electronics

MDPI

Article

A Method for Predicting the Remaining Life of Rolling Bearings Based on Multi-Scale Feature Extraction and Attention Mechanism

Changhong Jiang [1], Xinyu Liu [1], Yizheng Liu [2], Mujun Xie [1,*], Chao Liang [1,*] and Qiming Wang [1]

[1] School of Electrical and Electronic Engineering, Changchun University of Technology, Changchun 130000, China
[2] Jilin Province Dengxi Technology Co., Ltd., Changchun 130022, China
* Correspondence: xiemujun@ccut.edu.cn (M.X.); liangchao@ccut.edu.cn (C.L.)

Abstract: In response to the problems of difficult identification of degradation stage start points and inadequate extraction of degradation features in the current rolling bearing remaining life prediction method, a rolling bearing remaining life prediction method based on multi-scale feature extraction and attention mechanism is proposed. Firstly, this paper takes the normalized bearing vibration signal as input and adopts a quadratic function as the RUL prediction label, avoiding identifying the degradation stage start point. Secondly, the spatial and temporal features of the bearing vibration signal are extracted using the dilated convolutional neural network and LSTM network, respectively, and the channel attention mechanism is used to assign weights to each degradation feature to effectively use multi-scale information. Finally, the mapping of bearing degradation features to remaining life labels is achieved through a fully connected layer for the RUL prediction of bearings. The proposed method is validated using the PHM 2012 Challenge bearing dataset, and the experimental results show that the predictive performance of the proposed method is superior to that of other RUL prediction methods.

Keywords: rolling bearing; residual life prediction; multi-scale feature extraction; attention mechanism

Citation: Jiang, C.; Liu, X.; Liu, Y.; Xie, M.; Liang, C.; Wang, Q. A Method for Predicting the Remaining Life of Rolling Bearings Based on Multi-Scale Feature Extraction and Attention Mechanism. *Electronics* **2022**, *11*, 3616. https://doi.org/10.3390/electronics11213616

Academic Editor: Davide Astolfi

Received: 10 October 2022
Accepted: 2 November 2022
Published: 5 November 2022

1. Introduction

As a key component of mechanical equipment, rolling bearings play a role in bearing load and transferring kinetic energy and are known as the "joints of industrial equipment". However, rolling bearings have been operating under high loads for a long time, which has led to a variety of failures [1]. Once rolling bearing failure occurs, it not only causes economic interest loss, but even safety accidents. Some statistics indicate that bearing failures in machinery and equipment account for 30% to 40% of all failures [2]. Therefore, accurate prediction of the remaining useful life (RUL) of rolling bearings is an inevitable requirement for reducing equipment maintenance costs and ensuring the reliable operation of the equipment.

At present, prediction methods for the RUL of rolling bearings can be divided into two main types [3]: RUL methods based on mechanistic modeling [4], and data-driven RUL methods [5]. The RUL method based on mechanistic modeling is based on the failure mechanism of the equipment [6]. However, in practical engineering applications, the performance degradation mechanism of bearings is more complex, and it is difficult to establish an accurate mechanistic model. The data-driven RUL prediction method can extract the degradation characteristics of the equipment from a large amount of monitoring data and build the corresponding RUL prediction model. Therefore, data-driven RUL methods are more suitable for complex mechanical systems. The data-driven RUL approach consists of two key steps [7]: firstly, the construction of health indicators that can represent

the trend of bearing degradation, and secondly, the establishment of an effective RUL prediction model.

Traditional lifespan prediction methods mainly use signal analysis methods to construct health indicators. For example, in [8], the peak and root mean square (RMS) values of wavelet coefficients are fed into a recurrent neural network (RNN) to predict the remaining life of the bearing. In [9], health indicators were constructed by extracting the time and frequency domain features of the bearing signals, and the extracted health indicators were input into a deep autoencoder (DAE), which effectively predicted the RUL of the bearings. Although these methods of constructing health metrics can infer correlations and causal relationships hidden in the data, this requires the manual extraction of bearing features and relies on empirical knowledge [10], which lacks adaptiveness. To avoid the above, we can use the method of deep learning (DL) to directly learn the mechanical degradation features from the original data.

In recent years, DL theory has been extensively applied in the fields of data exploitation, image processing, and target recognition [11–13]. Deep learning-based RUL prediction abandons the traditional RUL method of manually extracting features by building a deep architecture neural network to obtain multi-leveled degradation features in the original time series. Convolutional neural networks (CNNs) have a good ability to extract degradation features from equipment and are widely used in the field of health monitoring and management of mechanical equipment. In the literature [14], the degradation features of bearings were learned by CNN; then, these features were constructed into health indicators by non-linear mapping. The literature [15] formed a convolutional autoencoder structure by fusing CNN models and autoencoders to better extract the degradation features of electric valves. However, ordinary CNN struggles to extract the degradation information of the device in a complex environment. As the number of layers in the network increases, model degradation will occur during training. At the same time, the elements of the convolutional kernel of ordinary CNNs are closely aligned with each other, and the perceptual field is fixed. To acquire a wider perceptual field and extract more feature information, the convolutional kernel size must increase, thus, also increasing the model parameters.

To address the above issues, some scholars have proposed the dilated convolution operation [16,17]. Bearing vibration signals belong to time series data, where RNNs have been used to handle time series information with good results. In [18], the health metrics of the device are fed into the RNN, and the RUL prediction of the device is achieved. However, RNNs can lead to the problem of gradient disappearance when processing long-sequence information [19]. To overcome this problem, some scholars have introduced Long Short-Term Memory (LSTM) networks with gating units. LSTM can learn long-term dependent information and effectively handle long-sequence data. The literature [20] combines CNN and LSTM to predict the remaining lifetime of rolling bearings. The attention mechanism was first applied to machine translation and is now applied extensively in the handling of various time series [21]. By calculating the attention probabilities of different features, the attention mechanism assigns different weights to different features in the model, reinforces more important features, and suppresses relatively unimportant features, which helps to improve the prediction performance of the model. In [22], a recurrent neural network based on an attention mechanism is proposed to predict the remaining life of a bearing.

The above methods have produced good results when predicting RUL for bearings; however, they all perform only single-scale feature extraction, which will inevitably result in the omission of certain important information. Moreover, the above methods do not consider the differences in the contribution of various features to the RUL prediction task, which will introduce adverse effects to the prediction results. In this paper, we propose a rolling bearing remaining life prediction method based on multi-scale feature extraction and an attention mechanism to extract temporal and spatial features from the normalized bearing vibration signals. The method then employs an attention mechanism to achieve a reasonable allocation of attention resources to the model and to enhance the influence of key information on bearing RUL prediction. The mapping of bearing degradation features

to remaining life labels is realized through a fully connected layer to achieve the RUL prediction of bearings. The effectiveness of the proposed method in this paper is validated on the PHM2012 bearing dataset.

