

Research Article

Evaluation and Prediction Method of Rolling Bearing Performance Degradation Based on Attention-LSTM

Yaping Wang, Chaonan Yang, Di Xu, Jianghua Ge , and Wei Cui

Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China

Correspondence should be addressed to Jianghua Ge; gejianghua@hrbust.edu.cn

Received 4 December 2020; Revised 13 April 2021; Accepted 17 April 2021; Published 20 May 2021

Academic Editor: Yongfang Zhang

Copyright © 2021 Yaping Wang 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.

It is significant for the evaluation and prediction of the performance degradation of rolling bearings. However, the degradation stage division of the rolling bearing performance is not obvious in traditional methods, and the prediction accuracy is low. Therefore, an Attention-LSTM method is proposed to improve the evaluation and prediction of the performance degradation of rolling bearings. First, to reduce the uncertainty of the manual intervention, performance degradation characteristic indexes of rolling bearings are evaluated and screened by the correlation, the monotonicity, and the robustness. Second, the original characteristic indicator curve is divided into the Health Indicator (HI) curve and the residual curve by means of fixed-window averaging to quantitatively and intuitively reflect the deterioration degree of the rolling bearing performance. Finally, the Attention mechanism is combined with the LSTM model, and a scoring function is established to enhance the prediction accuracy. The scoring function is used to adjust the intermediate output state weight of the LSTM model and improve the prediction accuracy. The appropriate network structure and the parameter configuration are determined, and the prediction model of rolling bearing degradation performance is established. Compared with other models, the method proposed by this paper makes full use of the historical data and is more sensitive to the key information in the long time series, and the e_{RMSE} index and the e_{MAE} index of the two sets of experimental data are minimum, and the prediction accuracy of rolling bearing degradation performance is higher. The model has the strong robustness and the generalization ability, which has the important engineering practical value for the prediction of the equipment health state.

1. Introduction

The reliability and the stability of the rotating machinery are increasingly required because of the increasing complexity and the precision of its equipment system. Additionally, the health management technology for the mechanical equipment has attracted immense attentions [1, 2]. As one of the core components of the rotating machinery, manufacturing errors or the structural deformation of the rolling bearing will lead to unreasonable changes in the contact force and slight faults during the bearing operation process. This will greatly affect of the bearing performance and thus affects the efficiency of rotating machinery [3]. Therefore, the rolling bearing predictive maintenance can effectively reduce risks of the rotating machinery equipment and prevent significant losses.

The assessment of equipment performance degradation is one of the key components of predictive maintenances, which can be used to track changes in mechanical equipment from the installation to the current performance degradation state. However, the deterioration of the mechanical equipment performance cannot be obtained directly. Therefore, the degradation state of equipment health level is characterized by constructing a Health Indicator (HI) curve of the equipment in the actual performance degradation assessment. Lin et al. [4] constructed an HI for the decay state of gear cracks and used the percentage of residual signals as the health factor. Widodo A [5] extracted multiple features from the equipment monitoring signal and then used principal component analysis (PCA) to downscale the feature set and construct an HI by calculating the deviation between the degraded state and the feature vector of the equipment

health state. Jin et al. [6] calculated the energy value of the wavelet decomposition coefficient of the monitoring signal and then fused it to construct an HI based on the Mahalanobis distance between energy values. Yuna Pan [7] used the distance between the feature vector in the degraded state and the support vector in the healthy state as the HI after computing the wavelet packet decomposition of the device monitoring signal. Yu et al. [8] constructed multiple HIs in order to monitor the performance degradation of rolling bearings. Yu [9] used an HI to calculate the similarity between probability density functions described by two different hidden Markov models, which were used to monitor the performance degradation trend of rolling bearings. In addition, the use of machine learning algorithms to fuse multiple time-domain or frequency-domain features has become prevalent; for example, neural networks have been used to fuse time-/frequency-domain features to construct the HI [10]. In essence, the degradation of mechanical equipment performance is a slow and constantly changing process, so the characteristic indicator of the signal cannot be directly taken as the HI value. However, HI curves are constructed by using multiple feature fusion methods or artificial intelligence-extracted feature coefficients for sequential data having the poor generalization ability and the lacking universality. Therefore, in the approach proposed in this paper, the characteristic curve of the signal is further decomposed into an HI curve and the residual curve to better reflect the performance degradation stage of the equipment.

The performance decay prediction technology is the core of predictive maintenance, which can detect and isolate early faults to determine the current state of equipment and predict the trend of performance changes. At present, prediction methods for the mechanical equipment performance degradation can be divided into two categories: the model-driven method and the data-driven method [11]. In the model-driven method, the performance degradation of mechanical equipment is assessed primarily from changes in the physical information of the equipment, and its approach is largely based on parameters such as the mechanical stress and the strain [12, 13]. Wang et al. [14] proposed a mechanical operating state prediction method based on a probability model to predict the operating state of wind turbines. Cubillo et al. [15] proposed a physical model to predict the performance degradation of rotating machinery. Although model-based methods can analyze the nature and accurately predict the degradation trend of rotating mechanical properties, establishing specific physical models is challenging in many engineering applications, which increases the difficulty of prediction. The data-driven method can monitor the mechanical equipment in real time and easily update and adjust the model parameters and the performance degradation trend of the equipment [16, 17]. Lingli Cui et al. [18] established a new no-track Kalman filter method to predict the performance status of rolling bearings at various stages in response to problems such as the inability to determine the running status of rolling bearings. Shi Xiaoxue et al. [19] proposed an adaptive genetic particle filter method to

predict the declining trend of rolling bearing performance in view of the low prediction accuracy for rolling bearings. Liu et al. [20] used a model-based particle filter method to monitor the performance degradation of rolling bearings for a series of problems caused by the operational state of the bearings. To address the low prediction accuracy of rolling bearing performance degradation, Fafa Chen et al. [21] proposed an evaluation and prediction method based on wavelet packet information entropy and multicore correlation vector machine to monitor the rolling bearing performance status in real time. However, these methods have problems such as low prediction accuracy and poor generalization ability when high-dimensional data are obtained.

With the improvement of the computer processing speed, artificial intelligence technology is undergoing rapid development. With the aim of utilizing the data explosion of time series prediction [22], the artificial intelligence method can build a model adaptively for an unknown mechanical equipment state and improve the prediction accuracy of the model through repeated training. Among such methods, recurrent neural networks (RNNs) [23, 24] and long- and short-term memory (LSTM) [25, 26] are often used to predict the deterioration of mechanical equipment performance. Feng Li et al. [27] proposed a quantum recursive encoder-decoder neural network to predict the performance degradation trend of mechanical equipment for cases in which the traditional neural network produces low-accuracy predictions. Qiaoping Tian et al. [28] proposed an approach based on the LSTM prediction model to monitor the degradation of bearing performance and then predicted the remaining service life of the bearing, as well as the effects on the operation and maintenance costs of mechanical equipment. Mengfu He et al. [29] established an LSTM model to predict the performance degradation of the bearing and assess the effect of the bearing's operating state on the normal operation of the mechanical system. Jianjing Zhang et al. [30] proposed a prediction method based on the LSTM model, which was able to effectively monitor the performance degradation state of the mechanical system. Sheng Xiang et al. [31] proposed a novel long- and short-term memory neural network with weight amplification for accurate prediction of the remaining gear life in order to ensure the healthy operating conditions of gears. Xiang et al. [32] proposed LSTM networks based on attention ordered neurons for gear remaining life prediction, and the experiments show the superiority of the new gear RUL prediction methodology based on LSTM-AON compared to the current prediction methods. Although these methods are widely used in the field of prediction, there are still some deficiencies. These models cannot memorize longer time series because of the large amount of data and their limited memory ability. Furthermore, when the same weight is assigned to all information regardless of importance, the model forgets important information, which reduces the prediction accuracy of mechanical equipment performance degradation. The biggest advantage of Attention is that it gives more weight to important information, which can effectively improve prediction accuracy. Y Qin [33]

proposed, to accurately predict the RUL of the rolling bearing, a new kind of gated recurrent unit neural network with dual attention gates, namely, gated dual attention unit (GDAU), and the experimental results show that the proposed GDAU can effectively predict the RULs of rolling bearings, and it has higher prediction accuracy and convergence speed than the conventional prediction methods. Therefore, this paper reports a method that combines the Attention mechanism [34] with the LSTM model to predict the performance degradation of mechanical equipment. In a comparison with other methods, the proposed model has the minimum mean squared error in the prediction of the performance degradation of rolling bearings, and the accuracy of the prediction for mechanical equipment is effectively improved. Y Qin [35] proposed a Macroscopic-Microscopic Attention in LSTM Networks Based on Fusion Features for Gear Remaining Life Prediction, and the experimental results showed that the method can predict the remaining life of gears and bearings with higher prediction accuracy than the traditional prediction methods.

