of China under Grant No. 51975191.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Zhang Jiushi, Jiang Yuchen, Shimeng Wu, Li Xiang, Luo Hao, Yin Shen (2022) Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliability Eng. Syst. Safety* 221:108297. <https://doi.org/10.1016/j.res.2021.108297>
- Feng T, Guo L, Gao H et al (2022) A new time-space attention mechanism driven multi-feature fusion method for tool wear monitoring. *Int J Adv Manuf Technol*. <https://doi.org/10.1007/s00170-022-09032-3>
- Liu Lu, Song Xiao, Zhou Zhetao (2022) Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture. *Reliability Eng. Syst. Safety* 221:108330. <https://doi.org/10.1016/j.res.2022.108330>
- Lei Y, Li N, Guo L, Li N, Yan T, Lin J (2018) Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech Syst Signal Process* 104:799–834. <https://doi.org/10.1016/j.ymssp.2017.11.016>
- Park P, Jung M, Di Marco P (2020) Remaining useful life estimation of bearings using data-driven ridge regression. *Appl Sci* 10:8977. <https://doi.org/10.3390/app10248977>
- Jha RK, Swami PD (2022) Failure prognosis of rolling bearings using maximum variance wavelet subband selection and support vector regression. *J Braz Soc Mech Sci Eng* 44:49. <https://doi.org/10.1007/s40430-021-03345-2>
- Pengfei L, Jizhong T (2019) A novel bearing health indicator construction method based on ensemble stacked autoencoder. In: 2019 IEEE International Conference on Prognostics and Health Management (ICPHM). 1–9 <https://doi.org/10.1109/ICPHM.2019.8819405>
- Huang G, Hua S, Zhou Q, Li H, Zhang Y (2020) Just another attention network for remaining useful life prediction of rolling element bearings. *IEEE Access* 8:204144–204152. <https://doi.org/10.1109/ACCESS.2020.3036726>
- Gao T, Li Y, Huang X, Wang C (2020) Data-driven method for predicting remaining useful life of bearing based on bayesian-theory. *Sensors* 21:182. <https://doi.org/10.3390/s21010182>
- Han Tian, Pang Jiachen, Tan Andy C.C. (2021) Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network. *J. Manuf. Syst.* 61:576–591. <https://doi.org/10.1016/j.jmsy.2021.10.011>
- Li P, Jia X, Feng J, Zhu F, Miller M, Chen LY, Lee J (2020) A novel scalable method for machine degradation assessment using deep convolutional neural network. *Measurement* 151:107106. <https://doi.org/10.1016/j.measurement.2019.107106>
- Ji-Yan Wu, Min Wu, Chen Z, Li X-L, Yan R (2021) Degradation-aware remaining useful life prediction with LSTM autoencoder. *IEEE Trans Instrum Meas* 70:1–10. <https://doi.org/10.1109/TIM.2021.3055788>
- Luo J, Zhang X (2021) Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction. *Appl Intell* 52:1076–1091. <https://doi.org/10.1007/s10489-021-02503-2>
- Chang Hua Hu, Pei H, Si XS, Dang Bo Du, Pang ZN, Wang Xi (2020) A prognostic model based on dbn and diffusion process for degrading bearing. *IEEE Trans Industr Electron* 67(10):8767–8777. <https://doi.org/10.1109/TIE.2019.2947839>
- Wang CX, Xiong R, Tian J, Jiahuan Lu, Zhang C (2022) Rapid ultracapacitor life prediction with a convolutional neural

- network. *Appl Energy* 305:117819. <https://doi.org/10.1016/j.apenergy.2021.117819>
16. Yang Y (2021) A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Appl Energy*. <https://doi.org/10.1016/j.apenergy.2021.116897>
  17. Liu R, Yang B, Hauptmann AG (2019) Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network. *IEEE Trans Industr Inf* 16(1):87–96. <https://doi.org/10.1109/TII.2019.2915536>
  18. Ding N, Li H, Yin Z, Zhong N, Zhang Le (2020) Journal bearing seizure degradation assessment and remaining useful life prediction based on long short-term memory neural network. *Measurement* 166:108215. <https://doi.org/10.1016/j.measurement.2020.108215>
  19. Deutsch J, He D (2018) Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Trans. Syst Man, and Cybernetics: Syst* 48(1):11–20. <https://doi.org/10.1109/TSMC.2017.2697842>
  20. Huang K, Liu X, Fu S, Guo D, Xu M (2021) A lightweight privacy-preserving CNN feature extraction framework for mobile sensing. *IEEE Trans Depend Sec Comput.* 18(3):1441–1455. <https://doi.org/10.1109/TDSC.2019.2913362>
  21. Bonhage A, Eltaher M, Raab T, Breuß M, Raab A, Schneider A (2021) A modified mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the North German Lowland. *Archaeol Prospect* 28:177–186. <https://doi.org/10.1002/arp.1806>
  22. Ismael AM, Şengür A (2021) Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst Appl* 164:114054. <https://doi.org/10.1016/j.eswa.2020.114054>
  23. Chandrashekar HM, Karjigi V, Sreedevi N (2020) Spectro-temporal representation of speech for intelligibility assessment of Dysarthria. *IEEE J. Selected Topics Signal Proc.* 4(2):390–399. <https://doi.org/10.1109/JSTSP.2019.2949912>