The rest of the paper is organized as follows: in Section 2, the network structure of the bearing RUL prediction method is constructed, and a flow chart of the bearing RUL prediction method is given. In Section 3, the experimental data are firstly pre-processed, followed by the construction of quadratic labels, and finally, the experimental results of the proposed method and the comparison tests are given. Section 4 concludes the whole paper.

2. Basic Theory

2.1. Convolutional Neural Networks

CNN, as an important branch of deep learning, is extensively used in fault diagnosis [23] and the lifetime prediction of mechanical equipment [24]. CNN comprises an input layer, convolutional layer, pooling layer, fully connected layer, and output layer. Figure 1 shows the basic structure of CNN. The functions of each layer are as follows.

(1) Input layer: utilized mainly for data entry.

(2) Convolutional layer: It has the advantages of local area connectivity and weight sharing. The convolution layer is composed of a group of convolution kernels, which are the main tools for feature extraction. The specific operations are shown below.

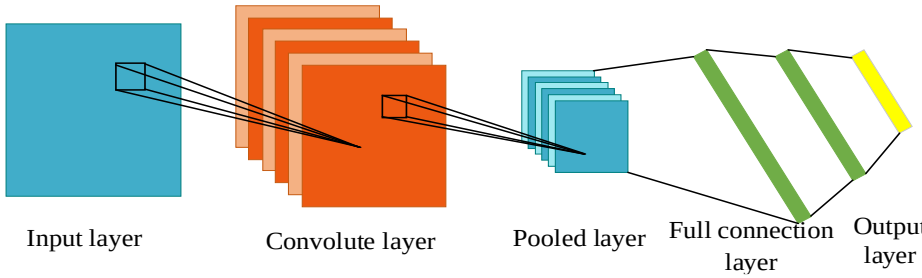


Figure 1. Basic structure of CNN.

$$y^{l(m,n)} = G_m^l * x^{l(r^n)} = \sum_{n'=0}^{W-1} G_m^{l(n')} x^{l(n+n')} \tag{1}$$

where W denotes the convolution kernel size, $G_m^{l(n')}$ denotes the n'th weight of the mth convolution kernel of the lth layer, and $x^{l(r^n)}$ denotes the nth local receptive field of layer l.

(3) Pooling layers: Generalize the output of convolutional layers at specific neighboring locations in the form of non-linear down-sampling to reduce the computational effort of the model, thereby increasing the computational speed of the network and making the feature representation translation invariant. This article adopts max pooling, the specific operations of which are shown below.

$$p^{l(m,n)} = \underset{(n-1)H+1 \le t \le nH}{max} \left\{ a^{l(m,t)} \right\} \tag{2}$$

where $p^{l(m,n)}$ represents the output value of the pooling layer, $a^{l(m,t)}$ represents the activation value, and H denotes the width size of the pooling domain.

(4) Fully connected layer: It maps the feature space extracted from the data after convolution and pooling to the sample space. The specific operations are shown below.

$$h^l = \sigma^l \left(\left(W^l \right)^T \times v^{l-1} + b^l \right) \tag{3}$$

where h^l denotes the output characteristics of the lth hidden layer, σ^l is the activation function of the lth layer, W^l denotes the connection weight between neurons in layer l and neurons in layer l–1, v^{l-1} is the output vector of layer l–1, and b^l is the offset.

(5) Output layer: mainly used to output the final prediction results.

2.2. Dilated Convolution

The elements of the ordinary convolutional kernel are arranged close to each other, and the obtained perceptual field is fixed. Therefore, if we want to obtain more perceptual fields and more feature information, we can only increase the size of the convolution kernel, which also causes an increase in the model parameters. To overcome these problems, some experts propose the operation of dilated convolution [25,26]. This convolution operation adds a certain void rate between each convolution kernel element but does not increase the parameters of the convolution kernel. The comparison of conventional convolution and dilated convolution is shown in Figure 2. As can be observed from Figure 2, the dilated convolution can obtain a larger perceptual field while preventing the parameters of the convolution kernel from increasing, so it has been used in many fields.

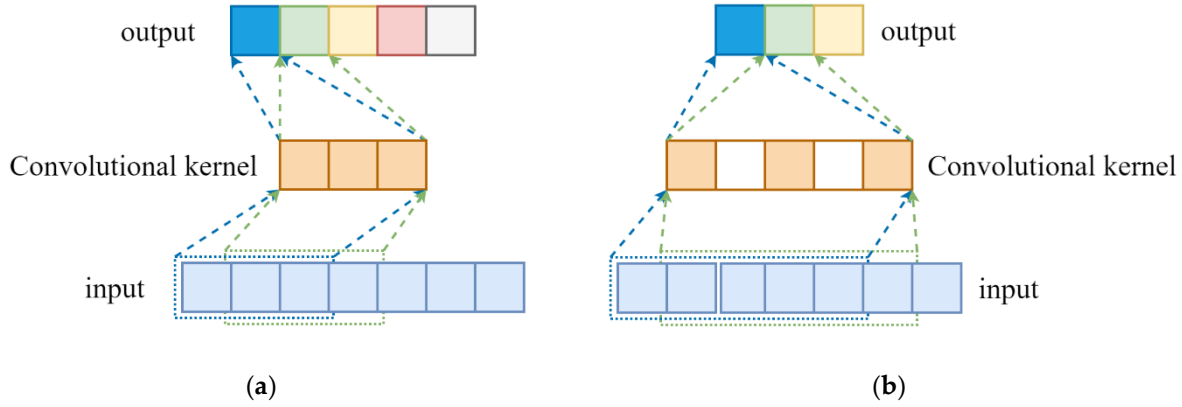


Figure 2. Comparison of ordinary convolution kernel and dilated convolution kernel. (**a**) Conventional 1 × 3 convolution kerel; (**b**) Dilated convolution 1 × 3, dilation rate is 2.

2.3. LSTM Networks

LSTM networks take into account the connection between outputs and inputs in a time series and have been applied extensively in the health management prediction of mechanical equipment [27,28]. Figure 3 shows the structure of an LSTM network. The LSTM network updates the network state mainly by forgetting gate f_t, input gate i_t, and output gate o_t. The cell state c_t and the output state h_t in the LSTM network are obtained by updating the cell state c_{t-1} and the output state h_{t-1} at the previous moment. The specific update process is as follows.

$$i_t = \sigma(\omega_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$o_t = \sigma(\omega_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$f_t = \sigma\left(\omega_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{6}$$

$$\widetilde{c}_t = \tanh(\omega_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$c_t = f_t * c_{t-1} + \widetilde{c}_t * i_t \tag{8}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{9}$$

where $\widetilde{c}_t$ denotes the candidate state, x_t denotes the input time series signal, h_t denotes the output updated by the network at time, and the Sigmoid and tanh functions are denoted by σ and $tanh$, respectively. ω_i, ω_o, ω_f, and ω_c denote the matrix weights of the input

gate, output gate, forgetting gate, and cell state, respectively; b_i, b_o, b_f, and b_c denote the offset of input gate, output gate, forgetting gate and unit state, respectively. "*" denotes the operation of multiplying the corresponding elements of two matrices of the same order, "·" denotes the ordinary product operation.