In summary, an Attention-LSTM-based method for the evaluation and the prediction of the rolling bearing performance decay is proposed to resolve the lack of evaluation indexes and the low prediction accuracy of the equipment performance degradation. First, feature indicators that have the better comprehensive performance are selected from a series of candidates. Second, the selected characteristic indicators are used to divide the original curve into the HI curve and the residual curve for the stage division and the performance evaluation. Third, the Attention-LSTM prediction model is established according to the data characteristics. Finally, the influence of different structural parameters on the model is compared and analyzed, and the final model structure and parameters are determined to evaluate and predict the deterioration of rolling bearing performance. This method provides some insights and strategies for the fault prediction and the health management of the mechanical equipment.

The rest of this article is arranged as follows. In Section 2, an evaluation method of the rolling bearing health stage is proposed to analyze the performance evaluation index of the bearing. The Attention-LSTM model is proposed to predict the degradation of rolling bearing performance in Section 3. An experimental study of the Attention-LSTM-based evaluation and prediction method for rolling bearing performance decay is introduced in Section 4. Finally, the summary of this paper and the future research direction are presented in Section 5.

2. Evaluation of Performance Degradation of Rolling Bearings

In order to evaluate the running state of the rolling bearing and the trend of performance degradation, the vibration signal of the bearing is extracted and reprocessed to obtain an effective candidate feature indicator of performance degradation. The monotonicity, the robustness, and the correlation of the signal are used to screen out the best feature indicators. From these parameters, signal feature

indicators that have better comprehensive performance are selected and will determine the accuracy of the subsequent evaluation and prediction of performance degradation. Then, the preferred feature indicators are separated into an HI curve and the residual curve. The HI curve is further normalized to determine the overall trend of bearing performance changes; the residual curve reflects the stability of the system during the operation phase of the bearing, and the result of the performance degradation evaluation of the rolling bearing is obtained.

2.1. Selection of HIs. Multiple potential indicators, such as root-mean-square, peak-to-peak value, and slope, are generated through time-domain analysis of the signal. A series of characteristic indicators and calculation formulas are listed in Table 1 and represent potential performance degradation characteristic indicators.

This section primarily analyzes several characteristic indicators that reflect signal characteristics. Certain characteristic indicators are difficult to solve, and the calculation process takes a long time, which is not suitable for the online data analysis. Therefore, it is essential to select appropriate characteristic indicators for the follow-up analysis. Signal feature indicators have the better comprehensive performance to provide significant data support for the subsequent performance degradation evaluation and the prediction. A single evaluation index is one-sided and cannot reflect the comprehensive performance of signal characteristics. However, the candidate feature indicators in Table 1 are comprehensively evaluated using a linear weighting method based on multiple metrics. Signal characteristic indicators that have the excellent monotonicity, the robustness, and the correlation have the better comprehensive performance.

The correlation reflects the degree of the correlation between the sequence of condition monitoring indicators and the decline in equipment performance. The value range of correlation is $[0, 1]$. The closer the correlation value to 1, the higher the correlation between the index and the life time, and it reflects that the index can well describe the performance degradation process of the equipment. The formula is as follows.

The correlation reflects the degree of correlation between the sequence of condition monitoring indicators and the deterioration of equipment performance and takes a value in the range $[0, 1]$.

$$\text{Corr}(X, T) = \frac{\left| K \sum X_T(t_k)t_k - \sum_k X_T(t_k) \sum_k t_k \right|}{\sqrt{\left[K \sum_k X_T(t_k)^2 - (\sum_k X_T(t_k))^2 \right] \left[K \sum_k t_k^2 - \left(\sum_k t_k \right)^2 \right]}} \quad (1)$$

where k is the total number of sampling points, $X_T = (X_T(t_1), X_T(t_2), \dots, X_T(t_K))$ is the condition monitoring feature sequence, and $T = (t_1, t_2, \dots, t_K)$ is the corresponding monitoring time series.

Monotonicity reflects the consistency of equipment performance degradation. As the equipment degradation process is irreversible and unavoidable, the condition

TABLE 1: Alternative performance degradation indicators.

Feature	Equation
Mean value	$\bar{X} = (1/n) \sum_{i=1}^n x_i$
Peak-to-peak value	$x_{p-p} = \max(x_i) - \min(x_i)$
Kurtosis	$\beta = (1/n) \sum_{i=1}^n x_i^4$
Peak indicator	$C_f = (x_p/x_{\min})$
Margin indicator	$CL_f = (x_{\max}/x_r)$
Average amplitude	$ \bar{x} = (1/n) \sum_{i=1}^n x_i $
Root mean square	$x_{\text{rms}} = \sqrt{(1/n) \sum_{i=1}^n x_i^2}$
Peak value	$x_p = \max(x_i)$
Waveform index	$s_f = x_{\text{rms}}/ \bar{x} $
Pulse indicator	$I_f = (x_{\max}/ \bar{x})$
Kurtosis	$K_r = (\beta/x_{\text{rms}}^4)$
Square root amplitude	$x_r = [(1/n) \sum_{i=1}^n \sqrt{ x_i }]^2$

monitoring indicators reflecting its performance degradation should have a monotonic degradation trend. The value range of the monotonicity is $[0, 1]$. When an indicator tends to increase or decrease monotonically with time during the equipment performance degradation, its monotonicity takes the value 1. When an indicator is constant or varies randomly with time, its monotonicity takes the value 0. The formula is as follows:

$$\text{Mon}(X) = \frac{1}{K-1} \left| \sum_k \delta(X_T(t_{k+1}) - X_T(t_k)) - \sum_k \delta(X_T(t_k) - X_T(t_{k+1})) \right|, \quad (2)$$

where $X_T = (X_T(t_1), X_T(t_2), \dots, X_T(t_k))$ is the sequence of condition monitoring indicators; $\delta(\cdot)$ is the step function; k is the total number of sampling points.

Robustness is a description of the volatility of condition monitoring indicators and reflects the robustness of condition monitoring indicators to disturbances such as external points. The value range of robustness is $[0, 1]$. If the metric exhibits a smooth pattern of change over time, the greater the value of its robustness, and the less uncertainty there will be in the performance degradation prediction results. The formula is as follows:

$$\text{Rob}(X) = \frac{1}{K} \sum_k \exp\left(-\left|\frac{X_R(t_k)}{X(t_k)}\right|\right), \quad (3)$$

where k is the total number of sampling points, and $X_T(t_k)$ is the sequence of residuals corresponding to the sequence of condition monitoring indicators.

The condition monitoring indicators are optimised by considering the attributes of the monotonicity, the correlation, the predictability, and the robustness, and a sequence of key condition monitoring indicators is selected to reflect the degradation process of the equipment for performance prediction. In this study, a weighted linear combination is used to determine the quality of the indicators to be tested:

$$\begin{aligned} \max J &= \omega_1 \text{Corr}(X, T) + \omega_2 \text{Mon}(X) + \omega_3 \text{Rob}(X) \\ \text{s.t.} &\begin{cases} \omega_i > 0 \\ \sum_i \omega_i = 1, i = 1, 2, 3, \end{cases} \end{aligned} \quad (4)$$

where J is the linear weighting of the three evaluation indicators and is linearly and positively correlated with each indicator, and the value of J is limited to the range of $[0, 1]$. The higher the value of J , the better the overall performance of the evaluation indicator. The indicator with the best comprehensive performance is selected from many characteristic indexes through the above method. ω_i is the attribute weight of a single detection indicator. In this study, the following weighting formula is used to assign weights:

$$\omega_{ij} = \begin{cases} \frac{1}{2} + \frac{\sqrt{-2 \ln[(2(j-1))/n]}}{6}, & 1 < j \leq \frac{n+1}{2}, \\ \frac{1}{2} - \frac{\sqrt{-2 \ln[2 - ((2(j-1))/n)]}}{6}, & \frac{n+1}{2} < j \leq n, \end{cases} \quad (5)$$

where $\omega_i = 1$, n is the number of attributes, i is the i -th attribute, j is the queue level (the queuing level is an arrangement of each attribute according to its importance. Different attributes of equal significance can be at the same level), and further normalization processing can obtain the attribute weight $w_i = (w_1, w_2, w_3)$.

