

## Research Article

# Rotating Machinery Remaining Useful Life Prediction Scheme Using Deep-Learning-Based Health Indicator and a New RVM

Gang Zhang , Weige Liang , Bo She, and Fuqing Tian

*Department of Weaponry Engineering, Naval University of Engineering, Wuhan, China*

Correspondence should be addressed to Weige Liang; [lwinger@outlook.com](mailto:lwinger@outlook.com)

Received 4 August 2020; Revised 23 March 2021; Accepted 27 March 2021; Published 7 May 2021

Academic Editor: Carlo Rainieri

Copyright © 2021 Gang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Remaining useful life (RUL) prediction plays a significant role in developing the condition-based maintenance and improving the reliability and safety of machines. This paper proposes a remaining useful life prediction scheme combining deep-learning-based health indicator and a new relevance vector machine. First, both one-dimensional time-series information and two-dimensional time-frequency maps are input into a hybrid deep-learning structure network consisting of convolutional neural network (CNN) and long short-term memory network (LSTM) to construct health indicator (HI). Then, the prediction results and confidence interval are calculated by a new RVM enhanced by a polynomial regression model. The proposed method is verified by the public PRONOSTIA bearing datasets. Experimental results demonstrate the effectiveness of the proposed method in improving the prediction accuracy and analyzing the prediction uncertainty.

## 1. Introduction

Rotating machinery has played an essential role in industrial applications. However, most rotating machinery operates under severe working conditions which may cause different types of faults. Therefore, timely maintenance is vital for the reliability of the rotating machinery [1–4]. Industrial Internet of Things (IoT) and data-driven techniques have been transforming the scheduled maintenance into predictive maintenance. Remaining useful life (RUL) prediction is a critical component of a predictive maintenance scheme, which will reduce the cost of unplanned maintenance and enhance the reliability, safety, and availability of the rotating machinery [5].

The data-driven techniques for RUL of machinery mainly consist of two steps: health indicator (HI) construction and remaining useful life prediction based on the constructed HI [6–10]. The HI is a quantitative value that represents the degradation process of the monitored machinery, including root mean square (RMS) [11], kurtosis [12], and entropy [13]. However, many traditional HIs have poor monotonic trend, which is against the prediction accuracy. For example, the HI curve with an excellent

monotonic trend will be well correlated with the degradation process, making the RUL able to be predicted by extrapolating the historical data. However, most HI curves do not show an evident trend until severe degradation starts, which is terrible to make maintenance scheme and reduce the prediction accuracy. Besides, many HI construction methods do not consider the historical data of similar machinery that contains tremendous degradation information.

Recently, the deep-learning network has shown great potential in dealing with big data [14–16]. Motivated by the strong power of the deep learning, researchers have done many related works about remaining useful life prediction based on deep-learning method. Zhu et al. [17] presented a deep-learning method for RUL through a multiscale convolutional neural network, the input of which was time-frequency maps. Liao et al. [18] proposed an enhanced restricted Boltzmann machine with a novel regularization term to construct a new HI that is suitable for RUL. The input data of the network was time-series features. Xia et al. [19] presented a two-stage automated approach to estimate the RUL, in which the autoencoded deep neural networks were used to classify different degradation stages and a

shallow neural-networks-based regression model was used to predict the remaining useful life. Zhang et al. [20] constructed a new HI called “waveform entropy.” Then, the new HI and some traditional HIs were input into the long short-term memory network to identify the bearing remaining useful life. Al-Dulaimi et al. [21] proposed a hybrid deep neural network framework for RUL estimation. This framework used an end-to-end RUL prediction scheme, the output of which was the RUL value. Although these deep-learning-based RUL prediction methods have shown great performance, they still may be confronted with some problems: (1) The deep-learning-based RUL prediction scheme does not provide any confidence limit, which is not beneficial for people to make a maintenance scheme. (2) Most of the deep-learning networks process only one-type input data, missing some important degradation information.

Relevance vector machine (RVM) is an artificial intelligence method to learn the machinery degradation patterns from available data instead of building statistical models. It can deal with the prognostic issues of sophisticated machinery whose degradation process is challenging to be interrelated by the statistical model [22]. What is more, the RVM-based RUL prediction method also gives a confidence interval to provide uncertainty estimation and probability significance. Therefore, the RVM has been attracting more and more attention in the RUL prediction of machinery [23–25]. The relevance vector of RVM is sparsity and the hyperparameters are simple, which is beneficial for the online remaining useful life prediction [26]. However, the long-term prediction accuracy of RVM is poor. Therefore, a new RVM prediction method with the sparsity characteristic and the accurate long-term prediction ability has been proposed.

There are many sources of uncertainty in RUL prediction, such as measurement error, randomness of load, degradation feature extraction error, and modeling error, which need to be quantified and managed during the prediction process, and the confidence interval of forecast results is given to facilitate the planning of maintenance. At present, the research on the uncertainty of RUL mainly focuses on statistical data-driven methods. The statistical data-driven method is based on the theory of probability and statistics. Through statistical or random model, the probability distribution of the remaining life can be solved naturally, which is easy to quantify the uncertainty of the prediction results of the remaining useful life. Liao et al. [27] constructed a multiphase degradation model with jumps based on Wiener process, which is formulated to describe the multiphase degradation pattern. All the parameters of the model were assumed to be random variable, which led to the uncertainty of the final remaining useful life. Gao et al. [28] proposed a right-time prediction method to reduce the prognostics uncertainty of mechanical systems under unobservable degradation. Wang et al. [29] presented a probabilistic framework for remaining useful life prediction of bearings. In the proposed model, the Markov chain Monte Carlo method is investigated in posterior sampling for predicting RUL and outputting uncertainty. Most of the

existing deep-learning methods can only achieve point prediction and cannot provide uncertainty of prediction results, which greatly limits the practical application of deep learning in RUL prediction field [30]. Some researchers tried to establish RUL prediction model based on Bayesian neural network to solve the uncertainty problem [30, 31]. Although Bayesian neural network can be used to solve the uncertainty problem of RUL prediction, the disadvantage of high training cost limits the practical application of Bayesian neural network.

Although the deep-learning-based HI construction methods and RVM-based RUL prediction methods have been widely studied, the methods combining them are relatively lacking. To fill the research gap, a new RUL prediction scheme that combines a new deep-learning structure-based HI construction method and a new RVM-based RUL prediction method is proposed. The new RUL prediction scheme can not only learn the degradation process features from different types of data and get RUL prediction result automatically but also provide a confidence interval (CI).

The contributions of this paper can be summarized as follows:

- (1) A new deep-learning structure that can deal with one-dimensional time-series data and two-dimensional image data simultaneously is proposed to construct HI. The constructed HI has better performance compared with other deep-learning-based HI construction methods.
- (2) The proposed systematic approach integrates deep-learning-based HI and a new RVM-based prediction method into a framework to realize the goal of estimating RUL automatically and provide a confidence interval.
- (3) A new RVM model is proposed by combining traditional RVM and polynomial regression model, improving the long-term prediction accuracy.

The paper is organized as follows: Section 2 provides the theoretical backgrounds. Section 3 introduces a new RUL prediction scheme combining deep-learning-based HI and a new RVM. Section 4 demonstrates the effectiveness of the presented RUL estimation scheme with an experimental bearing dataset. The conclusions are presented in Section 5.

## 2. Theoretical Background

The proposed RUL prediction scheme mainly consists of four functional layers, which are time-series information learning layer, time-frequency map information learning layer, fully connected layer, and RUL prediction layer. The hybrid deep-learning structure consists of two parallel paths followed by a fully connected multilayer neural network to use the information contained in the original data fully. The two parallel paths are time-series information learning layer constructed by long short-term memory (LSTM) neural network and time-frequency map information layer made up of convolutional neural network (CNN), respectively.

The LSTM is used to extract temporal features, while the CNN is utilized to extract spatial features, which are then fused by fully connected layer to construct an HI. Finally, the HI is put into RUL prediction layer to get the remaining useful time and its confidence intervals. The theoretical background of each layer is introduced as follows.

**2.1. Time-Series Information Learning Layer.** The time-series information learning layer mainly consists of the long short-term memory network. The long short-term memory network is a state-of-the-art sequence data processing method. It develops from the recurrent neural network with a memory cell, which overcomes the problem of gradient vanishing or exploding. Figure 1 represents the hidden layer replaced by memory cells in LSTM network.