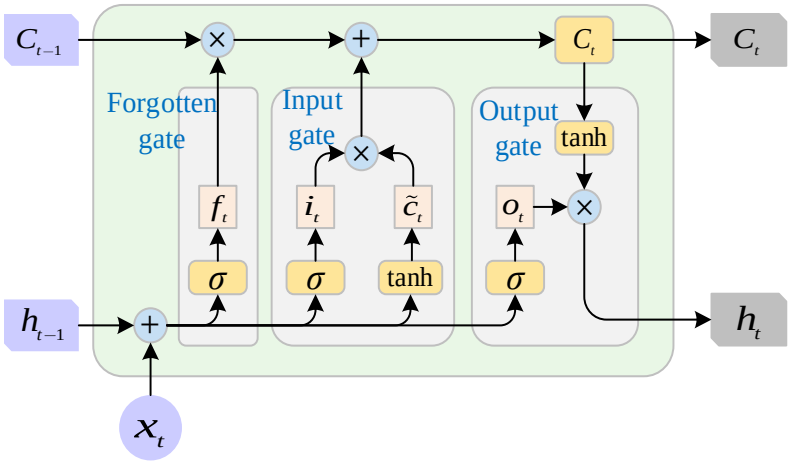


Figure 3. Basic structure of LSTM.

2.4. Attentional Mechanisms

Similar to the human visual mechanism, the attention mechanism can give more attention to key information that is beneficial to the task and less attention to unimportant information, thus, enabling the extraction of effective features [29]. The attention mechanism is not an exact model but an idea, and therefore, it can be combined with many network models. The current mainstream attention mechanisms can be divided into the following three types: channel attention, spatial attention, and self-attention. The channel attention mechanism aims to automatically obtain the importance of each feature channel by means of network learning, and finally assign different weight coefficients to each channel to reinforce the important features to suppress the unimportant ones [30]. The core idea of the channel attention mechanism is to help the network focus on the information related to the current input, assign different weights to different features, and multiply the input vector with the weights to achieve the importance assignment. The implementation process of the channel attention mechanism can be divided into two parts: the generation of attention weights and the assignment of weights. This is shown in the following equation.

$$A = h(X) \tag{10}$$

$$Z_1 = A \times Z \tag{11}$$

where X is the input vector, $h(\cdot)$ is the attention mechanism network, Z_1 is the output vector, A is the attention weight, and Z is the feature vector of the input vector X.

3. Rolling Bearing RUL Prediction Based on Multi-Scale Feature Extraction and Attention Mechanism

3.1. Network Model Construction

The network model of the rolling bearing remaining life prediction method based on multi-scale feature extraction and attention mechanism proposed in this paper is shown in Figure 4. Firstly, in order to extract more comprehensive bearing degradation indexes from the original data, this paper uses dilation convolution and long-short time neural network to extract the spatial and temporal features of bearings, where dilation convolution has a large sensory field and does not increase the optimization parameters of the network, while LSTM has a good ability to extract temporal features. Next, global average pooling (GAP) is used to structurally regularize the network to prevent overfitting and to give each channel

an actual category meaning. Then, the channel attention mechanism is used to implement adaptive weight assignment for bearing degradation features. Finally, a fully connected layer is used to implement the mapping of bearing degradation features to remaining life labels.

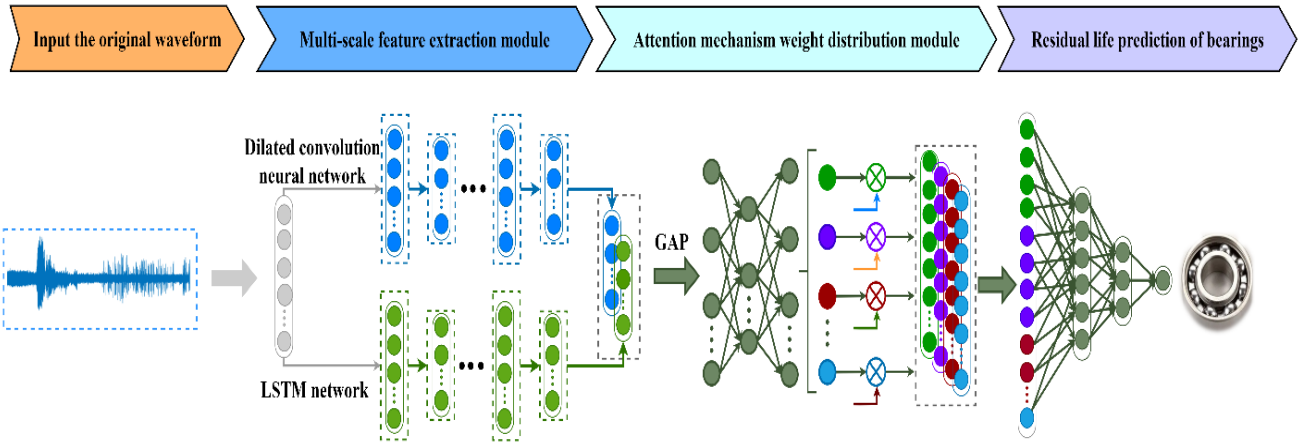


Figure 4. Network model of rolling bearing remaining life prediction method based on multi-scale feature extraction and attention mechanism.

3.2. Prediction Process of Bearing RUL Based on Multi-Scale Feature Extraction and Attention Mechanism

Figure 5 shows the flow chart of the bearing RUL prediction method for bearings based on multi-scale feature extraction and attention mechanism designed in this paper; the specific steps are as follows:

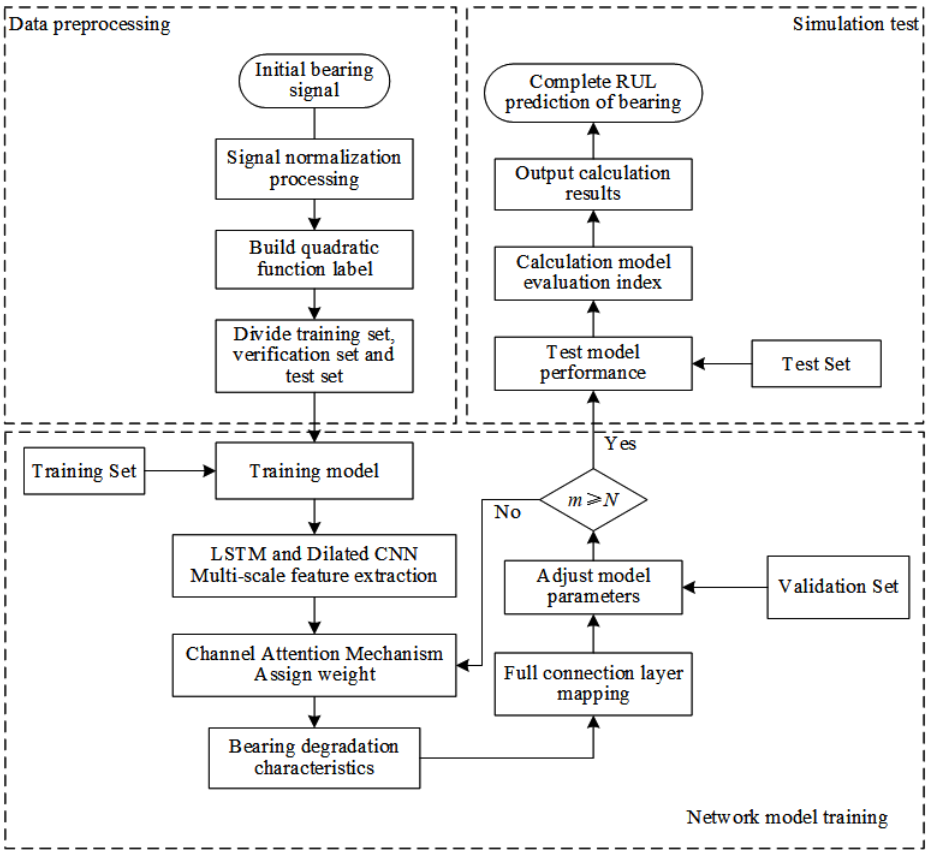


Figure 5. Flow chart of the RUL prediction method in this paper.

Step 1: Obtain the bearing vibration signal and normalize the original signal.

Step 2: Construct the quadratic degradation labels corresponding to the bearing vibration data and divide the normalized bearing vibration signal into the training set, test set, and validation set.

Step 3: Input the training set bearing vibration data into the Dilated CNN and LSTM network for adaptive extraction of spatial and temporal features; adjust the network parameters (including the learning rate, the number of iterations and the size of the convolution kernel).

Step 4: The weights are assigned to the bearing degradation features extracted by the multi-scale feature extraction module through the channel attention mechanism.

Step 5: A fully connected layer is used to implement the mapping of bearing degradation features to the remaining life labels for the RUL prediction of bearings.

Step 6: The validation set verifies the model training effect and fine-tunes the model parameters according to the validation results.

Step 7: The test set tests the performance of the trained model and calculates the model evaluation metrics, outputs the settlement results, and ends the process.

4. Test Validation

4.1. Test Data

The bearing vibration data for validating the proposed method in this paper are obtained from the PHM 2012 bearing dataset of the PRONOSTIA platform. The platform provides realistic bearing degradation data that can be used to validate various algorithms regarding bearing health assessment, remaining life prediction, and fault diagnosis. The PRONOSTIA experimental platform is shown in Figure 6. The stage allows the bearing to rotate at high speed and is fitted with two DYTRAN high-frequency accelerometers type 3035B to collect the bearing signals in both the horizontal and vertical directions. The vibration signal is sampled every 10 s with a sampling time of 0.1 s and a sampling frequency of 25.6 kHz so that 2560 data are recorded per sample. At the start of bearing rotation, all bearings are healthy and free of defects. The bearings underwent accelerated degradation during rotation, and once the amplitude of the bearing signal was monitored to exceed 20 g, the bearings were considered damaged, and the experiment was over.

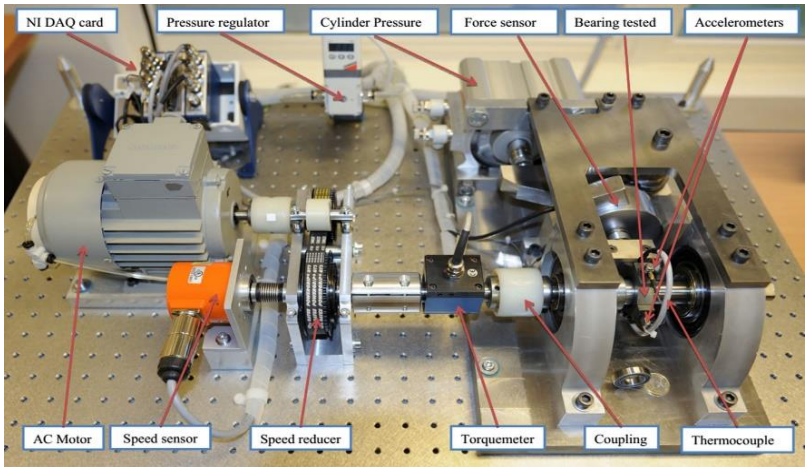


Figure 6. PRONOSTIA experimental platform.

The PHM 2012 bearing data set was collected under three different operating conditions. The specific information on the bearings under these three operating conditions is shown in Table 1. The article selects the bearing vibration data collected under operating condition 1 for experimental verification. Although the PHM 2012 bearing data set contains vibration data in both the horizontal and vertical directions, according to some experts, vibration signals in the horizontal direction provide more useful information than those

in the vertical direction [31]. Therefore, only monitoring data collected in the horizontal direction are used in the article.

Table 1. PHM2012 data presentation.

Working Condition	Condition 1	Condition 2	Condition 3
Number of bearing	1–1, 1–2, 1–3 1–4, 1–5, 1–6, 1–7	2–1, 2–2, 2–3 2–4, 2–5, 2–6, 2–7	3–1, 3–2, 3–3
Load (N)	4000	4200	5000
Speed (r/min)	1800	1650	1500

The time domain signals of bearing 1–1 and 1–3 are shown in Figure 7a,b. From Figure 7, it can be seen that the amplitude of the bearing vibration signal changes significantly with time, and the signal shows a tendency to disperse, which is beneficial to the extraction of health feature information with degradation trend.