2.2. Feature Separation Method for Fixed Window Averaging Processing. The performance degradation of rolling bearings is a continuous and slow process, and the characteristic index cannot be directly used as the HI value. Therefore, the filtered performance degradation index curve can be decomposed into the HI curve and the residual curve of the rolling bearing. First, the fixed-window averaging method is used to process the selected characteristic value curve, and the characteristic value within a certain period of time is averaged to represent the current HI value, thereby obtaining the HI curve of the full life cycle of the rolling bearing. The performance of the rolling bearing at a given moment is determined by its comprehensive operation over a period of time, rather than by a single eigenvalue at a single moment. The characteristic index $X(t_k)$ of the rolling bearing is composed of the overall trend curve $X_T(t_k)$ and the residual curve $X_R(t_k)$.

$$X(t_k) = X_T(t_k) + X_R(t_k). \quad (6)$$

Fixed-window averaging is used to select the eigenvalues of r time points before and after time t , and the HI value at that time is determined by averaging. The formula of fixed-window averaging is as follows:

$$x_{Tt} = \frac{\sum_{i=1}^r (x_{t-i} + x_{t+i}) + x_t}{2r + 1}, \quad (7)$$

where x_{Tt} is the HI value of the rolling bearing at time t , x_t is the characteristic value of the rolling bearing at time t , and r is the radius of the sliding window.

Averaging the characteristic value curve can markedly improve the monotonicity of the HI curve, but it reduces its sensitivity, and as a result, it cannot reflect the abrupt transition point in the process of performance degradation. Therefore, the window ΔT plays a decisive role in determining the HI curve. When the radius r of the window ΔT increases, the mean value does not affect the monotonicity of the HI curve, but it leads to a gentler change in the curve. Thus, averaging improves the monotonicity of the HI curve and reduces its sensitivity to the original eigenvalue. It can be seen that the radius r of the window ΔT is positively correlated with the monotonicity level of the HI curve and negatively correlated with its sensitivity. The monotonicity level of the HI curve can be used as the threshold to determine the radius r of the window ΔT ; on the basis of satisfying monotonicity requirements, the HI curve has the greatest sensitivity to abrupt transition points in the process of performance decline. The process of setting r is as follows:

- Parameter Initialization.* According to the characteristics of the data $X = x_1, x_2, \dots, x_n$ to be processed, the monotonicity coefficient threshold a' of the sequence to be processed and the initial value $r = 1$ of the window radius are inputted.
- Average Processing.* Equation (6) is used to average the X rows of the sequence with the r window radius to improve its monotonicity.
- Determination of Whether the Threshold Is Exceeded.* Formula (2) is used to calculate the monotonicity level a of sequence X . If $a < a'$, when the monotonicity level of sequence X does not meet the requirement, the window radius increases by 1, and then steps a, b, and c continued; if $a \geq a'$, the monotonicity level of sequence X meets the requirement, and the process proceeds to step d.
- Output of the Result.* The window radius r is determined. For the established radius r , the processing sequence X is averaged, which is defined as the sequence $X_T(t_k)$.

Then, the residual curve $X_R(t_k)$ is

$$X_R(t_k) = X(t_k) - X_T(t_k). \quad (8)$$

The residual curve $X_R(t_k)$ can be used to assess whether the bearing performance in this stage is in a dynamic equilibrium or nonequilibrium state.

The overall process of performance degradation assessment of rolling bearings is summarized in Figure 1.

2.3. Experimental Verification. In this study, experimental data of the full life cycle [36] of rolling bearings were used for verification; the data were obtained from the University of

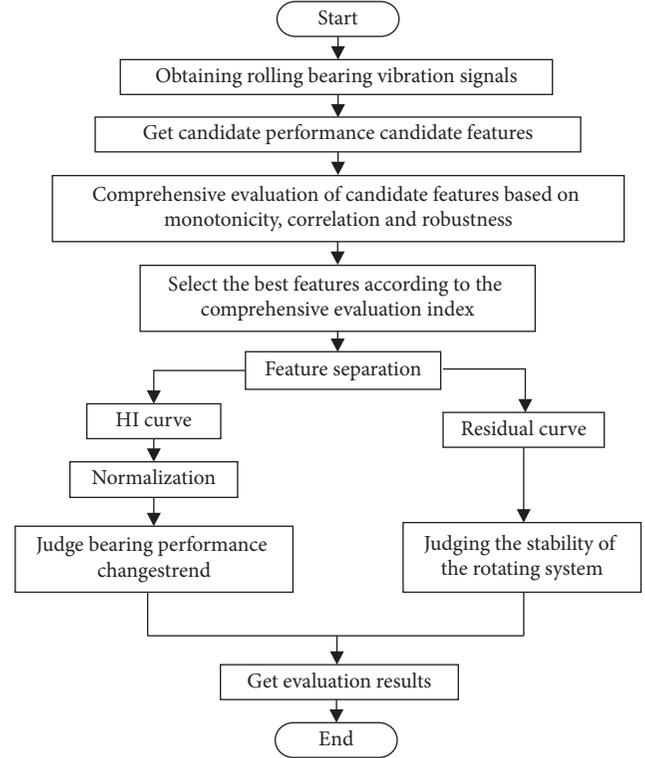


FIGURE 1: Performance degradation assessment flowchart.

Cincinnati. The schematic diagram of the test bed is shown in Figure 2. The test bed is equipped with four bearings, all of which are Rexford ZA-2115 bearings. Each track has 16 rolling bodies, the roller diameter is 3.31 mm, the pitch circle diameter is 28.15 mm, the contact angle is 15.17° , and the rotating speed is 2000 r/min. The experiment is an accelerated degradation experiment of rolling bearing, with 26.66 kN of radial load applied. Eight acceleration sensors are used for data acquisition, and the acceleration sensor model is the PCB353B33 piezoelectric sensor. The experiment took 33 days to complete, and data sampling was conducted for 1 s every 5 minutes. A total of 2,156 groups of data were collected in this experiment, with each group having a length of 20,480 points. At a later stage of the full life cycle experiment, a rolling body defect appeared in the fourth bearing. The collected data set was processed according to the details in Figure 1. The experimental data verification results are shown in Figures 3 and 4.

Three comprehensive evaluation indicators—monotonicity, correlation, and robustness—were used to quantitatively evaluate candidate indexes in the above table. Because the degradation of equipment is unavoidable in the process of its operation, the monitoring indicator of the equipment performance degradation process has a monotonous change trend, and it is ideal to have an extremely high correlation. Therefore, from formula (5), monotonicity has the first position in the queue, the correlation is second, and the robustness is last, as shown in Table 2:

Thus, the attribute evaluation results of different performance degradation characteristics listed in Table 1 are shown in Table 3.

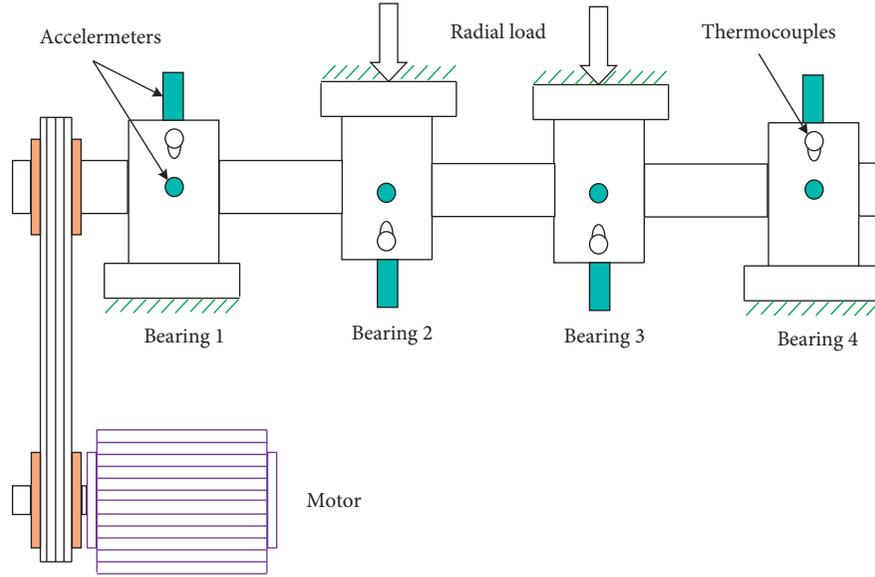


FIGURE 2: Schematic diagram of the test bench.