The memory cell of LSTM mainly consists of an input gate, output gate, and forget gate. Equations (1)–(6) represent the network update process at time  $t$  [32]:

$$i^t = \sigma(W_i x^t + U_i h^{t-1} + b_i), \quad (1)$$

$$o^t = \sigma(W_o x^t + U_o h^{t-1} + b_o), \quad (2)$$

$$f^t = \sigma(W_f x^t + U_f h^{t-1} + b_f), \quad (3)$$

$$a^t = \tanh(W_c x^t + U_c h^{t-1} + b_c), \quad (4)$$

$$c^t = f^t \otimes c^{t-1} + i^t \otimes a^t, \quad (5)$$

$$h^t = o^t \otimes \tanh(c^t). \quad (6)$$

In the above equations,  $i$ ,  $o$ ,  $f$ ,  $a$ ,  $c$ , and  $h$  represent input gate, output gate, forget gate, the output value of input gate, the state value of memory cell, and the output value of hidden layer, respectively.  $W_i$ ,  $U_i$ ,  $W_o$ ,  $U_o$ ,  $W_f$ ,  $U_f$ ,  $W_c$ , and  $U_c$  are the weights matrices.  $b_i$ ,  $b_o$ ,  $b_f$ , and  $b_c$  are the bias values.  $\otimes$  is point multiplication operation.

**2.2. Time-Frequency Map Information Learning Layer.** The time-frequency map information learning layer is made up of a deep convolutional neural network, consisting of a convolutional layer and a pooling layer.

In the convolutional layer, local features are generated by convolutional kernels from the feature maps. Then, the convolutional results are input into the activation layer to construct the feature maps of the current layer, whose equation process is as follows [33]:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right). \quad (7)$$

In the above equation,  $x_j^l$  is the  $j$ th feature map of the  $l$ th layer.  $x_i^{l-1}$  is the  $i$ th feature map of the  $(l-1)$ -th layer.  $k_{ij}^l$  is the convolutional kernel with size of  $S \times S$ .  $b_j^l$  is the bias of the  $l$ th layer.  $M_j$  is the feature map of the convolutional layer.  $f$  is the activation function.

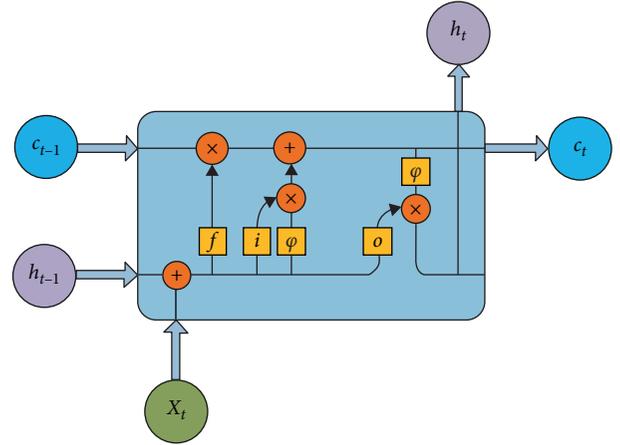


FIGURE 1: Structure of LSTM.

In the pooling layer, the feature is extracted from feature maps with the subsampling method to increase computational efficiency. The max-pooling method is given as

$$x_j^{l+1}(c, d) = \max_{0 \leq p, q < m} \{x_j^l(c * m + p, d * m + q)\}. \quad (8)$$

In the above equation,  $x_j^l$  and  $x_j^{l+1}$  are the  $j$ th input feature map of layer  $l$  and the  $j$ th output feature map of layer  $l+1$ .  $m$  is the pooling filter size,  $c$  and  $d$  are the value after convolution, and  $p$  and  $q$  are the moving step length.

**2.3. Fully Connected Layer.** The fully collected layer is added after the time-series information learning layer and time-frequency map information learning layer. The features leaning from the above two layers are flattened to construct the fully connected layer, which can be represented by the following equation:

$$O = f\left(\sum_{j=1}^d x_j^F \omega_j + b\right). \quad (9)$$

In the above equation,  $O$  is the final output value.  $x_j^F$  is the  $j$ th neuron.  $\omega_j$  represents the weights between the  $j$ th neuron and the output node.  $b$  is the bias.  $f$  is the activation function.

**2.4. RUL Prediction Layer.** This layer can filter the unwanted measurement noise and manage the uncertainty in prognostics. The RUL prediction layer is constructed with a new relevance vector machine (RVM) combining the traditional RVM method with polynomial models.

RVM is a kernel function algorithm based on Bayesian inference framework [29]. The RVM model of the given dataset  $\{x_i, t_i\}_{i=1}^N$ ,  $x_i \in R^d$ ,  $t_i \in R$ , is

$$t = \Phi \omega + \varepsilon, \quad (10)$$

where  $\omega = (\omega_0, \dots, \omega_N)^T$ , and  $\omega$  is the weight of the RVM model;  $\Phi = [\phi_1, \phi_2, \dots, \phi_N]^T$  is the design matrix;  $\phi_i(x_i) = [1, K(x_i, x_1), \dots, K(x_i, x_N)]$ ,  $i = 1, 2, \dots, N$ , and  $N$  is sample number;  $K$  is the kernel function;  $\varepsilon \sim N(0, \sigma^2)$  is

the Gaussian distributed random error with the mean of 0 and the variance of  $\sigma^2$ .

According to the Bayesian inference, the likelihood of the dataset  $p(t|x)$  satisfies  $N(t|y(x), \sigma^2)$  distribution, which can be written as

$$p(t|\omega, \sigma^2) = (2\pi\sigma^2)^{-(N/2)} \exp\left(\frac{-\|t - \Phi\omega\|^2}{(2\sigma^2)}\right). \quad (11)$$

Maximum-likelihood estimation of  $\omega$  and  $\sigma^2$  from (11) will generally lead to severe overfitting, so a Gaussian prior over the weights is defined to smooth the functions as

$$p(\omega|\alpha) = \prod_{i=0}^N N(\omega_i|0, \alpha^{-1}) = \prod_{i=0}^N \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(\frac{\omega_i^2 \alpha_i}{2}\right). \quad (12)$$

In the above equation,  $p(\omega|\alpha)$  is the Gaussian prior probability over the weights,  $\omega$  is the weight of the RVM model, and  $\alpha = \{\alpha_0, \alpha_1, \dots, \alpha_N\}$  is a vector of  $N + 1$  hyperparameters corresponding with weights  $\omega$ . These hyperparameters are the critical features of the model and are ultimately responsible for the sparsity properties.

The posterior over the unknowns could be computed with Bayes' rule, given the defined noninformative prior distribution.

$$p(\omega, \alpha, \sigma^2|t) = \frac{p(t|\omega, \alpha, \sigma^2)p(\omega, \alpha, \sigma^2)}{\int p(t|\omega, \alpha, \sigma^2)p(\omega, \alpha, \sigma^2)d\omega d\alpha d\sigma^2}. \quad (13)$$

Equation (13) cannot be computed directly, but it can be decomposed as

$$p(\omega, \alpha, \sigma^2|t) = p(\omega|t, \alpha, \sigma^2)p(\alpha, \sigma^2|t). \quad (14)$$

The posterior distribution of the weights is

$$p(\omega|t, \alpha, \sigma^2) = \frac{p(t|\omega, \sigma^2)p(\omega, \alpha)}{p(t|\alpha, \sigma^2)}. \quad (15)$$

The posterior covariance and the mean of equation (15) are

$$\Sigma = (\mathbf{A} + \sigma^{-2} \Phi^T \Phi)^{-1}, \quad (16)$$

$$\boldsymbol{\mu} = \sigma^2 \Sigma \Phi^T \mathbf{t}, \quad (17)$$

where  $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ .