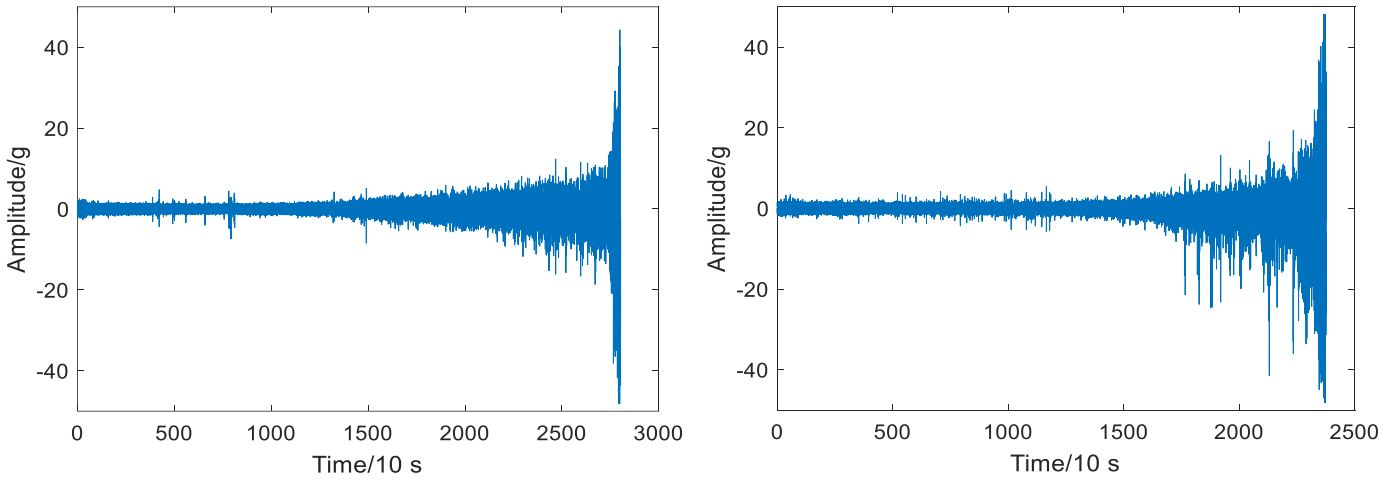


Figure 7. Time domain signal waveforms of bearing 1−1 and 1−3. (**a**) Time domain waveform of bearing 1−1; (**b**) Time domain waveform of bearing 1−3.

4.1.1. Data Preprocessing

To avoid the impact of inconsistent feature metric scales on prediction accuracy, the article uses the min-max normalization method to normalize the bearing signals. The min-max normalization is calculated as follows.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{12}$$

where x is the original bearing life signal, x_{min} is the minimum value in the original bearing life signal, x_{max} is the maximum value in the original bearing life signal, and x_{new} is the normalized bearing life signal.

4.1.2. Construction of Data Labels

After obtaining the raw vibration data of the bearings, they need to be divided into the training set, test set, and validation set. However, since the raw data do not have corresponding labels, degradation labels corresponding to the vibration data need to be constructed. At present, the commonly used degradation labels mainly include linear degradation labels and segmental degradation labels, as shown in Figure 8a,b below, respectively. The linear degradation label does not need to identify the degradation start point, and it is considered that the normal phase data also need to be predicted, which will

greatly improve the training time and is not conducive to network training; the segmental degradation label is trained only for the degradation phase, which reduces the prediction time consumption and also improves the prediction accuracy, but it needs to identify the degradation phase start point, which increases the labor cost.

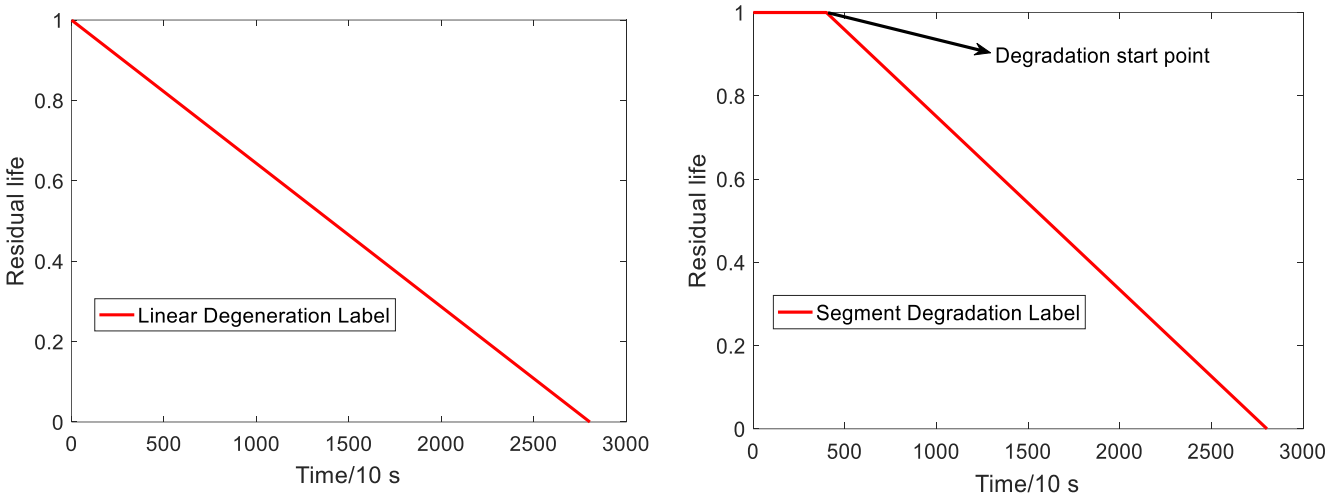


Figure 8. Two commonly used degradation labels. (**a**) Linear degeneration label; (**b**) Segment degradation label.

To address the above issues, the article uses the quadratic function indicator as the degradation label of the bearing, as shown in Figure 9. As can be observed from Figure 9, the label is more in line with the degradation trend of the bearing. In the early stage of bearing degradation, the degradation effect is not obvious, with a relatively gentle degradation trend, and in the late stage of bearing degradation, the bearing shows a rapid degradation trend. The label takes into account the entire degradation trend of the bearing and does not require the identification of the start of the degradation phase. The formula for the quadratic degradation label is as follows.

$$y_i = 1 - \frac{{t_i}^2}{T^2} \tag{13}$$

where y_i is the remaining life of the bearing at moment t_i, T is the time of complete bearing failure, and t_i is the sampling time.

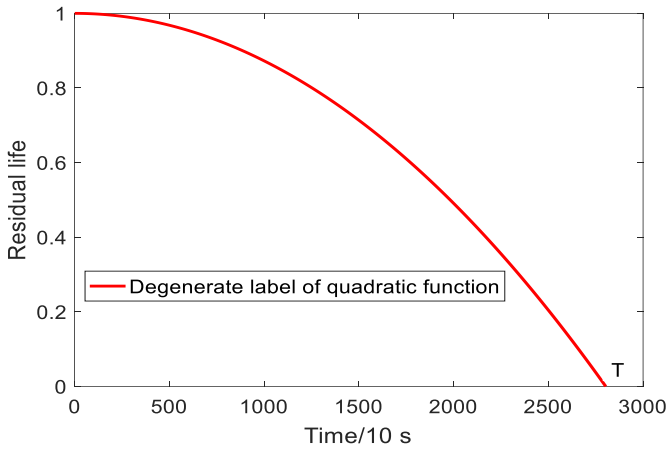


Figure 9. Quadratic function label.