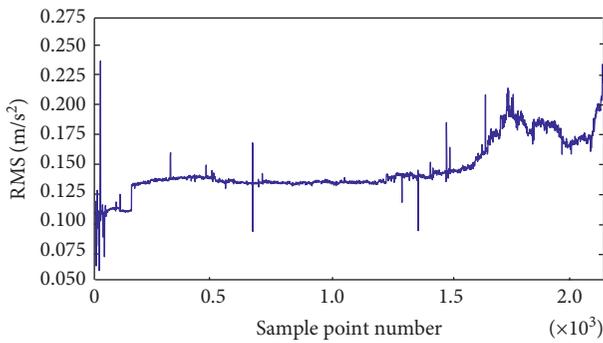


FIGURE 3: HI curve of the full bearing life cycle.

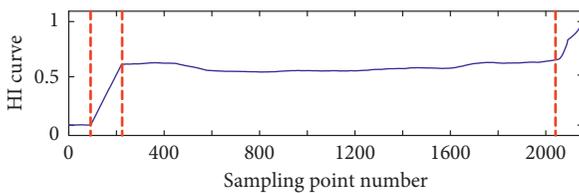


FIGURE 4: Residual curve of the full bearing life cycle.

TABLE 2: Weight calculation of each attribute.

Attribute	Mon	Corr	Rob
Queuing level	1	2	3
ω_{ij}	1	0.747	0.350
ω_i	0.477	0.356	0.167

It can be seen from Table 3 that the maximum linear weighting is for the root-mean-square value (RMS), so the RMS was selected as the performance decay index of the bearing.

The vibration signal of the bearing was obtained through tests during its full life cycle. Then, the RMS of each sampling

TABLE 3: Linear weighted value of each indicator.

Features	Evaluation indicators		Linear weighted	
	Corr	Mon	Rob	J
Square root amplitude	0.0044	0.5016	0.9844	0.4052
Root mean square	0.9007	0.5072	0.9816	0.7265
Peak to peak value	0.2762	0.5114	0.8660	0.4867
Peak value	0.0167	0.4947	0.8522	0.3842
Kurtosis	0.0064	0.5039	0.8637	0.3869
Waveform index	0.0058	0.5118	0.9950	0.4123
Margin index	0.0018	0.5104	0.8214	0.3813
Peak indicator	0.0004	0.5086	0.8239	0.3803
Average amplitude	0.0723	0.5049	0.9841	0.4309
Mean value	0.0532	0.5066	0.8856	0.4084
Pulse indicator	0.3452	0.4823	0.8432	0.4938
Kurtosis index	0.0081	0.5178	0.9236	0.4041

point was extracted to obtain the RMS for the full bearing life cycle, as shown in Figure 5.

As shown in Figure 3, after many experiments and comparisons, finally set the radius r to 3 and a' to 0.5, and the RMS for the full bearing life cycle was averaged using fixed windows to obtain the HI curve for the life cycle.

As shown in Figure 4, the residual curve of the full bearing life cycle was obtained from the difference between the original data and the data after averaging.

As shown in Figures 4 and 5, the combination of this bearing's full life cycle HI curve and residual curves clearly allows the full life cycle performance decline of the rolling bearing to be divided into four stages. The bearings are assembled for normal usage and operation in the first stage with excellent system stability. In the second stage, bearing damage occurs and the HI value rises from 0 to around 0.5, and the system is less stable at this stage from the residual curve, while the bearing performance plateaus later in the stage. In the third stage, the system reaches dynamic

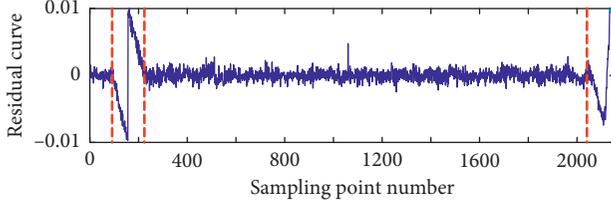


FIGURE 5: RMS value of the full bearing life cycle.

equilibrium, a new stage of steady state is formed, and the bearing runs smoothly until damage to the bearing occurs again. In the fourth stage, damage to the bearing occurs again, the system is unable to achieve a new dynamic equilibrium, and the bearing fails.

3. Prediction of Rolling Bearing Performance Decay Based on Attention-LSTM

In this paper, an Attention-LSTM performance decay prediction model is proposed based on the Attention and LSTM network models. The performance degradation prediction model based on Attention-LSTM was constructed from the following aspects.

3.1. LSTM Model. LSTM is a variant network of RNN [37]. The LSTM model has three gate structures: the forget gate, the input gate, and the output gate. These gate structures will selectively add or delete some information and allow information to pass through selectively. Here, i_t , f_t , o_t and c_t are used to represent the three gate structures and nerve cell states corresponding to t time. The details are provided as follows:

- (1) Forget gate f_t :

The forget gate determines the amount of information passed. The specific expression is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \quad (9)$$

Among them, σ is the sigmoid function; W_f and b_f are the weight and bias, respectively; h_{t-1} is the output at the previous moment; x_t is the input at the current moment.

- (2) Input gate i_t :

Input gate i_t is used to determine what kind of characteristic data information will be stored in the cell memory unit. The specific expression is as follows:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ c'_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \\ c_t &= f_t c_{t-1} + i_t \tilde{c}_t. \end{aligned} \quad (10)$$

Among them, W_i , W_c and b_i , b_c are the weight and bias of the input gate and the cell memory unit, respectively; \tanh is the activation function; c_t is the output of the cellular memory unit; c_{t-1} is the output of the cellular memory unit at the previous moment; c'_t is the output of \tanh .

- (3) Output gate o_t :

The output gate o_t controls the output of the cell state c_t at this moment, the output value is h_t , and the network determines the size of the output value through a Sigmoid function.

- (4) Determination of the final output h_t :

As shown in Figure 6, the final output of the LSTM is determined jointly by the output gate o_t and the unit state c_t . The cell state c_t at this moment is processed through the \tanh function ($\tanh(c_t) \in (0, 1)$) and multiplied by the output of the sigmoid function to finally compute the output value c_t at this moment. The specific expression is as follows:

$$\begin{aligned} o_t &= \sigma(W_o [h_{t-1}, x_t] + b_o), \\ h_t &= o_t^* \tanh(c_t). \end{aligned} \quad (11)$$

Among them, W_o and b_o are the weight and bias of the output gate, respectively.

3.2. Attention-LSTM Performance Degradation Prediction Model

3.2.1. Attention-LSTM Model Structure Design. The conventional coder-decoder model processes the input sequence in such a way that the encoder encodes the input sequence into a fixed-length hidden vector h to which the same weight is assigned. The hidden vector h is output through the decoder. The input sequence is given the same weight, which eventually leads to the degradation of the model performance. The Attention is a mechanism used to enhance the effectiveness of the encoder-decoder model. The Attention mechanism originated in the field of the human vision. The essence of the Attention mechanism is to imitate the thought process of the brain while paying attention to certain factors. When a certain scene frequently appears on multiple occasions, people will focus on that part of the information and pay more attention to it. The Attention mechanism assigns corresponding weights to the hidden vectors h at different moments of the input sequence, merges the hidden vectors into new hidden vectors according to their importance, and then inputs them into the decoder.

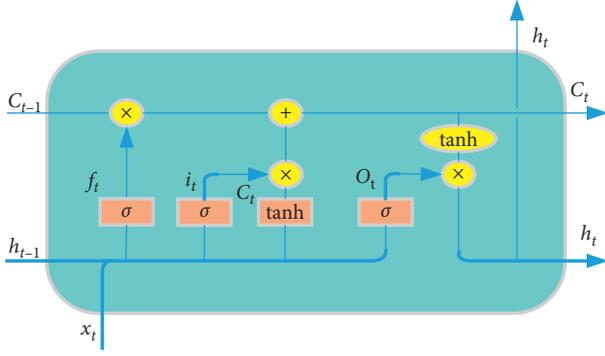


FIGURE 6: LSTM prediction model.