As can be seen from equations (16) and (17), the values of the hyperparameter  $\alpha$  and noise variance  $\sigma^2$  need to be obtained in order to get the values of  $\boldsymbol{\mu}$  and  $\Sigma$ . The maximum-likelihood estimation method is used to obtain the estimated values of  $\alpha$  and  $\sigma^2$ . The probability distribution of the output value  $t^*$  is calculated by the new input value  $x^*$  as

$$p(t^*|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = N(t^*|y^*, \sigma_*^2), \quad (18)$$

where

$$y_* = \boldsymbol{\mu}^T \Phi(x^*), \quad (19)$$

$$\sigma_*^2 = \sigma^2 + \Phi(x^*)^T \Sigma \Phi(x^*). \quad (20)$$

In the training process, most  $\alpha$  values tend to be infinite, so the corresponding weights have posterior distributions, whose mean and variance are both zero, suggesting that those parameters and corresponding kernel functions play no role in regression analysis, which represents the sparsity of the RVM. The inputs data corresponding to the nonzero weights is called relevance vector (RV). The 95% upper and lower confidence interval can be calculated as

$$t_{\text{upper}}^* = t^* + 1.96 \sqrt{\sigma_*^2 + K(x, x^*)^T \sum K(\mathbf{x}, x^*)}, \quad (21)$$

$$t_{\text{lower}}^* = t^* - 1.96 \sqrt{\sigma_*^2 + K(x, x^*)^T \sum K(\mathbf{x}, x^*)}. \quad (22)$$

where  $t_{\text{upper}}^*$  and  $t_{\text{lower}}^*$  are the upper bound and lower bound of the predicted value  $t^*$ , respectively.  $K$  is the kernel function;  $\mathbf{x} = \{x_i\}_{i=1}^m$  represents the set of the relevance vectors,  $x$  represents the relevance vector, and  $x^*$  is the input data.

The polynomial models are suitable for long-term RUL prediction. Polynomial regression belongs to the least-square curve fitting family. Specifically speaking, it estimates the coefficients of a polynomial function to approximate the curve closely. The mathematical expression of polynomial regression is as follows:

$$y = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n, \quad (23)$$

where  $y$  is the response variable,  $x$  is the predictor variable, and  $a_0, a_1, a_2, \dots, a_n$  are model coefficients that can be estimated by curve fitting methods.

In this paper, we take advantage of the RVM and polynomial model, the response variable  $y$  is RV, and  $x$  is the corresponding running time. The coefficients of the polynomial model  $a_0, a_1, a_2, \dots, a_n$  are determined by  $x$  and relevance vector  $y$  of the test bearing. Then the predicted value is calculated by equation (23). The variance value  $\sigma_*^2$  is calculated by equation (20) and the 95% upper and lower confidence interval can be calculated by equations (21) and (22).

**2.5. Time-Frequency Analysis.** In the process of performance degradation of rolling bearings, vibration acceleration signals have nonstationary characteristics. Time-frequency analysis includes both time-domain information and frequency-domain information, which can effectively characterize the characteristics of nonstationary signals. Continuous wavelet transform is a time-frequency analysis method commonly used in state monitoring of rotating machinery. The calculation formula is as follows:

$$\begin{aligned} U(\alpha, \beta) &= \langle x(t), \Psi_{\alpha, \beta} \rangle, \\ &= \int_{-\infty}^{\infty} x(t) \bar{\Psi}_{\alpha, \beta}(t) dt \\ &= \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{\alpha}} \bar{\Psi}\left(\frac{t - \beta}{\alpha}\right) dt, \end{aligned} \quad (24)$$

where  $\alpha$  is the scale parameter;  $\beta$  is the transformation parameter;  $x(t)$  is the original vibration acceleration signal;  $\psi(t) \in L^2(R)$  is the mother wavelet function;  $\bar{\psi}(t)$  is the complex conjugate of  $\psi(t)$ . There is a standard or universal method for the selection of the mother wavelet function. In this paper, Morlet wavelet, which is similar to the impact signal of rolling bearing, is chosen as the mother wavelet. After the continuous wavelet transform, the one-dimensional vibration acceleration signal is mapped to the two-dimensional coefficient matrix, and the time-frequency diagram of the vibration signal is obtained.

### 3. RUL Scheme Combining Deep-Learning-Based HI and a New RVM

A hybrid deep-learning structure that can learn temporal features and spatial features simultaneously is proposed to take advantage of mutual information from multidimensional features for degradation assessment and RUL prediction. What is more, the training set is constructed with historical whole lifetime monitoring data. Then the training set consisting of different HI curves is used to train the RVM. The sparsity of RVM regression is highly dependent on the choice of kernel functions. The common kernel functions are classified into local kernels and global kernels. In local kernels, only the data points that are close or in proximity of each other have an effect on the kernel values.

In contrast, a global kernel allows data points that are far away from each other to affect the kernel values as well. Furthermore, the common global kernels are polynomial function, spline function, and so forth [34]. Different types of kernels perform distinctly in the interpolation and extrapolation ability. A multikernel RVM-based prediction method is proposed to make full use of the superiorities of different kernels by combining them with the particle swarm optimization (PSO) algorithm. Figure 2 shows the proposed RUL scheme.

First, the time-series information including time-domain features, frequency-domain features, and time-frequency map information of the whole lifetime is extracted from the original vibration signal. Different information is processed by different information learning layer. Then, a fully connected layer is used to combine different features learned from the time-series information learning layer and time-frequency map information learning layer together. The HI is constructed by a three-layer neural network using combined information. Finally, the constructed HI curve is used to predict the RUL with the RUL prediction layer, which is constructed by the RVM and polynomial model.

At the inspection time  $T_k$ , the future HI can be predicted with the constructed polynomial curve. When the polynomial curve reached the failure threshold, the bearing is considered to fail. According to the concept of the first hitting time [31], the RUL of the bearing can be defined as

$$\text{RUL}(T_k) = \inf\{t: f(t + T_k) \geq \theta\}. \quad (25)$$

In the above equation,  $\text{RUL}(T_k)$  is the remaining useful life at inspection time  $T_k$ ,  $f(t + T_k)$  is the predicted HI at the time  $t + T_k$ , and  $\theta$  is the failure threshold.

## 4. Experimental Results and Analysis

In this section, the run-to-failure data acquired from accelerated degradation tests of rolling element bearings are used to verify the effectiveness and superiority of the proposed RUL scheme in practical applications. The experimental data comes from PROGNOSTIA in the IEEE PHM 2012 Data Challenge [35]. The experimental platform mainly consists of three parts, a rotatory part, a degradation generation part, and a signal acquisition part, which is shown in Figure 3.

*4.1. Data Description.* In this experiment, 17 rolling element bearings working under three different conditions are tested. The experimental conditions are listed in Table 1. Under each condition, two bearings' data are used as a training dataset, while others are testing datasets, which are listed in Table 2. The acceleration sensor is installed in the outer layer of the rolling element bearing. The sampling frequency is 25.6 kHz and every sampling process includes 2560 points. The sampling process is repeated every 10 s.

*4.2. HI Construction.* The whole lifetime data of the first bearing is selected to be analyzed. The acceleration signal on the horizontal direction shown in Figure 4 shows that the vibration amplitude increases as the experiment cycle increases, but it is hard to determine accurately when the incipient fault occurred. Therefore, different features are extracted from the original signal including 10 time-domain features, 12 frequency-domain features, and 1 time-frequency domain feature, which are listed in Table 3.

The training data can be presented by  $I_{\text{train}} = \{x_t, y_t\}$ , in which  $x_t \in R^{N \times N}$  is the time-frequency map with size  $N \times N$  at time  $t$ .  $y_t \in [0, 1]$  represents the performance degradation degree of bearing at time period  $t$ ,  $y_t = (t/T)$ , and  $T$  is the life cycle period of bearing for training. It was verified that the relationship between training data label and running cycle does not affect the final result of health indicator construction. Therefore, the linear model is selected here to construct the training dataset label. Here, bearing 1\_1 is taken as an example. The time period when the bearing fails completely is 2800 period, and the degradation degree corresponding to the 1400 period is  $y_t = (1400/2800) = 0.5$ .

In the proposed deep learning network, the convolution structure mainly refers to classical AlexNet network and time-series information learning layer constructed by stacking three-LSTM-layer network. The literature shows that the network structure can effectively extract the characteristics of time-series data. The CNN and LSTM connected layer is used to connect the information extracted by time-series information learning layer and time-frequency map information learning layer together, which can get degradation information comprehensively. Finally, a fully connected layer is constructed to output the final result. Detailed network parameters can be seen in Table 4.