4.2. Evaluation Indicators

To quantitatively evaluate the prediction effect, the article uses the root mean square error (RMSE) and the mean absolute error (MAE) between the predicted value of RUL and the true value of RUL as evaluation indicators. The smaller these two evaluation indicators, the smaller the difference between the predicted and true values, and the higher the prediction accuracy. The formulae for calculating RMSE and MAE are as follows.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2} \tag{14}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{y}_i - y_i\right| \tag{15}$$

where y_i denotes the true remaining life of the rolling bearing, and $\hat{y}_i$ denotes the predicted value of the remaining life of the rolling bearing. m is the number of samples.

4.3. Test Results

To verify the effectiveness of the proposed method, the paper takes the data under the PHM 2012 bearing dataset working condition 1 for the experiment, and uses bearing 1–1 as the training set, bearing 1–2 as the validation set, and other bearings under working condition 1 as the test set. The network structure of the RUL prediction method proposed in the article is shown in Table 2. For the hyperparameters of the network model, the method of multiple experiments is adopted to determine them. Specifically, the batch size is set as 64, the number of iterations is set as 50, the learning rate is 0.001, and the optimizer is selected as Adam. Since the proposed method is supervised learning, the mean square error function (MSE) is selected as the loss function of regression prediction in this paper, and the MSE function is calculated as follows.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{16}$$

where n is the number of samples, y_i denotes the real life of the bearing, and $\hat{y}_i$ denotes the predicted life of the bearing.

Table 2. Specific structure of the network model.

Layers	Operating	Parameters Size
1–1	Convolution	Filter = 3, kernel_size = 5, dilation = 3
	Dropout	0.2
	Max-Pool	Pool_size = 2
	Convolution	Filter = 6, kernel_size = 5, dilation = 3
	Dropout	0.2
	Max-Pool	Pool_size = 2
1–2	LSTM	Hidden_size = 1500, num_layers = 2, dropout = 0.5
2	Channel attention	/
	Flatten	5286
3	Fully connected 1	1000
	Fully connected 2	500
	Fully connected 3	100
	Fully connected 4	1

The hyperparameters of the network play an important role in the training of the whole network, so a reasonable selection of the hyperparameters of the network can improve the overall RUL prediction effect. First, the number of batches is set to 64 according to the device configuration. Second, to verify whether the network has reached the convergence state, the loss function curve is visualized in this paper, and the network model training loss is shown in Figure 10. Figure 10 shows that the value of the loss function dropped to below 0.005 after the network reached 10 iterations; thus, it can be concluded that the training has reached the convergence state, so the number of iterations set to 50 is reasonable.

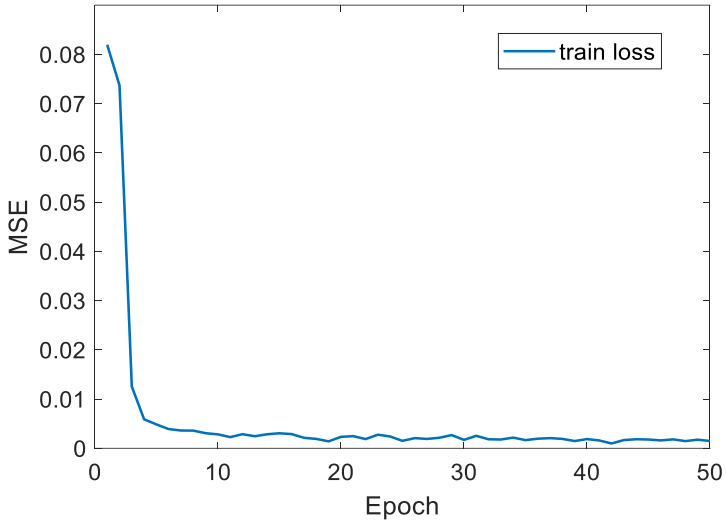


Figure 10. Network model training loss graph.

The main parameters that have an impact on the prediction performance of the network model are the convolutional kernel size and the learning rate. Among them, the convolutional kernel realizes the extraction of bearing degradation features, and too large a convolutional kernel size leads to the loss of local information, while too small a convolutional kernel cannot capture the global features. Therefore, in this paper, the convolutional kernel sizes of 3, 5, 7, 9, and 11 are selected as alternatives in turn, and the other parameters are kept unchanged to perform parameter optimization. Similarly, the learning rate is the most important parameter in the optimizer; too small a learning rate will greatly increase the training time, while too large a learning rate will cause the training process to fluctuate greatly, which is not conducive to model convergence. Therefore, the learning rates of 0.01, 0.05, 0.001, 0.005, and 0.0005 are selected as alternatives in the article, and the remaining parameters are kept constant to perform parameter optimization. The evaluation metric is chosen as the average of RMSE of the five test bearings. Figure 11 shows the learning rate and convolutional kernel size optimization search process. In Figure 11, "Lr" denotes the learning rate and "Ks" denotes the convolutional kernel size. The average RMSE is the smallest when the convolutional kernel size is 5, which means that the prediction effect is optimal at this time, so the convolutional kernel size is 5. It can also be observed that the prediction effect is optimal when the network learning rate is 0.001, so the learning rate is 0.001 in this paper.

The RUL prediction results of the method proposed in the article on the training bearing 1−1 are shown in Figure 12a. As can be observed from Figure 12a, the method better fits the training set. It can be concluded that the model learns the degraded information contained in the training set. Further, after proposing new labels, the bearing prediction has a good effect both in the early and late stages, thus, also validating the effectiveness of our proposed method. The RUL prediction results of the proposed method on test bearing 1−3 are shown in Figure 12b. As can be observed from Figure 12b, the method in the article fits the degradation trend of the bearing very well, has good monotonicity and prediction

accuracy, and can almost perfectly predict the final failure life of the bearing at the final moment.

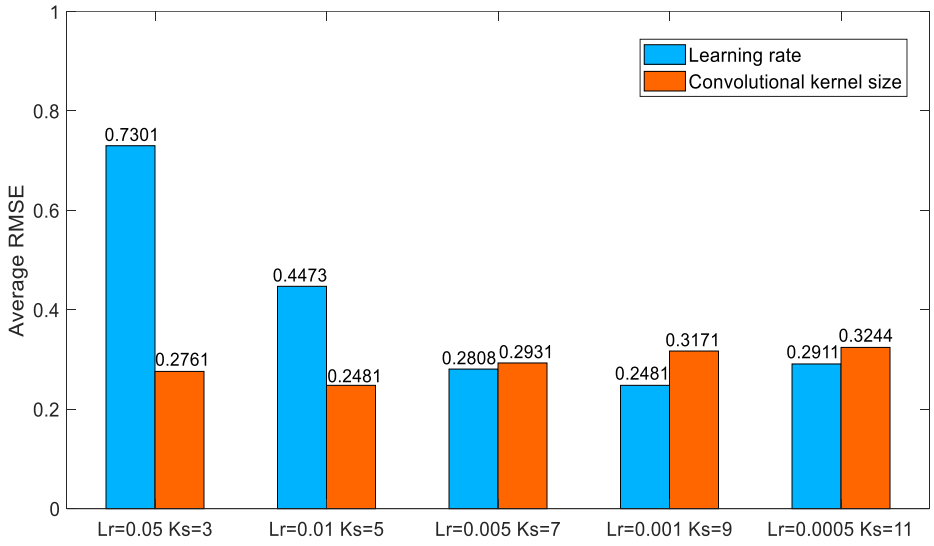


Figure 11. Learning rate and convolutional kernel size finding process.