As shown in Figure 7, C_i is the weighted sum of the hidden layer h in the encoder; weight W_{ij} is related to the state h_j of the encoder at each time and the state h_{i-1} of the decoder at the previous time.

Although LSTM has certain advantages and a memory function, it cannot improve the expression of important information when processing long multidimensional time series. Furthermore, some important information will be ignored in the practical application process, which will lead to the performance degradation of the model and render it unable to achieve the ideal prediction accuracy. Therefore, the Attention mechanism is integrated into LSTM to improve the prediction accuracy of rolling bearing performance degradation prediction. The below describes how the Attention mechanism works in the LSTM model.

As shown in Figure 8, the Attention model is placed at the output of the LSTM in the Attention-LSTM framework, where $x_{t-1}, x_{t-2}, \dots, x_t$ is the time series to be predicted. The time series $x_{t-1}, x_{t-2}, \dots, x_t$ is input to obtain the intermediate state h_t through its forget gate, input gate, and output gate in the LSTM framework. Then, each intermediate state is used as the input of the Attention model to obtain the weight of each intermediate state W_{ij} through the corresponding dimensional transformation and the fully connected layer processing in the Attention model. Finally, the corresponding eigenvector can be obtained by multiplying each calculated weight by its corresponding intermediate state. The predicted results are obtained through the output of the softmax layer.

In the prediction of rolling bearing performance degradation, the intermediate state h_t of the LSTM layer determines the influence of the output state on the prediction result and then provides the corresponding weight to the Attention layer to achieve the best prediction accuracy. Therefore, a scoring function S_{ij} is defined (calculate the score between h_{i-1} and h_j), which is realized through the full connection layer. The importance of the output state is calculated by the scoring function S_{ij} . The higher the importance, the greater the value obtained. The specific expression is as follows:

$$S_{ij} = v \tanh(W \cdot h_j + U \cdot h_{i-1}' + b), \quad (12)$$

where v, w , and U are weighting factors, and b is a bias factor.

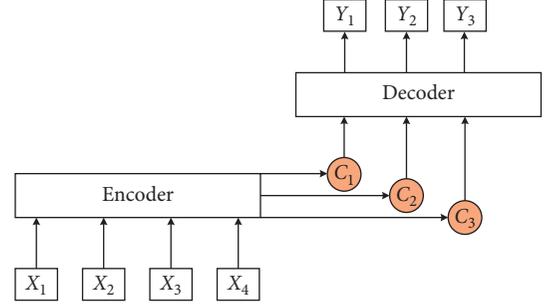


FIGURE 7: Encoder-decoder model combined with Attention mechanism.

The scoring function S_{ij} is used to calculate the weight W_{ij} of the LSTM output value h_t . The specific expression is as follows:

$$W_{ij} = \frac{\exp(S_{ij})}{\sum_{k=t-n}^t \exp(S_{ik})}. \quad (13)$$

The feature vector c is obtained by weighting, and then the prediction result at the next moment is calculated. The specific expression is as follows:

$$C = \sum_{j=t-n}^t \exp(W_{ij} h_j). \quad (14)$$

3.2.2. Activation Function Selection. In neural networks, the model has a certain degree of the nonlinear expression ability and solves nonlinear problems with complex data characteristics by adding activation functions. The function of optimizing the cyclic body gate structure by using different activation functions adjusted in the LSTM model. The output activation function plays an important role in the prediction model by ensuring the stability of the cell state and the model training.

In this section, the above five activation functions are compared. The functions were selected by using the aforementioned bearing fatigue life experiment data from the University of Cincinnati. Figure 9 shows the variation in the loss values of bearing data. The gray, red, blue, green, and purple curves in the figure represent the changes in loss values during the training of the sigmoid, tanh, ReLU, Leaky ReLU, and ELU activation functions, respectively. It is clear from Figure 9 that both ELU and sigmoid have faster convergence rates and are more stable as activation functions.

Through the experimental results in Figure 9 and Table 4, the ELU function was selected as the activation function of the LSTM network in this analysis.

3.2.3. Model Training. The goal of the loss function is to minimize the loss of the neural network by optimizing parameters (weights) of the network. The target (actual) value is matched with the predicted value through the neural network, and the loss can be calculated through the loss

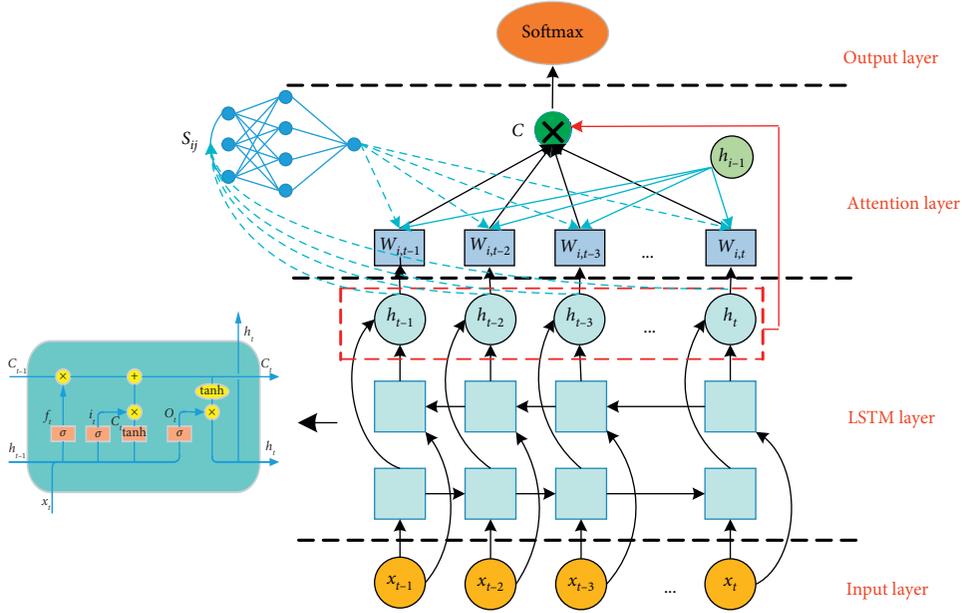


FIGURE 8: Attention-LSTM prediction model.

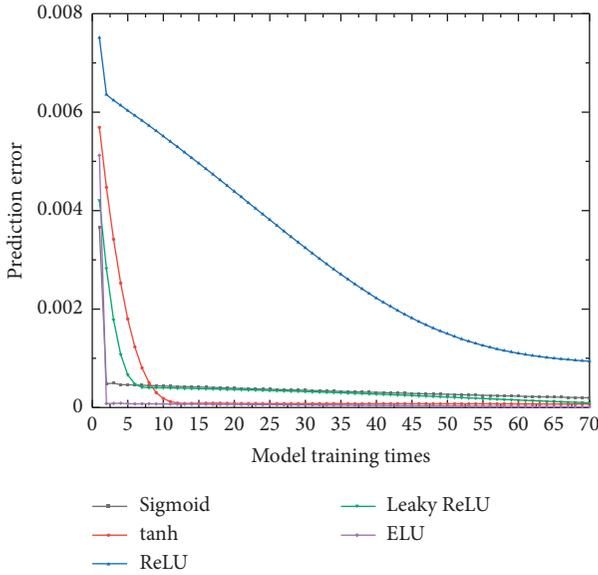


FIGURE 9: Loss value changes during the training process using the first set of bearing data.

function. The optimizer is then used to optimize the network weight to minimize the loss. In the process of the model training, the mean squared error (MSE) is used as the loss function of the LSTM model to reflect the difference between the real value and the predicted value through training. The specific expression is as follows:

$$e_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (15)$$

where y_i is the real value of the data i , \hat{y}_i is the value predicted by the LSTM model, and n is the data volume.

3.2.4. Model Evaluation. The model uses root mean square error (RMSE) and mean absolute error (MAE) to evaluate the prediction results. The specific expression is as follows:

Root mean square error (RMSE) formula:

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (16)$$

Mean absolute error (MAE) formula:

$$e_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (17)$$

Among them, y_i is the i -th predicted value; \hat{y}_i is the corresponding actual value. The prediction effect of the model is better when the root mean square error and the mean absolute error are smaller.