*4.3. RUL Prediction.* The HI constructed in Section 3 is used to predict the remaining useful life by the RUL prediction layer. Figure 5 shows the complete degradation prediction

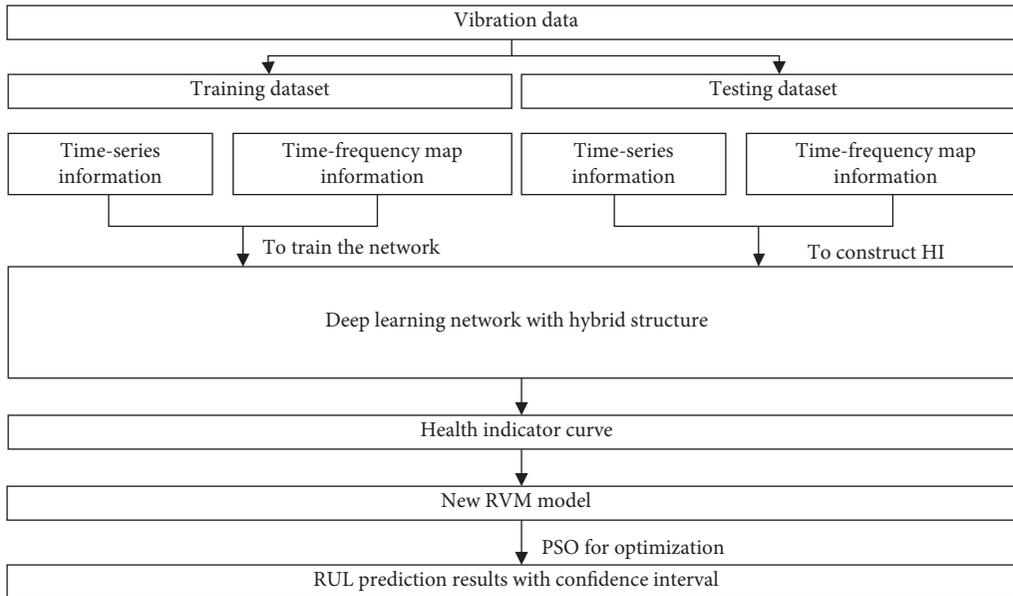


FIGURE 2: The proposed RUL scheme.

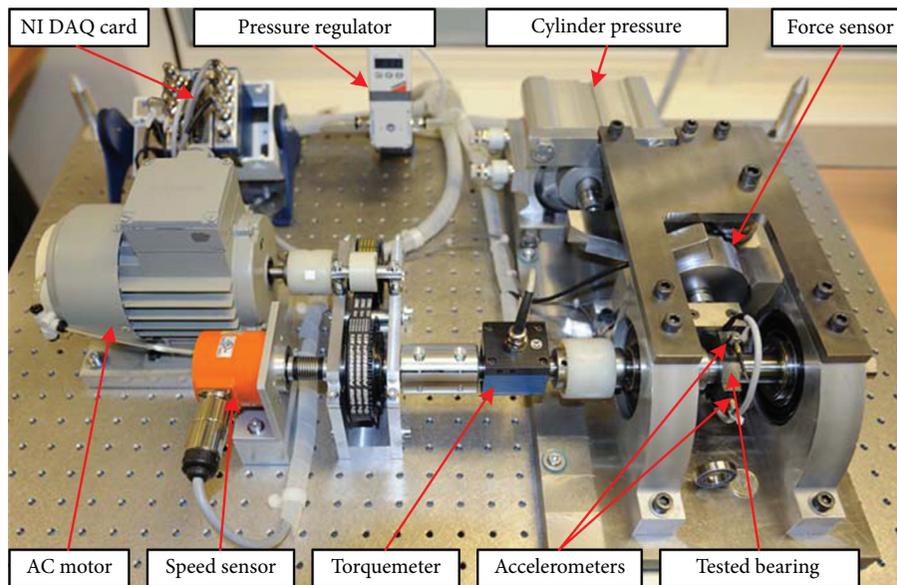


FIGURE 3: The experiment platform.

TABLE 1: The experimental conditions.

Working condition	Rotate speed (rpm)	Load (N)
1	1800	4000
2	1650	4200
3	1500	5000

process of bearing1\_5 at inspection time  $T = 2302$  running period to describe the prognostics procedure in detail. First, all the HIs constructed with hybrid deep-learning networks, which are shown by hollow blue dots in Figure 5, are input into RVM to perform regression analyses with different kernel parameter values. Next, the kernel parameter value is selected by the PSO algorithm. As is shown in Figure 5, the relevance

TABLE 2: Dataset.

Dataset	Condition 1	Condition 2	Condition 3
Training set	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Testing set	Bearing1_3	Bearing2_3	
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	Bearing3_3
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	

vector is calculated by RVM with the optimized kernel parameter value. The polynomial model is constructed by fitting the RVs. Subsequently, the HI values at the future running

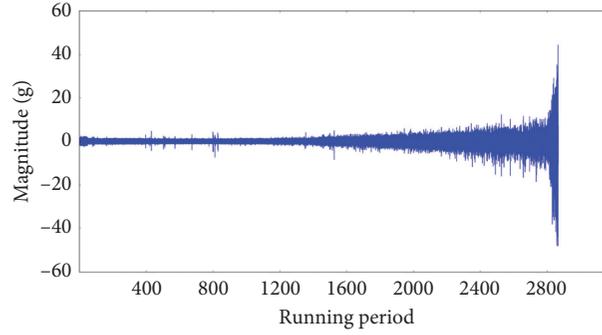


FIGURE 4: The acceleration signal on the horizontal direction.

TABLE 3: Features extracted from the vibration signal.

Type	Feature
Time-domain features	A1: root mean square
	A3: peak-peak value
	A5: peak factor
	A7: clearance factor
	A9: standard deviation
Frequency-domain features	B1: mean value
	B3: skewness
Time-frequency features	B5–B12: entropy of different frequency band
	C1: time-frequency map
	A2: kurtosis
	A4: shape factor
	A6: impulse factor
	A8: mean absolute
	A10: crest factor
	B2: standard deviation
	B4: kurtosis

TABLE 4: Parameters of hybrid deep learning network.

Network layer	Parameters				
	Input picture size	Number of channels	Convolution kernel size	Step size	Number of nodes
Input layer	[100 × 100]	3	—	—	—
Convolutional layer 1	—	96	[11 × 11]	4	—
Pooling layer 1	—	96	[3 × 3]	2	—
Convolutional layer 2	—	256	[5 × 5]	1	—
Pooling layer 2	—	256	[3 × 3]	2	—
Convolutional layer 3	—	384	[3 × 3]	1	—
Convolutional layer 4	—	384	[3 × 3]	1	—
Convolutional layer 5	—	256	[3 × 3]	1	—
Pooling layer 5	—	256	[3 × 3]	2	—
CNN flatten layer	—	7424 (6400 + 1024)	—	—	—
LSTM layer 1	—	—	—	—	80
LSTM layer 2	—	—	—	—	60
LSTM layer 3	—	—	—	—	30
LSTM flatten layer	—	—	—	—	30
CNN + LSTM connected layer	—	—	—	—	7454 (7424 + 30)
Fully connected layer 1	—	4096	—	—	—
Fully connected layer 2	—	1000	—	—	—
Output layer	—	1	—	—	—

periods are predicted with the constructed polynomial model. When the predicted HI values reached the predefined failure threshold, the bearing is considered to be a failure. The RUL and the confidence interval can be computed by equation (22).

## 5. Results and Analysis

Figure 6 represents the HIs of bearing 1\_3 constructed with the different methods to illustrate the superiority of the proposed method compared with the deep-learning method

of single-structure network and traditional HI. Figures 6(a)–6(d) are traditional HIs, which are commonly used in the RUL prediction domain. It can be seen in Figure 6(a) that the RMS increases significantly at the end of the experiment. The kurtosis and crest factor are sensitive to the incipient degradation process with more background noise, which cannot present the degradation process of the whole lifetime clearly as shown in Figures 6(b) and 6(c). In Figure 6(d), the peak-peak-value-based HI has similar trends to RMS, which is insensitive to the incipient degradation process. RUL

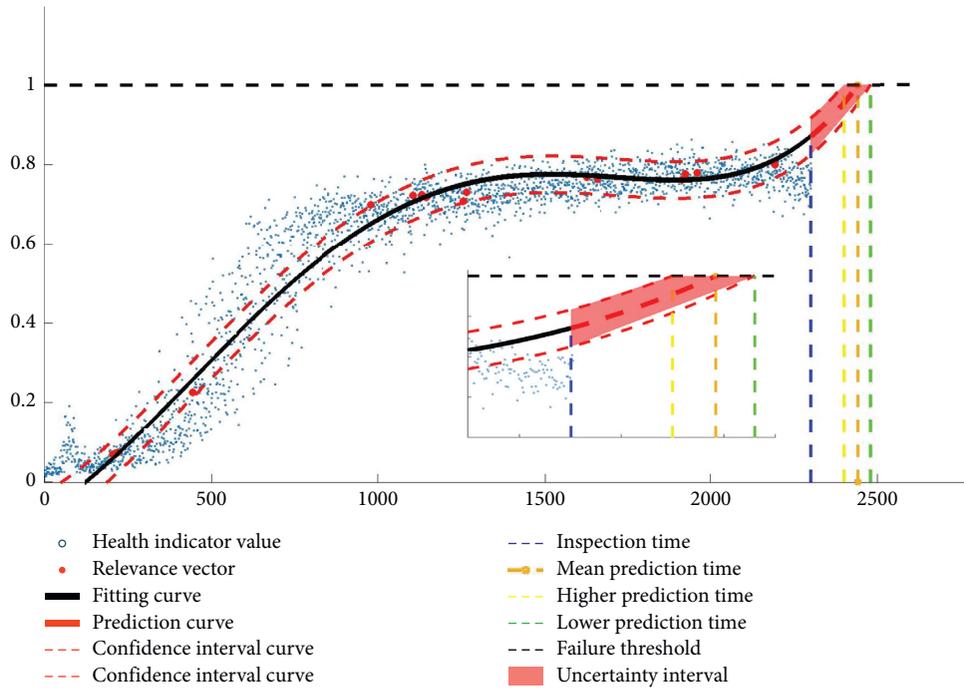


FIGURE 5: RUL prediction result of bearing 1\_5 at inspection time  $T = 2302$ .