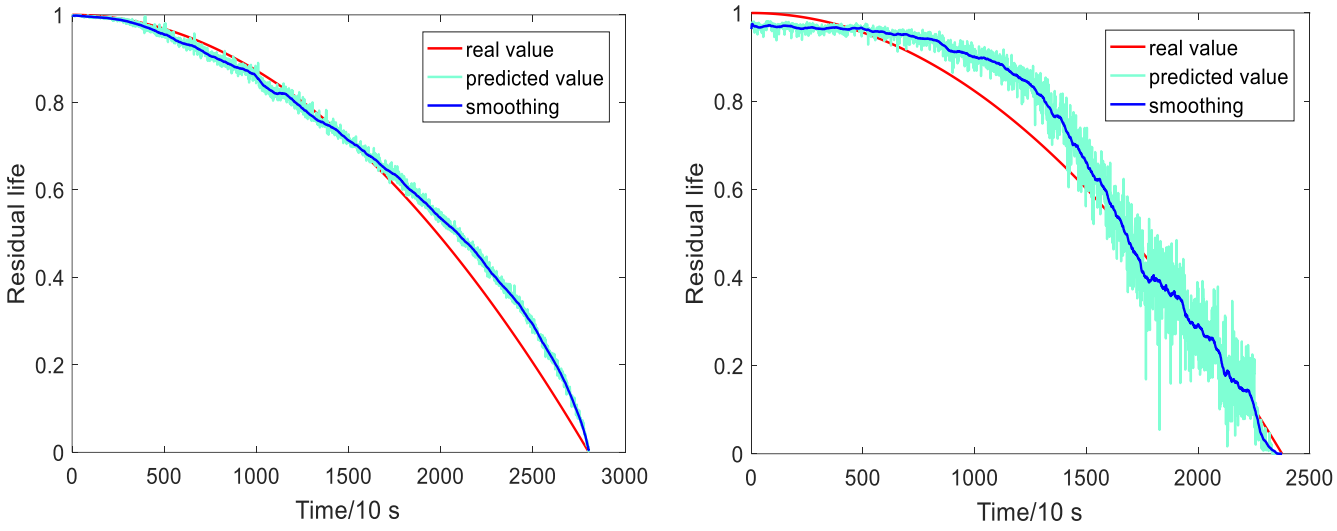


Figure 12. Prediction effect of the method in this paper on bearing 1−1 and 1−3. (**a**) Prediction effect on bearing 1−1; (**b**) Prediction effect on bearing 1−3.

4.4. Comparison Test

To verify the effectiveness and superiority of the proposed method, the residual network (ResNet), CNN-LSTM, and temporal convolutional network (TCN) are selected for comparison tests. Among them, ResNet has residual connectivity, which reduces the risk of overfitting due to the increase in network depth. The CNN-LSTM model can extract both spatial and temporal features and is widely used in RUL prediction. the TCN model has long-term memory capability and achieves better results in time series prediction. The parameters of the comparison methods selected in this paper are consistent with those of the proposed models. The RUL prediction results of each prediction method on test bearing 1−3 are shown in Figure 13, from which it is obvious that the curve of the RUL prediction method in this paper has the best fit with the curve of the real bearing life. This indicates

that the overall prediction effect of the proposed method is better than other comparison methods.

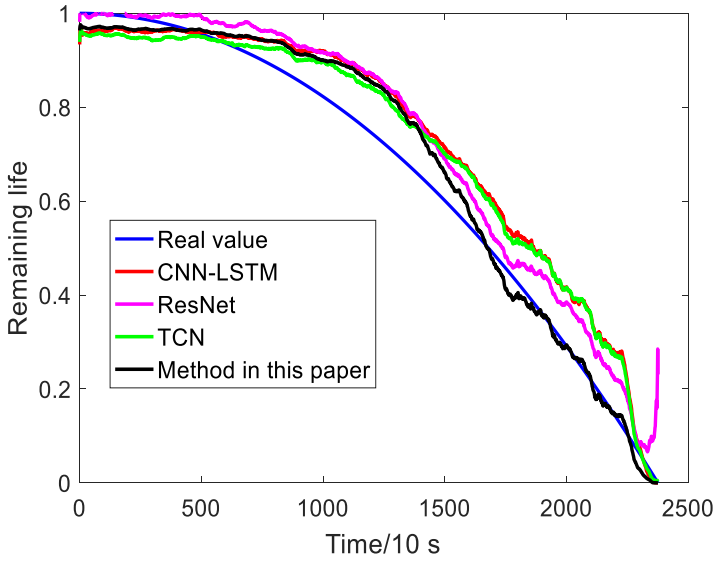


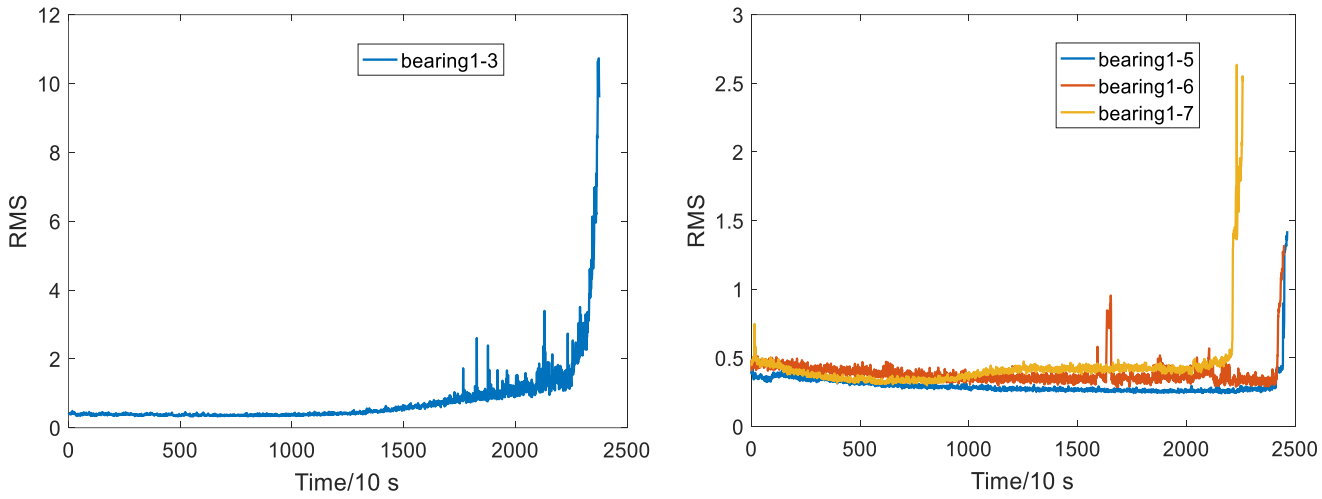
Figure 13. RUL prediction results of each method on test bearing 1−3.