4. Experimental Verification and Result Analysis

4.1. Cincinnati Data. The experimental data in this section are the fatigue life of bearings and are from the University of Cincinnati. In order to carry out the model building and the data processing more accurately and quickly, the prediction model of the rolling bearing performance degradation was built on the basis of the currently popular deep learning framework Keras (which uses Tensorflow as the engine for calculation) in the pyCharm2019.3.2 software environment.

4.1.1. Parameter Settings. The parameters were set according to the previous preliminary research on the Attention-LSTM prediction model and are shown in Table 5.

TABLE 4: Experimental prediction error of different activation functions.

Activation function	Sigmoid	Tanh	ReLU	Leaky ReLU	ELU
Prediction error	0.062476	0.108454	0.080745	0.116402	0.056929

TABLE 5: Attention-LSTM prediction model parameters.

Parameter name	Parameter value
Number of attention-layer nodes	4
Number of input layer nodes	4
Number of output layer nodes	4
Proportion of training sequence	67%
Proportion of test sequence	33%
Number of middle layers	4
Number of middle layer nodes	6
Activation function	ELU
Loss function	MSE
Optimizer	Adam
Dropout layers	4
Learning rate	0.001

4.1.2. Data Verification. The structure and parameters of the above model were used to verify the prediction of bearing fatigue life from the experimental data from the University of Cincinnati. Figure 10 shows the prediction effect of the Attention-LSTM-based prediction model of the rolling bearing performance degradation with the bearing data set.

The black line represents characteristic values of the rolling bearing over its full life cycle. The red line represents the data of the model training set, whose data length is 67% of the full life cycle of the rolling bearing. The green line represents the model test data set, whose length is 33% of the full life cycle of the rolling bearing. From the prediction results, it can be concluded that the performance degradation prediction model of the rolling bearing based on the Attention-LSTM can accurately predict the full life cycle operation state of bearing. When the bearing failure occurs in the entire life cycle, the vibration signal oscillates, and the amplitude increases. The predicted curve is highly fitted with the actual curve trajectory, and there is no overfitting or underfitting. Compared with other prediction models, this model has higher prediction accuracy, faster learning of fault feature information, faster convergence speed, and stronger generalization ability.

Figure 11 shows the prediction effect of the rolling bearing performance degradation prediction based on multiple prediction models in the bearing data set.

According to the analysis in Figure 11, the rolling bearing performance degradation prediction model based on the multilayer perceptron, BP neural network, and RNN model has a poor prediction effect in the bearing data set, and the prediction effect in the later stage of the whole life cycle is quite different from the actual data. Therefore, these models are not suitable for a long-term data prediction. The prediction effect of LSTM is more stable than other rolling bearing performance degradation prediction models, and the model prediction ability is better. However, in the later stage of failure, the prediction curve of this model and the true curve still show insufficient fitting, and the prediction effect is poor compared with the Attention-LSTM model.

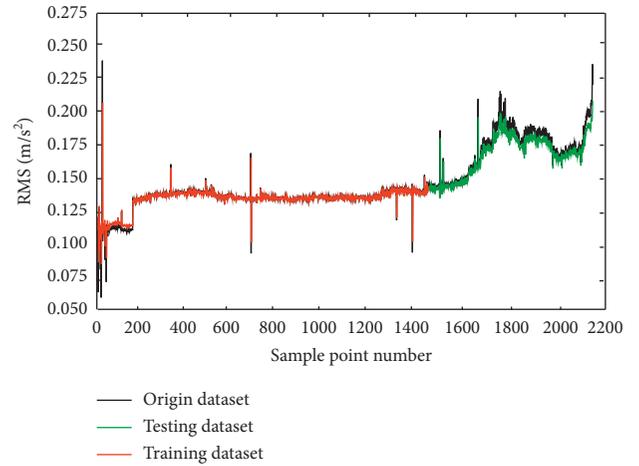


FIGURE 10: The prediction ability of the Attention-LSTM model.

As shown in Table 6, the prediction comparison effects of the five prediction models can be concluded that the Attention-LSTM-based prediction model, in which e_{RMSE} and e_{MAE} is the smallest, has the best prediction effect. The second is LSTM, which is better than the rest of the prediction models. The BP neural network model has the worst prediction effect, and both the e_{RMSE} and e_{MAE} of the training set and test set are the largest compared with the other four prediction models; especially the prediction curve fitting in the later stage of the full life cycle is insufficient and fails to meet the requirements of the rolling bearing performance recession prediction.

4.2. Laboratory Data Verification

4.2.1. Test Bench Construction. The Attention-LSTM model proposed in this paper for the evaluation and the prediction of the rolling bearing performance degradation was verified by a performance decay test bed. Cylindrical roller bearings were used in the experiment. The laboratory-scale rolling bearing test bed is shown in Figure 12.

The motor of this test bed needs to run continuously for a long time, so a DC motor is used and equipped with a motor controller with the overload protection. The protection circuit can prevent the motor from burning out if the bearing is locked. During the test, the motor speed was set at 1500 r/min. The model of data acquisition instrument is CoCo-80, which is a handheld device that integrates a vibration data acquisition instrument (VDC), dynamic signal analyzer (DSA), and long time data recorder.

In this test, two bearing vibration signals were collected: the bearing in the motor end was seriously damaged, and the bearing in the free end was slightly worn. Therefore, the full life cycle data set of the bearing in the motor end was used for the data validation. The data set includes all data for the

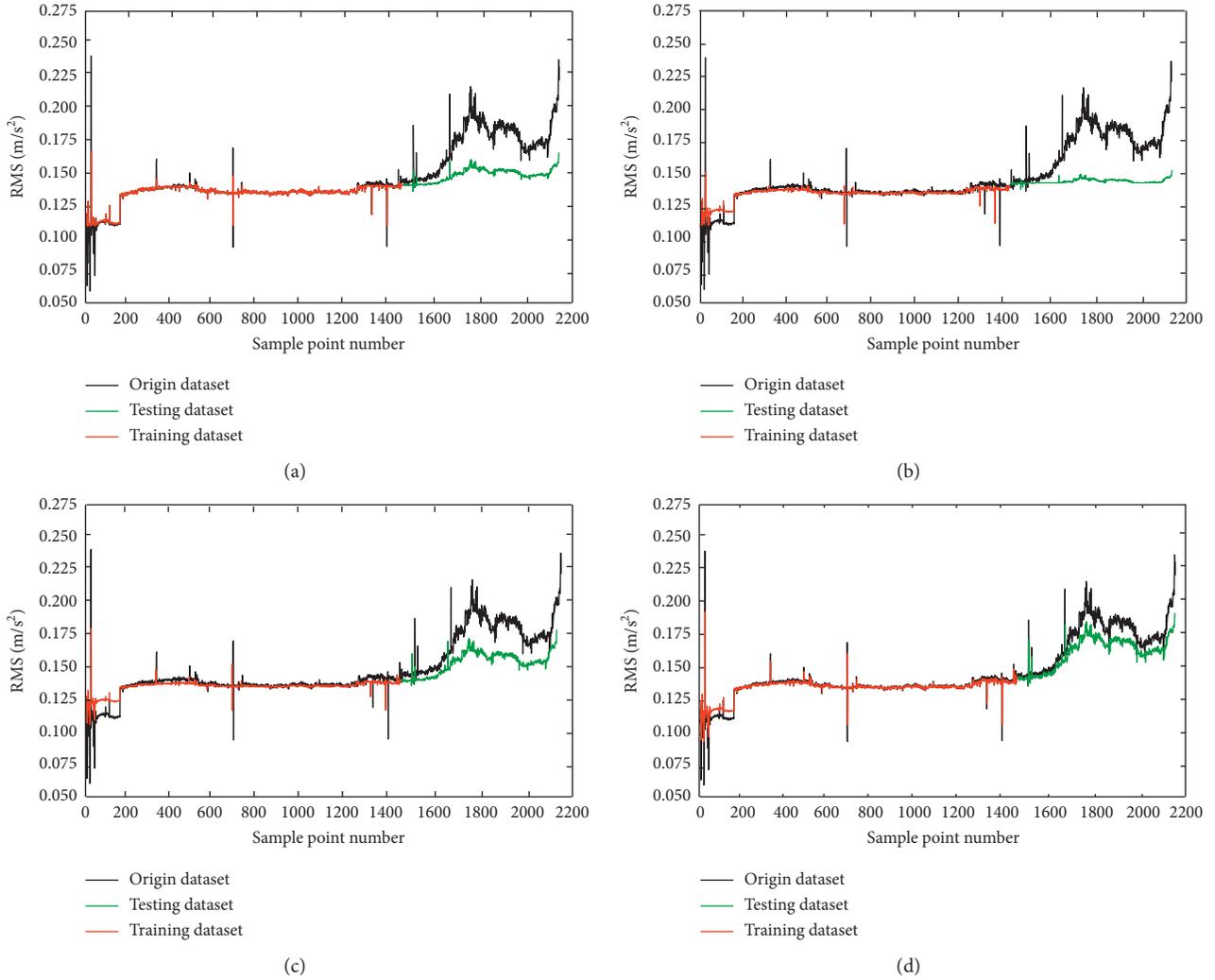


FIGURE 11: Multiple model prediction effect diagram. (a) MLP. (b) BP neural network. (c) RNN. (d) LSTM.