  24. Ukil S, Ghosh S, Obaidullah SM et al (2020) Improved word-level handwritten Indic script identification by integrating small convolutional neural networks. *Neural Comput & Appl* 32:2829–2844. <https://doi.org/10.1007/s00521-019-04111-1>
  25. Ahamed P, Kundu S, Khan T et al (2020) Handwritten Arabic numerals recognition using convolutional neural network. *J Ambient Intell Human Comput* 11:5445–5457. <https://doi.org/10.1007/s12652-020-01901-7>
  26. Li D, Yang L (2021) Remaining useful life prediction of lithium battery based on sequential CNN–LSTM method. *ASME J Electrochem En Conv Stor* 18(4):041005. <https://doi.org/10.1115/1.4050886>
  27. Wang H, Bai X, Tan J (2020) Uncertainty quantification of bearing remaining useful life based on convolutional neural network. *IEEE Symposium Series on Comput. Intell. (SSCI) 2020:2893–2900*. <https://doi.org/10.1109/SSCI47803.2020.9308463>
  28. Yang B, Liu R, Zio E (2019) Remaining useful life prediction based on a double-convolutional neural network architecture. *IEEE Trans Industr Electron* 66(12):9521–9530. <https://doi.org/10.1109/TIE.2019.2924605>
  29. Ren L, Sun Y, Wang H, Zhang L (2018) Prediction of bearing remaining useful life with deep convolution neural network. *IEEE Access* 6:13041–13049. <https://doi.org/10.1109/ACCESS.2018.2804930>
  30. Cheng Cheng, Ma Guijun, Zhang Yong et al (2020) A deep learning based remaining useful life prediction approach for bearings. *IEEE/ASME Trans. Mechatr.* 25(3):1243–1254
  31. Nectoux P, Gouriveau R, Medjaher K, E. Ramasso E et al (2012) PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In: *proceedings of international conference prognostic health management*, pp 1–8
  32. J. Lee, H. Qiu, G. Yu, J. Lin (2007) Rexnord Technical Services. IMS, University of Cincinnati, “Bearing Data Set” [EB/OL]. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA: <http://ti.arc.nasa.gov/project/prognostic-data-repository>.
  33. Liang G, Naipeng L, Feng J et al (2017) A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* 240(31):98–109. <https://doi.org/10.1016/j.neucom.2017.02.045>
  34. Mosallam A, Medjaher K, Zerhouni N (2016) Data driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *J Intell Manuf* 27(5):1037–1048. <https://doi.org/10.1007/s10845-014-0933-4>
  35. Zhao Bo, Zhang X, Li H, Yang Z (2020) Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. *Knowl-Based Syst* 199(8):105971. <https://doi.org/10.1016/j.knsys.2020.105971>
  36. Cheng H, Kong X, Chen G, Wang Q, Wang R (2021) Transferable convolutional neural network based remaining useful life prediction of bearing under multiple failure behaviors. *Measurement* 168(15):108286. <https://doi.org/10.1016/j.measurement.2020.108286>
  37. Wang C, Jiang W, Yang X, Zhang S (2021) RUL prediction of rolling bearings based on a DCAE and CNN. *Appl Sci* 11(23):11516. <https://doi.org/10.3390/app112311516>
  38. Guangquan Z, Zedong J, Cong H et al (2019) Bearing fault diagnosis based on wavelet packet energy entropy and DBN. *J. Electr Measur Instrum.* 33(02):32–38
  39. Wang F, Liu X, Deng G, Xiaoguang Yu, Li H, Han Q (2019) Remaining life prediction method for rolling bearing based on the long short-term memory network. *Neural Process Lett* 50:2437–2454. <https://doi.org/10.1007/s11063-019-10016-w>
  40. Jiahang L, Xu Z (2022) Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction. *Appl Intel* 52:1076–1091. <https://doi.org/10.1007/s10489-021-02503-2>
  41. Mengqi M, Jianbo Y, Zhihong Z (2022) A sparse domain adaption network for remaining useful life prediction of rolling bearings under different working conditions. *Reliability Eng Syst Safety* 219:108259. <https://doi.org/10.1016/j.res.2021.108259>
  42. Yang W, Yao Q, Ye K et al (2020) Empirical mode decomposition and temporal convolutional networks for remaining useful life estimation. *Int J Parallel Prog* 48:61–79. <https://doi.org/10.1007/s10766-019-00650-1>
  43. Qin Y, Chen D, Xiang S, Zhu C (2021) Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings. *IEEE Trans. Ind. Inform.* 17(9):6438–6447
  44. Zhu J, Chen N, Peng W (2019) Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Trans Ind Electron* 66(4):3208–3216. <https://doi.org/10.1109/TIE.2018.2844856>
  45. Wang B, Lei Y, Li N, Yan T (2019) Deep separable convolutional network for remaining useful life prediction of machinery". *Mech Syst Signal Process.* <https://doi.org/10.1016/j.ymsp.2019.106330>
  46. Hai Q, Lee J, Jing L, Gang Y (2006) Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *J Sound Vib* 289(4–5):1066–1090. <https://doi.org/10.1016/j.jsv.2005.03.007>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.