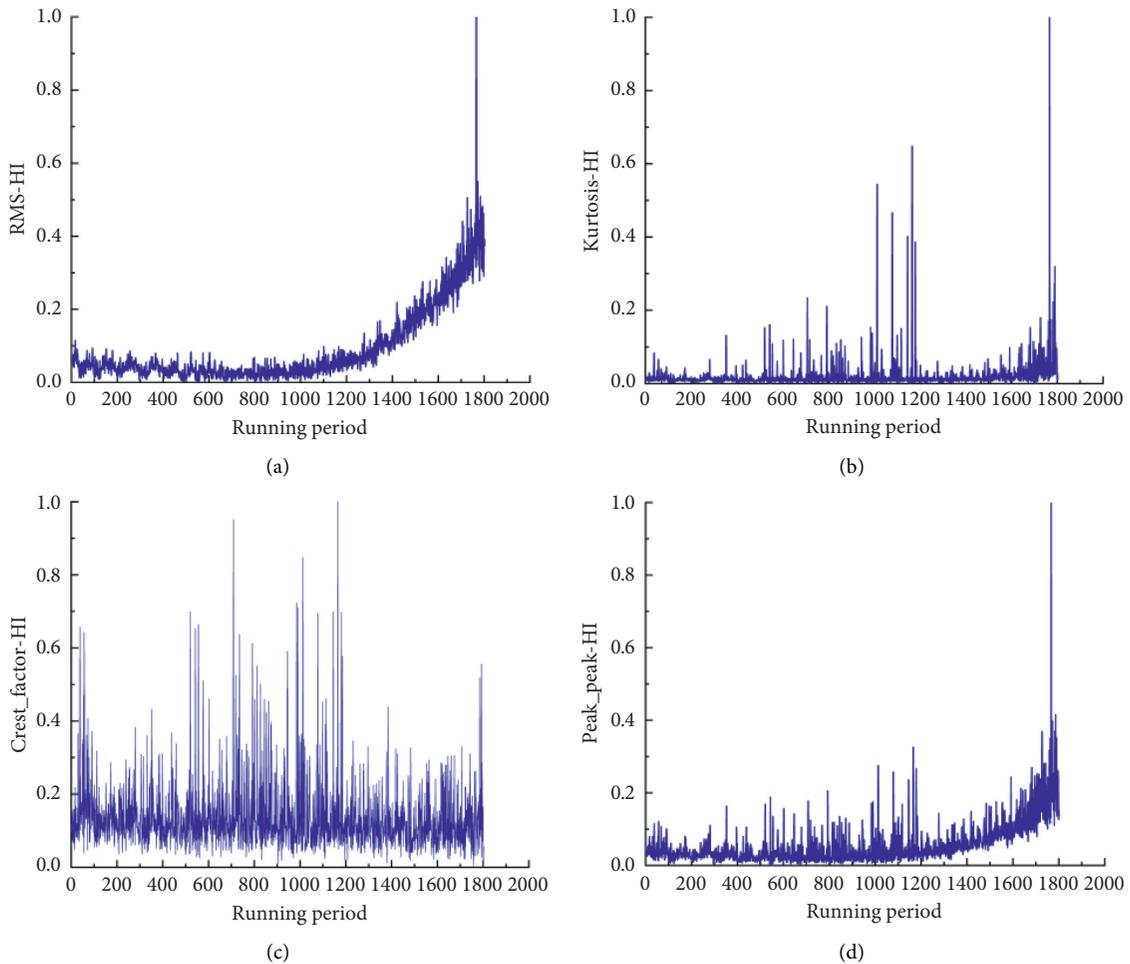


FIGURE 6: Continued.

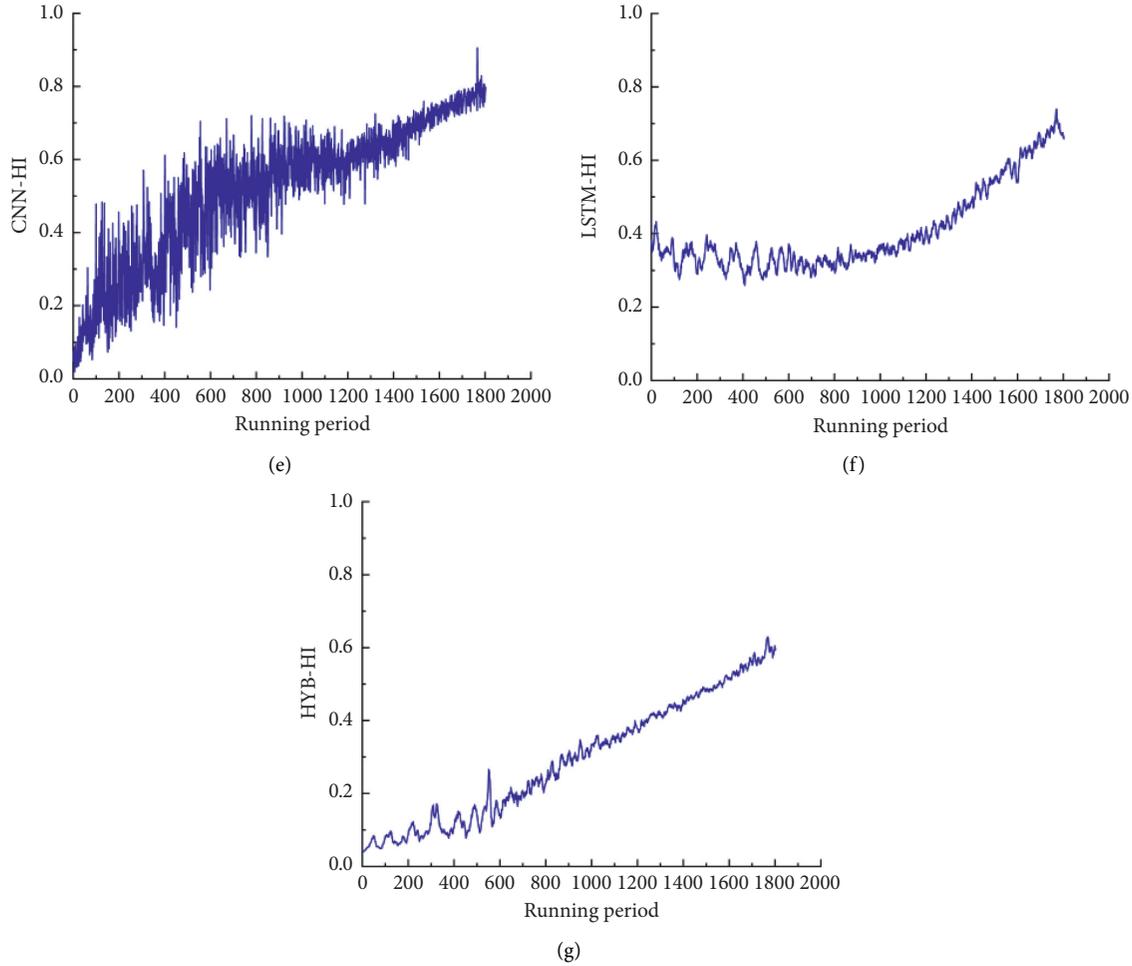


FIGURE 6: The HI curve constructed by different method. (a) Root-mean-square-based HI. (b) Kurtosis-based HI. (c) Crest-factor-based HI. (d) Peak-peak-based HI. (e) CNN-based HI. (f) LSTM-based HI. (g) Hybrid-structure-based HI.

prediction based on these HIs cannot provide timely maintenance suggestions.