TABLE 6: Errors in different prediction models.

Predictive model	Training set RMSE	Test set RMSE	Training set MAE (%)	Test set MAE (%)
MLP	0.060	0.026	6.754	8.647
BP neural network	0.062	0.035	6.886	9.095
RNN	0.067	0.021	6.757	8.346
LSTM	0.060	0.012	5.319	7.485
Attention-LSTM	0.052	0.009	4.170	5.988

bearing under test, from normal operation to serious damage. The sampling frequency was set at 20 kHz, and the sampling interval was 5 minutes. One data file was generated for each collection, resulting in 22040 data files, with a total collection time of 76.53 days.

4.2.2. Performance Degradation Prediction Analysis. The RMS for the full life cycle of the rolling bearing obtained above was used as the data set to verify the proposed performance degradation prediction method based on the Attention-LSTM. The multilayer perceptron, BP neural network, RNN, LSTM, and Attention-LSTM were

analyzed in this paper. The whole data set was divided into training and test data sets. The training set included the first 67% of the entire data set and was used to train the prediction model. The remaining 33% of the data formed the test set, which was used to test the model. The prediction effect of the model is shown in Figures 13. The specific test results are listed in Table 7.

As shown in Figure 13, prediction curves of the MLP, BP neural network, and RNN models all show a certain degree of deviation. In particular, the prediction results of the BP neural network and MLP have larger errors, and it is obvious from the figure that the prediction curves deviate more from



FIGURE 12: Full life cycle performance degradation test bench.

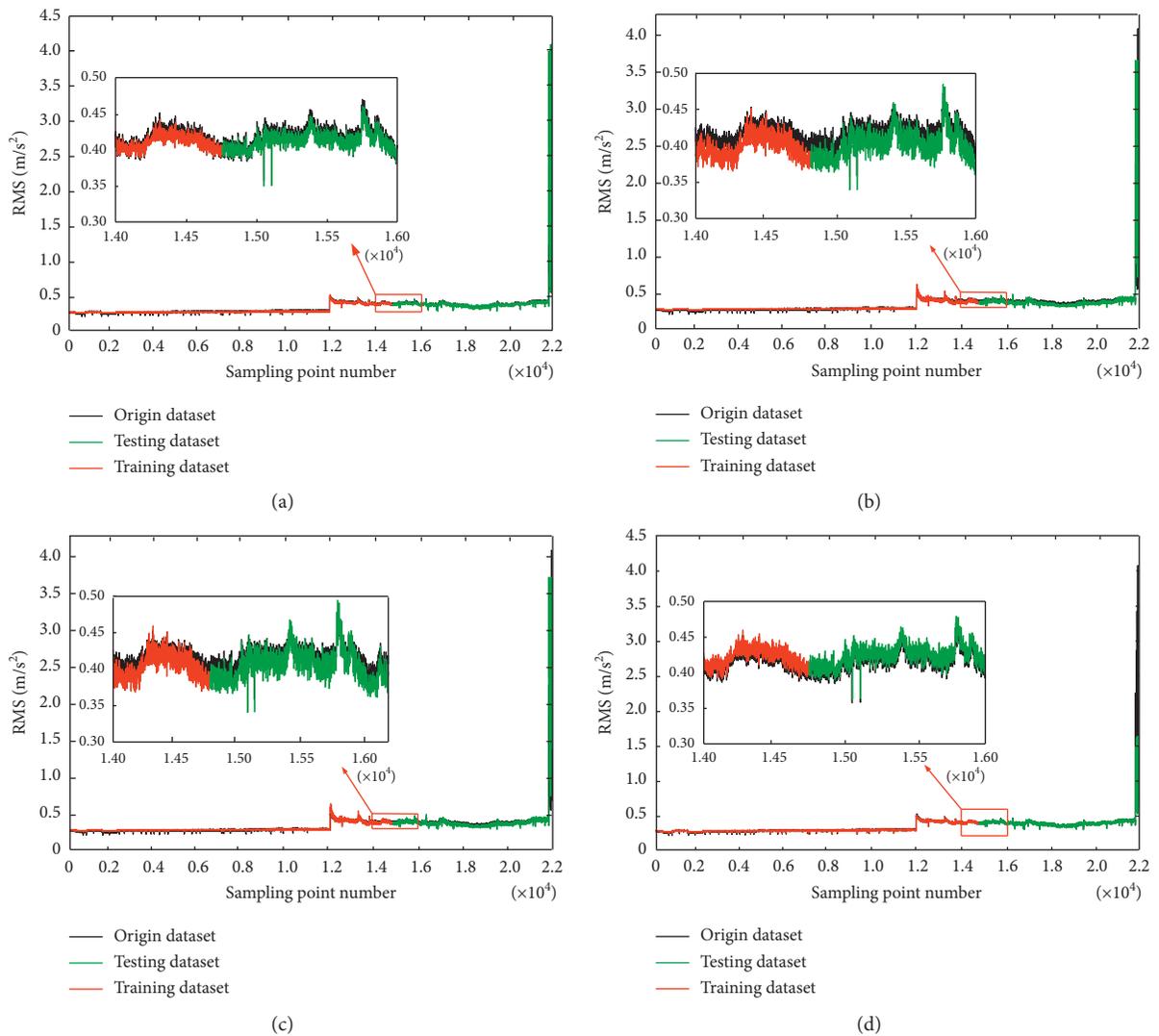


FIGURE 13: Continued.

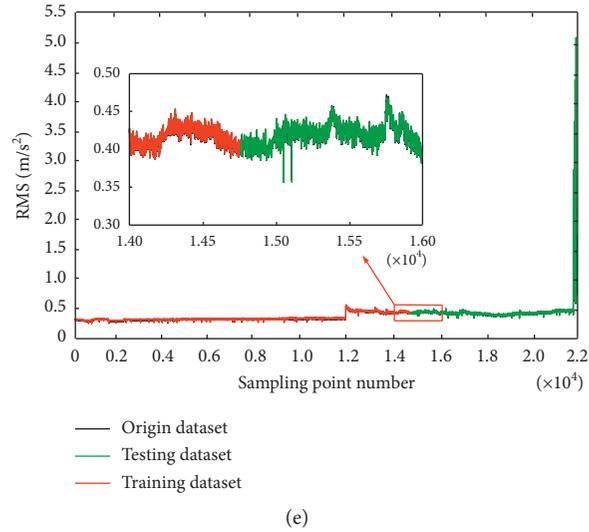


FIGURE 13: Multiple model prediction effect diagram. (a) MLP. (b) BP. (c) RNN. (d) LSTM. (e) Attention-LSTM.

TABLE 7: Errors in different prediction models.

Predictive model	Training set RMSE	Test set RMSE	Training set MAE/%	Test set MAE/%
MLP	0.017	0.234	8.792	8.555
BP neural network	0.018	0.246	9.916	7.903
RNN	0.012	0.125	8.841	8.227
LSTM	0.011	0.108	7.654	9.064
Attention-LSTM	0.008	0.086	6.915	7.029

the actual curves. The LSTM model fits well with actual curves in the early stage, but the fitting effect is less satisfactory in the later stage. In contrast, the Attention-LSTM model has the good prediction effect in each stage of the full life cycle, and the prediction curve greatly fits the actual curve and has the best prediction accuracy. Especially, at the later stage of the bearing failure, the prediction curve almost coincides with the actual curve, and the model performance is more stable.

From the analysis in Table 7, the Attention-LSTM prediction model proposed in this paper performed the best in both the training and test sets and also had a good generalization ability. The e_{RMSE} of the Attention-LSTM prediction results in both different data sets being the smallest. Compared to the MLP, BP neural network, RNN, and LSTM, the reduction of the e_{MAE} in the training set was 1.877%, 3.001%, 1.926%, and 0.739%, respectively; in the test set, it was 1.526%, 0.874%, 1.198%, and 2.035%. The Attention-LSTM prediction model improved the accuracy of the bearing performance degradation prediction after the experimental comparison and the validation.

5. Conclusions

In this paper, a new evaluation and prediction method is proposed to predict the performance degradation of rolling bearings. The proposed method more intuitively reflects the performance change of bearings at each declining stage and more accurately predicts the performance degradation

trend. Evaluation indexes in the paper are used to comprehensively evaluate multiple performance features and obtain feature indexes with the great comprehensive performance. The fixed window averaging method is used to divide the selected characteristic index curves into the HI curve and the residual curve, which can accurately divide the performance decline stage of the bearing. The Attention mechanism theory is applied to the LSTM network, which improves the sensitivity of the model to historical data and strengthens the feature extraction ability of the model. The attention-LSTM prediction model has been validated by experimental comparison and has a better fit to the real curve than other prediction models, with the smallest root mean square error and mean absolute error, which effectively improves the accuracy of rolling bearing performance degradation prediction.

In this study, there are some deviations in the fitting effect when the rolling bearing fails in a later stage of its life. In the future, the overall neural network model structure will be further improved to increase the fitting accuracy so that the predicted results can reflect the trend of the bearing performance degradation more comprehensively.

Data Availability

The experimental verification in this article includes the Open University data verification and experimental data verification. Among them, the Open University verification data was generated by the NSF I/UCR Intelligent

Maintenance System Center (IMS-<http://www.imscenter.net>) supported by Rexnord in Milwaukee, Wisconsin; the first group of bearing No. 4 in the experiment was selected. The collected data is verified, because the bearing from normal to damaged verifies the decline of the rolling bearing performance. The laboratory data comes from the data in the accelerated degradation test bed of rolling bearings in this research institute. The data collected through repeated experiments can predict the decline of rolling bearing performance.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Y.W. conceived the research, C.Y. and W.C. performed the experiments and wrote the paper. D.X. and Y.W. provided guidance for experiments and contributed to revising the paper. J.G. and Y.W. provided software support and data analysis. All authors have read and approved the final manuscript.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, grant number 51575143, and supported by Heilongjiang Provincial Natural Science Foundation of China grant number E2018046.

References

- [1] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, p. 2876, 2017.
- [2] S. Aouaouda and M. Chadli, "Robust fault tolerant controller design for Takagi-Sugeno systems under input saturation," *International Journal of Systems Science*, vol. 50, no. 6, pp. 1163–1178, 2019.
- [3] Z. Wang, J. Wang, W. Du et al., "Research on fault diagnosis of gearbox with improved variational mode decomposition," *Sensors*, vol. 18, no. 10, 2018.
- [4] D. Lin, M. Wiseman, D. Banjevic, and A. K. S. Jardine, "An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 993–1007, 2004.
- [5] A. Widodo and Bo-S. Yang, "Application of relevance vector machine and survival probability to machine degradation assessment," *Expert Systems with Application*, vol. 38, no. 3, pp. 2592–2599, 2010.
- [6] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. S. Chow, "Anomaly detection and fault prognosis for bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 9, pp. 2046–2054, 2016.
- [7] Y. Pan and J. Chen, "Application research of wavelet packet-support vector data description in bearing performance degradation evaluation," *Journal of Vibration and Shock*, vol. 4, pp. 171–218, 2009.
- [8] J. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2573–2588, 2011.
- [9] J. Yu, "Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring," *Mechanical Systems and Signal Processing*, vol. 83, pp. 149–162, 2017.
- [10] 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.
- [11] S. Zhao, Y. Zhang, S. Wang, B. Zhou, and C. Cheng, "A recurrent neural network approach for remaining useful life prediction utilizing a novel trend features construction method," *Measurement*, vol. 146, pp. 279–288, 2019.
- [12] 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.
- [13] J. Wertz, M.-H. H. Shen, O. Scott-Emuakpor, T. George, and C. Cross, "An energy-based torsional-shear fatigue lifing method," *Experimental Mechanics*, vol. 52, no. 7, pp. 705–715, 2012.
- [14] J. Wang, R. X. Gao, Z. Yuan, Z. Fan, and L. Zhang, "A joint particle filter and expectation maximization approach to machine condition prognosis," *Journal of Intelligent Manufacturing*, vol. 30, pp. 605–621, 2019.
- [15] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, "A review of physics-based models in prognostics: application to gears and bearings of rotating machinery," *Advances in Mechanical Engineering*, vol. 8, pp. 1–21, 2016.
- [16] A. Chadli, M. Chadli, S. X. Ding, and N. Braiek, "Design of robust fuzzy fault detection filter for polynomial fuzzy systems with new finite frequency specifications," *Automatica*, vol. 29, pp. 42–54, 2018.
- [17] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, 2019.
- [18] L. Cui, X. Wang, Y. Xu, H. Jiang, and J. Zhou, "A novel Switching Unscented Kalman Filter method for remaining useful life prediction of rolling bearing," *Measurement*, vol. 135, pp. 678–684, 2019.
- [19] X. Shi and Y. Wu, "Forecast of rolling bearing performance decline trend based on AGPF," *Computer Measurement & Control*, no. 10, pp. 228–231, 2017.
- [20] X. Liu, P. Song, C. Yang, C. Hao, and W. Peng, "Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1549–1558, 2018.
- [21] F. Chen, Y. Yang, B. Tang et al., "Performance degradation prediction of mechanical equipment based on optimized multi-Kernel relevant vector machine and fuzzy information granulation," *Measurement*, vol. 151, Article ID 107116, 2019.
- [22] G. Huang, H. Li, J. Ou, Y. Zhang, and M. Zhang, "A reliable prognosis approach for degradation evaluation of rolling bearing using MCLSTM," *Sensors*, vol. 20, no. 7, p. 1864, 2020.
- [23] Ruixin, H. Jiang, X. Li et al., "A reinforcement neural architecture search method for rolling bearing fault diagnosis," *Measurement*, vol. 154, 2020.
- [24] Y. Chi, S. Yang, and W. Jiao, "Multi-label classification method of rolling bearing fault based on LSTM-RNN," *Journal of Vibration, Measurement & Diagnosis*, vol. 40, no. 3, pp. 563–629, 2020.

- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [27] F. Li, Y. Cheng, Y. Chen, and B. Tang, "Quantum attention loop encoding and decoding neural network for state trend prediction," *China Mechanical Engineering*, vol. 31, no. 21, pp. 2573–2582, 2020.
- [28] Q. Tian and H. Wang, "An ensemble learning and RUL prediction method based on bearings degradation indicator construction," *Applied Sciences*, vol. 10, no. 1, 2020.
- [29] M. He, Y. Zhou, Li Yang et al., "Long short-term memory network with multi-resolution singular value decomposition for prediction of bearing performance degradation," *Measurement*, vol. 156, Article ID 107582, 2020.
- [30] J. Zhang, P. Wang, R. Yan et al., "Long short-term memory for machine remaining life prediction," *Journal of Manufacturing Systems*, vol. 48, pp. 78–86, 2018.
- [31] S. Xiang, Y. Qin, C. Zhu, Y. Wang, and H. Chen, "Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction," *Engineering Applications of Artificial Intelligence*, vol. 91, Article ID 103587, 2020.
- [32] S. Xiang, Y. Qin, C. Zhu, Y. Wang, and H. Chen, "LSTM networks based on attention ordered neurons for gear remaining life prediction," *ISA Transactions*, vol. 106, pp. 343–354, 2020.
- [33] Y. Qin, D. Chen, S. Xiang et al., "Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings," *IEEE Transactions on Industrial Informatics* no. 99, p. 1.
- [34] H. Zhang, Y. Liu, C. Wang et al., "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors (Basel, Switzerland)*, vol. 20, no. 6, 2020.
- [35] Y. Qin, S. Xiang, Y. Chai et al., "Macroscopic–microscopic attention in LSTM networks based on fusion features for gear remaining life prediction," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 12, pp. 10865–10875, 2019.
- [36] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.
- [37] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.