Figures 6(e)–6(g) described the deep-learning-based HI suggesting better monotonicity and trendability than the traditional HI. However, the single deep-learning structure cannot fully utilize the information contained in the original acceleration signal. It can be seen in Figure 6(e) that the HI constructed by the deep CNN could present the degradation process, but there exists too much background noise at the earlier stage, since the single CNN structure cannot learn and distinguish the time-frequency picture effectively at the earlier degradation stage. From Figure 6(f), it can be concluded that the HI constructed with LSTM is almost constant at the earlier degradation stage because the single LSTM structure is unable to learn the difference of the earlier degradation features between the training set and testing set. Figure 6(g) is the HI curve constructed with the proposed method, suggesting that the HI constructed with deep hybrid structure has better linearity and less background noise, which are beneficial to promoting prognostic accuracy.

Figure 7 shows the RUL prediction results of bearing 1\_3 to illustrate the influence of different deep learning structure

on the predicted results further. Figure 7(a) shows the RUL prediction results of LSTM-based HI, where the predicted RUL time is 4800 s and the 95% confidence interval is [2510, 7090] s. It can be seen in Figure 7(a) that the relevance vector in the earlier stage affects the polynomial type, which lowers the prediction accuracy. Figure 7(b) shows the RUL prediction results of CNN-based HI, where the predicted RUL time is 5190 s and the 95% confidence interval is [0, 10830] s. A larger confidence interval brings about too much uncertainty to the RUL results, which is terrible for making maintenance plan. The RVs in the middle period influence affect confidence interval. Figure 7(c) shows the RUL prediction results of hybrid-structure-based HI, where the predicted RUL time is 5600s, and the 95% confidence interval is [5398, 5803] s. The actual RUL time is 5730 s, which has been included in the confidence interval. What is more, the confidence interval is narrow, leading to the low uncertainty of the prediction results.

The RUL prediction results of all the test bearings are shown in Figure 8. It can be seen from Figure 8 that the method proposed in this paper can effectively predict the performance degradation trend and obtain a relatively

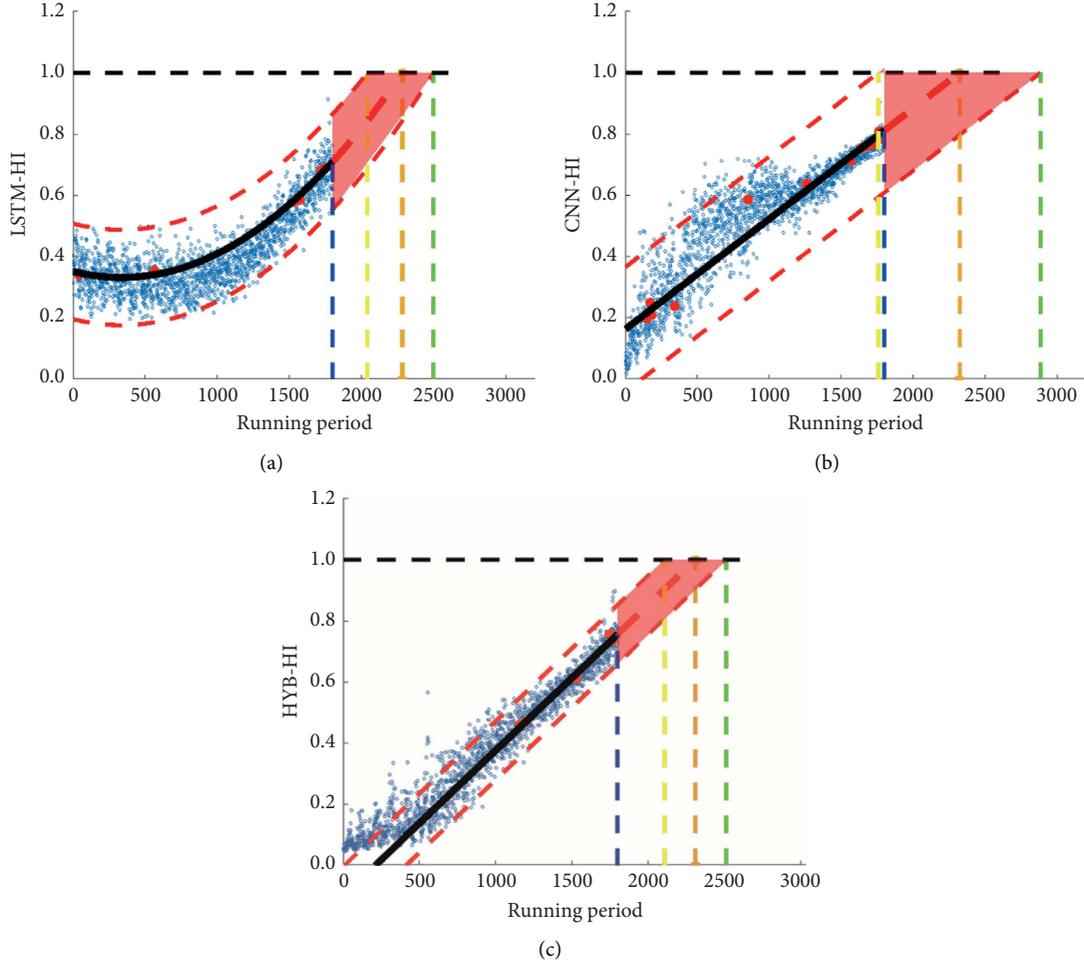


FIGURE 7: The RUL prediction results of bearing1\_3. (a) LSTM-based HI for RUL. (b) CNN-based HI for RUL. (c) Hybrid-structure-based HI for RUL.

narrow confidence interval. It can be concluded from Figure 8 that there exist 3 types of degradation curve: linear type (such as bearing1\_3, bearing1\_4, bearing1\_7, bearing2\_4, and bearing2\_5), exponential type (such as bearing2\_2), and S-shape type (such as bearing1\_5, bearing1\_6, bearing2\_1, bearing2\_3, and bearing2\_6). The polynomial model can effectively fit different types of degradation curve.

Different RUL prediction methods are compared with six other studies with the same dataset to illustrate the superiority of the proposed scheme, which are listed in Table 4. Column 1 shows the testing bearings. The prediction starting time is shown in column 2. For each testing bearing, the actual and predicted RUL times are displayed in columns 3 and 4, respectively. The predicted errors of the proposed method are shown in the final column, and the six comparative studies are shown in columns 5 to 10. The mean and SD of the percent errors and the scoring metrics are shown

in the last three rows. A scoring function to evaluate the final prediction results is defined as follows:

$$\text{Score} = \frac{1}{11} \sum_{i=1}^{11} A_i, \quad (26)$$

where

$$A_i = \begin{cases} \exp\left(-\ln(0.5) * \left(\frac{Er_i}{5}\right)\right), & \text{if } Er_i \leq 0, \\ \exp\left(+\ln(0.5) * \left(\frac{Er_i}{20}\right)\right) & \text{if } Er_i > 0. \end{cases} \quad (27)$$

$A_i$  is the prediction score of the  $i$ th test bearing and  $Er_i$  is the percent error of RUL prediction results for the  $i$ th testing dataset, and it can be calculated as

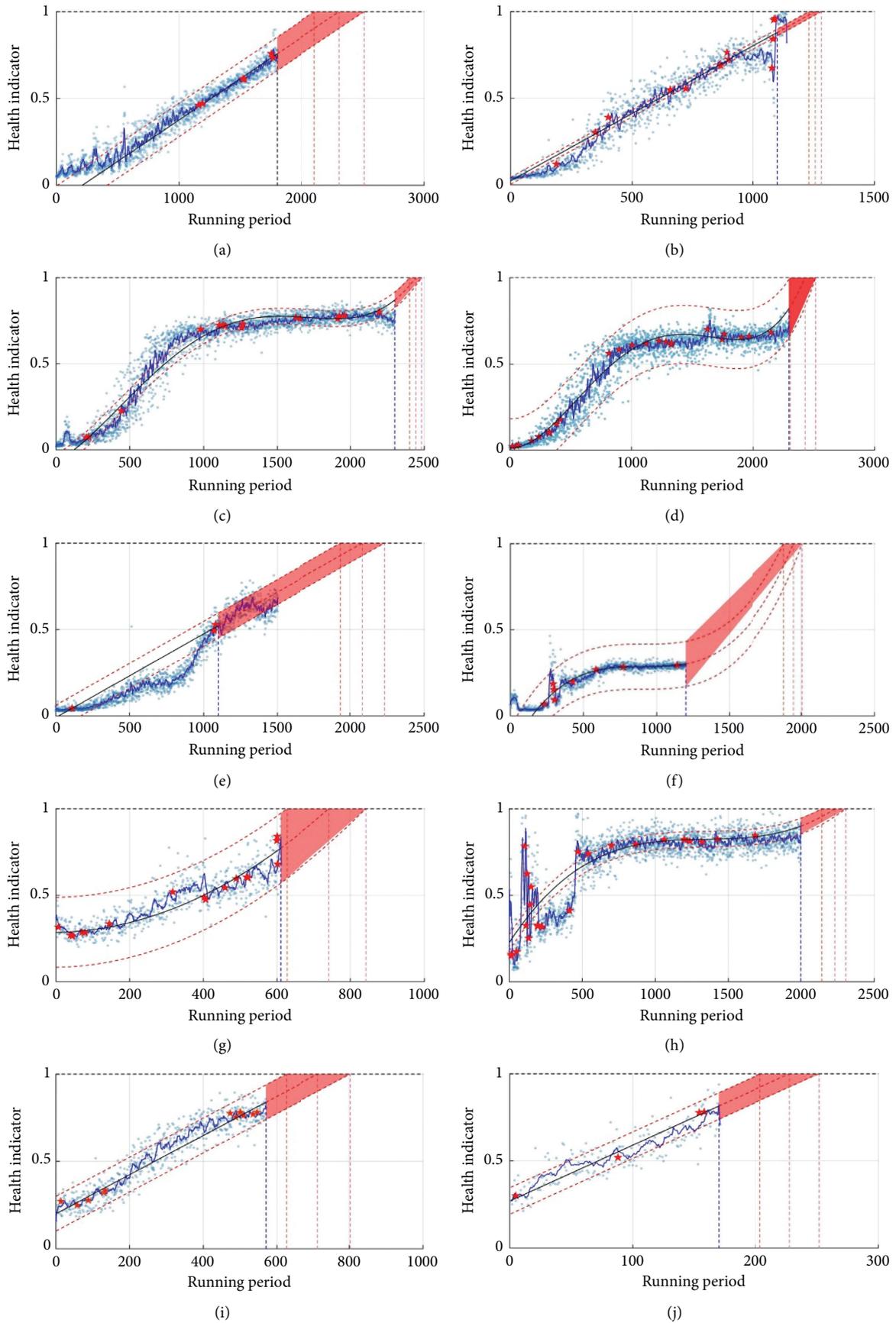


FIGURE 8: Continued.

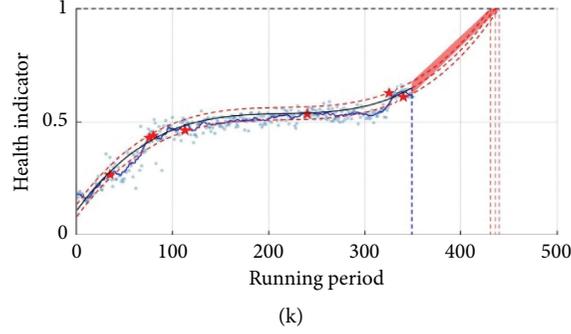


FIGURE 8: The RUL prediction results of all the test bearings. (a) Bearing 1\_3 performance degradation trend prediction. (b) Bearing 1\_4 performance degradation trend prediction. (c) Bearing 1\_5 performance degradation trend prediction. (d) Bearing 1\_6 performance degradation trend prediction. (e) Bearing 1\_7 performance degradation trend prediction. (f) Bearing 2\_3 performance degradation trend prediction. (g) Bearing 2\_4 performance degradation trend prediction. (h) Bearing 2\_5 performance degradation trend prediction. (i) Bearing 2\_6 performance degradation trend prediction. (j) Bearing 2\_7 performance degradation trend prediction. (k) Bearing 3\_3 performance degradation trend prediction.

TABLE 5: The RUL prediction results of different methods.

	Current time (s)	Actual RUL (s)	Predicted RUL (s)	Er (%)						
				Sutrisno	Hong	Lei	Guo	Yoo	Wang	Proposed method
1_3	18010	5730	5600	37	-1.04	-0.35	43.28	1.05	-1.05	2.27
1_4	11380	339	320	80	-20.94	5.6	67.55	20.35	-17.99	5.60
1_5	23010	1610	1410	9	-278.26	100	-22.98	11.18	21.74	12.42
1_6	23010	1460	1300	-5	19.18	28.08	21.23	34.93	6.16	10.96
1_7	15010	7570	9270	-2	-7.13	-19.55	17.83	29.19	7.79	-22.46
2_3	12010	7530	7460	64	10.49	-20.19	37.84	57.24	43.03	0.99
2_4	6110	1390	1310	10	51.8	8.63	-19.42	-1.44	1.44	5.76
2_5	20010	3090	2290	-440	28.8	23.3	54.37	-0.65	18.77	25.89
2_6	5710	1290	1430	49	-20.93	58.91	-13.95	-42.64	2.33	-10.85
2_7	1710	580	570	-317	44.83	5.17	-55.17	8.62	-3.45	1.72
3_3	3510	820	850	90	-3.66	40.24	3.66	-1.22	13.41	-3.66
Mean				100.27	44.28	28.18	32.48	18.96	12.47	9.32
SD				173.28	90.29	35.41	37.57	25.59	15.90	12.57
Score				0.31	0.36	0.43	0.26	0.57	0.62	0.64

$$Er_i = \frac{\text{ActRUL}_i - \text{RUL}_i}{\text{ActRUL}_i} \times 100\%, \quad (28)$$

where  $\text{ActRUL}_i$  represents the actual RUL of test bearing  $i$  and  $\text{RUL}_i$  represents the predicted RUL results of test bearing  $i$ .

The RUL prediction method proposed by Sutrisno et al. [36] has predicted the RUL based on vibration frequency signature anomaly detection and survival time ratio. However, the anomaly detection time point is decided by subjective criteria, and the prediction errors are large in the table. Hong et al. [37] have constructed the packet-EMD and SOM-based HI to predict RUL to improve the RUL accuracy compared to the previous work. However, it requires extracting more than 100 features to construct the HI, which is time-consuming. Lei et al. [38] have proposed a new HI construction method based on weighted minimum quantization error (WMQE) to predict RUL of bearings. These three methods use feature extraction, selection, and fusing, which rely on manual experience and time consumption.

Guo et al. [40] constructed a deep-learning-based method to construct HI with multiple features from the time domain, frequency domain, and time-frequency domain. This method showed superiority over SOM-based HI construction method, but it has a lower accuracy than Lei et al.'s. Yoo [41] has proposed a new method to construct the HI with CNN and the Gaussian process regression method for RUL prediction. This method improves the prediction accuracy and efficiency. Si et al. [42] has constructed HI with wavelet packet decomposition, empirical mode decomposition, and self-organizing map and used RVM combined with exponential degradation model to predict RUL, which improves the RUL accuracy effectively. However, the HI construction process is complicated and time-consuming.

In Table 5, the proposed method shows the lowest percent errors and low deviation, proving that the model is accurate and reliable on every tested bearing. What is more, this method does not design a sophisticated feature extraction algorithm based on human experience and realizes the intelligent RUL prediction. The performance

degradation processes of the test bearings are not consistent. Compared with the other methods, the prediction accuracy of the proposed method is not the highest in all test bearings, but the average error and score of the prediction results are the best. In the next step, the performance degradation process of different test bearings will be studied in depth to further improve the prediction accuracy of the remaining useful life of each type of test bearings.

## 6. Conclusions

This paper proposes a new RUL prediction scheme combining deep learning and a new RVM method. Firstly, different types of degradation data are input into the deep-learning network with a hybrid structure to construct the health indicator. Then the new RVM model consisting of RVM and a polynomial model is used to predict the RUL and calculate confidence interval. Finally, the proposed method is compared with different RUL prediction methods to verify the effectiveness.

The proposed deep-learning network with a hybrid structure could learn from different types of degradation data. The constructed health indicator curve has better monotonicity and trendability than the single-structure deep-learning network, such as CNN and LSTM. The RVM is widely used in RUL prediction. On the one hand, the RVM could reduce the redundancy of the degradation curve to enhance the prediction accuracy. On the other hand, the prediction results of RVM are profoundly affected by kernel function and the long-term prediction ability is reduced. The proposed method retains the advantage of RVM and overcomes the disadvantage by combining the polynomial model with RVM. The final RUL prediction results show that the proposed method can enhance prediction accuracy and narrow down the confidence interval.

Although the proposed RUL scheme improves the prediction results, it is time-consuming. In future work, it is expected to raise the computational efficiency by researching a better deep-learning structure.

## Data Availability

The experimental data are obtained from PROGNOSTIA in the IEEE PHM 2012 Data Challenge by Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebel-Morello, B.; Zerhouni, N. "PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests," presented at the IEEE Int. Conf. Prognostics Health Manage., Denver, CO, USA, 2012, 1–8.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This work has been supported in part by the National Natural Science Foundation of China (61640308).

## References

- [1] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: a systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
- [2] D. Wang, K.-L. Tsui, and Q. Miao, "Prognostics and health management: a review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.
- [3] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.
- [4] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: methods and applications," *Journal of Manufacturing Systems*, vol. 48, pp. 144–156, 2018.
- [5] Y. Lin, X. Li, and Y. Hu, "Deep diagnostics and prognostics: an integrated hierarchical learning framework in PHM applications," *Applied Soft Computing*, vol. 72, pp. 555–564, 2018.
- [6] H. Li, Y. Wang, B. Wang, J. Sun, and Y. Li, "The application of a general mathematical morphological particle as a novel indicator for the performance degradation assessment of a bearing," *Mechanical Systems and Signal Processing*, vol. 82, pp. 490–502, 2017.
- [7] L. Tao, C. Yang, Y. Cheng, C. Lu, and M. Ragulskis, "Machine component health prognostics with only truncated histories using geometrical metric approach," *Mechanical Systems and Signal Processing*, vol. 113, pp. 168–179, 2017.
- [8] A. Rai and S. H. Upadhyay, "Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering," *Mechanical Systems and Signal Processing*, vol. 93, pp. 16–29, 2017.
- [9] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 924–932, 2016.
- [10] Y. Qian, R. Yan, and R. X. Gao, "A multi-time scale approach to remaining useful life prediction in rolling bearing," *Mechanical Systems and Signal Processing*, vol. 83, pp. 549–567, 2017.
- [11] Y. Lei, N. Li, and J. Lin, "A new method based on stochastic process models for machine remaining useful life prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 12, pp. 2671–2684, 2016.
- [12] Z.-X. Zhang, X.-S. Si, and C.-H. Hu, "An age- and state-dependent nonlinear prognostic model for degrading systems," *IEEE Transactions on Reliability*, vol. 64, no. 4, pp. 1214–1228, 2015.
- [13] K. Rodney, Singleton, E. G. Strangas, and S. Aviyente, "Extended kalman filtering for remaining-useful-life estimation of bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 1781–1790, 2015.
- [14] J. Zhang, J. Yan, D. Infield, Y. Liu, and F.-s. Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Applied Energy*, vol. 241, pp. 229–244, 2019.
- [15] B. Wang, Y. Lei, N. Li, and T. Yan, "Deep separable convolutional network for remaining useful life prediction of machinery," *Mechanical Systems and Signal Processing*, vol. 134, Article ID 106330, 2019.
- [16] A. Listou Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov, and H. Zhang, "Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture,"

- Reliability Engineering & System Safety*, vol. 183, pp. 240–251, 2019.
- [17] J. Zhu, N. Chen, and W. Peng, “Estimation of bearing remaining useful life based on multiscale convolutional neural network,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3208–3216, 2019.
  - [18] L. Liao, W. Jin, and R. Pavel, “Enhanced restricted Boltzmann machine with prognosability regularization for prognostics and health assessment,” *IEEE Transactions on Industrial Electronics*, vol. 63, pp. 7076–7083, 2017.
  - [19] M. Xia, T. Li, T. Shu, J. Wan, C. W. de Silva, and Z. Wang, “A two-stage approach for the remaining useful life prediction of bearings using deep neural networks,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3703–3711, 2019.
  - [20] B. Zhang, S. Zhang, and W. Li, “Bearing performance degradation assessment using long short-term memory recurrent network,” *Computers in Industry*, vol. 106, pp. 14–29, 2019.
  - [21] A. Al-Dulaimi, S. Zabihi, A. Asif, and A. Mohammadi, “A multimodal and hybrid deep neural network model for Remaining Useful Life estimation,” *Computers in Industry*, vol. 108, pp. 186–196, 2019.
  - [22] A. Widodo and B.-S. Yang, “Application of relevance vector machine and survival probability to machine degradation assessment,” *Expert Systems with Applications*, vol. 38, no. 3, pp. 2592–2599, 2011.
  - [23] F. Chen, Y. Yang, B. Tang, B. Chen, W. Xiao, and X. Zhong, “Performance degradation prediction of mechanical equipment based on optimized multi-kernel relevant vector machine and fuzzy information granulation,” *Measurement*, vol. 151, pp. 1–12, 2020.
  - [24] M. Yan, X. Wang, B. Wang, M. Chang, and I. Muhammad, “Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model,” *ISA Transactions*, vol. 98, pp. 471–482, 2019.
  - [25] D. Kong, Y. Chen, N. Li, C. Duan, L. Lu, and D. Chen, “Relevance vector machine for tool wear prediction,” *Mechanical Systems and Signal Processing*, vol. 127, pp. 573–594, 2019.
  - [26] Y. Chang and H. Fang, “A hybrid prognostic method for system degradation based on particle filter and relevance vector machine,” *Reliability Engineering & System Safety*, vol. 186, pp. 51–63, 2019.
  - [27] G. B. Liao, H. P. Yin, M. Chen et al., “Remaining useful life prediction for multi-phase deteriorating process based on Wiener process,” *Reliability Engineering & System Safety*, vol. 27, pp. 1–11, 2021.
  - [28] Z. Y. Gao, J. W. Li, and R. X. Wang, “Prognostics uncertainty reduction by right-time prediction of remaining useful life based on hidden Markov model and proportional hazard model,” *Eksplotacja I Niezawodnosć-Maintenance and Reliability*, vol. 23, no. 1, pp. 154–164, 2021.
  - [29] T. Wang, Z. Liu, and N. Mrad, “A probabilistic framework for remaining useful life prediction of bearings,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.
  - [30] W. W. Peng, Z. S. Ye, and N. Chen, “Bayesian deep-learning-based health prognostics toward prognostics uncertainty,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2283–2293, 2019.
  - [31] M. Kim and K. Liu, “A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics,” *IJSE Transactions*, vol. 53, no. 3, pp. 326–340, 2020.
  - [32] H. Sepp and S. Jürgen, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 1997.
  - [33] Y. Cun, B. Boser, J. Denker et al., “Handwritten digit recognition with a back-propagation network,” *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 2, pp. 396–404, 1990.
  - [34] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, “Ensemble of relevance vector machines and boosted trees for electricity price forecasting,” *Applied Energy*, vol. 250, pp. 540–548, 2019.
  - [35] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, and N. Zerhouni, “PRONOSTIA: an experimental platform for bearings accelerated degradation tests,” in *Proceedings of the IEEE International Conference on Prognostics Health Management*, pp. 1–8, Denver, CO, USA, March 2012.
  - [36] E. Sutrisno, H. Oh, A. S. S. Vasana, and M. Pecht, “Estimation of remaining useful life of ball bearings using data driven methodologies,” in *Proceedings of the 2012 IEEE Conference on Prognostics and Health Management*, pp. 1–7, Denver, CO, USA, June 2012.
  - [37] S. Hong, Z. Zhou, E. Zio, and K. Hong, “Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method,” *Digital Signal Processing*, vol. 27, pp. 159–166, 2014.
  - [38] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, “A model-based method for remaining useful life prediction of machinery,” *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1314–1326, 2016.
  - [39] B. Wang, Y. Lei, N. Li, and N. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings,” *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 1–12, 2019.
  - [40] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, “A recurrent neural network based health indicator for remaining useful life prediction of bearings,” *Neurocomputing*, vol. 240, pp. 98–109, 2017.
  - [41] Y. Yoo and J. G. Baek, “A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network,” *Applied Sciences*, vol. 8, pp. 1–17, 2018.
  - [42] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, “Remaining useful life estimation - a review on the statistical data driven approaches,” *European Journal of Operational Research*, vol. 213, no. 1, pp. 1–14, 2011.