RMSE and MAE are used to evaluate the prediction effectiveness of each method in the article. The RMS and MAE prediction performance indexes of each prediction method are shown in Table 3. It can be observed from Table 3 that the prediction performance indexes of the RUL prediction method proposed in this paper are optimal for all five test bearings. This further reflects the effectiveness and superiority of the proposed method in this paper. This is because this paper not only adopts the expanded convolution with a wider feeling field, but also adopts the attention mechanism to assign weights to the importance of features. This avoids the interference of useless features and enhances the utilization of effective features. In summary, the method proposed in this paper has a better prediction effect than the existing advanced methods.

Table 3. RUL prediction results of different models.

Comparison Methods	**CNN-LSTM**		**ResNet**		**TCN**		**Proposed Method**	
Test Bearing	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
Bearing 1−3	0.0794	0.0981	0.0667	0.0814	0.0737	0.0851	0.0563	0.0705
Bearing 1−4	0.1754	0.2311	0.2035	0.2789	0.1588	0.2140	0.1443	0.1689
Bearing 1−5	0.3023	0.4151	0.3252	0.4335	0.3028	0.4146	0.2522	0.3467
Bearing 1−6	0.2770	0.3850	0.2730	0.3772	0.2763	0.3845	0.2333	0.3089
Bearing 1−7	0.2753	0.3801	0.2825	0.3849	0.2843	0.3968	0.2479	0.3455

In addition, by comparing Table 3, it can be found that although the prediction effects of the method proposed on test bearings 1−5, 1−6, and 1−7 are all better than other comparable models, they are all far inferior to bearing 1−3 in terms of prediction effects. To analyze the reasons causing such results, the root means square indicators of the initial signals of each test bearing are extracted separately, and the RMS indicators of each initial bearing signal are shown in Figure 14. As can be observed from Figure 14, the RMS variation trend of the vibration signals of bearings 1−5, 1−6, and 1−7 is steeper, so they belong to the sudden failure type, while the RMS variation trend of the vibration signals of bearing 1−3 is flatter, so bearing 1−3 belongs to the gradual failure type, which is why it leads to a large difference in the prediction effect. Subsequently, migration learning can be considered for introduction into the proposed method to reduce the difference in data distribution between different failed bearings, thus, improving the prediction accuracy.



(**a**) (**b**)

Figure 14. Comparison of RMSE metrics for each bearing. (**a**) RMS index of bearing 1–3; (**b**) RMS index of bearings 1–5, 1–6, and 1–7.

5. Conclusions

In this paper, a method for predicting the remaining life of rolling bearings based on multi-scale feature extraction and attention mechanism is proposed. Firstly, this paper takes the vibration signal of the bearing as the network input and normalizes it to perform feature extraction directly from the original dataset, reducing the loss of degradation features. Secondly, quadratic function labels are constructed for the dataset to avoid the identification of the starting point of the bearing degradation stage. Thirdly, the temporal and spatial features of the bearing vibration signals are extracted using a dilated convolutional neural network and a long- and short-term memory network, respectively. Finally, a channel attention mechanism is used to assign importance to the extracted degradation features, and the mapping of bearing degradation features to remaining life labels is achieved by a fully connected layer. The effectiveness and superiority of the proposed rolling bearing residual life prediction method is verified on the PHM 2012 bearing dataset, and the tests show that the proposed method has better prediction results compared with other advanced methods.

Author Contributions: Conceptualization, C.J. and M.X.; methodology, X.L.; software, X.L.; validation, X.L., M.X. and C.L.; formal analysis, C.L.; investigation, Y.L.; resources, C.J.; data curation, Q.W. and X.L.; writing—original draft preparation, X.L.; writing—review and editing, C.J. and M.X.; funding acquisition, C.J., Y.L. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Development Plan Project of Jilin Province, grant number 20220203041SF.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Changchun University of Technology.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, C.-C.; Liu, Z.; Yang, G.; Wu, C.-C.; Ye, Q. An Improved Fault Diagnosis Using 1D-Convolutional Neural Network Model. *Electronics* **2020**, *10*, 59. [CrossRef]
2. Gu, K.; Zhang, Y.; Liu, X.B.; Li, H.; Ren, M.F. DWT-LSTM-Based Fault Diagnosis of Rolling Bearings with Multi-Sensors. *Electronics* **2021**, *10*, 2076. [CrossRef]
3. Zhu, J.; Chen, N.; Peng, W.W. Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3208–3216. [CrossRef]
4. Tian, Z.G.; Liao, H.T. Condition based maintenance optimization for multi-component systems using proportional hazards model. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 581–589. [CrossRef]
5. Xie, Z.; Du, S.; Lv, J.; Deng, Y.; Jia, S. A Hybrid Prognostics Deep Learning Model for Remaining Useful Life Prediction. *Electronics* **2020**, *10*, 39. [CrossRef]
6. Zio, E.; Peloni, G. Particle filtering prognostic estimation of the remaining useful life of nonlinear components. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 403–409. [CrossRef]
7. Chen, C.C.; Zhang, B.; Vachtsevanos, G. Prediction of Machine Health Condition Using Neuro-Fuzzy and Bayesian Algorithms. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 297–306. [CrossRef]
8. Malhi, A.; Yan, R.Q.; Gao, R.X. Prognosis of Defect Propagation Based on Recurrent Neural Networks. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 703–711. [CrossRef]
9. Ren, L.; Sun, Y.Q.; Cui, J.; Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *J. Manuf. Syst.* **2018**, *48*, 71–77. [CrossRef]
10. Zhang, C.; Lim, P.; Qin, A.K.; Tan, K.C. Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics. *IEEE Trans. Neural Netw. Learn Syst.* **2017**, *28*, 2306–2318. [CrossRef]
11. Toldinas, J.; Venčkauskas, A.; Liutkevičius, A.; Morkevičius, N. Framing Network Flow for Anomaly Detection Using Image Recognition and Federated Learning. *Electronics* **2022**, *11*, 3138. [CrossRef]
12. Cui, J.-W.; Du, H.; Yan, B.-Y.; Wang, X.-J. Research on Upper Limb Action Intention Recognition Method Based on Fusion of Posture Information and Visual Information. *Electronics* **2022**, *11*, 3078. [CrossRef]
13. Liu, J.; Cong, R.; Wang, X.; Zhou, Y. Link-aware Frame Selection for Effificient License Plate Recognition in Dynamic Edge Networks. *Electronics* **2022**, *11*, 3186. [CrossRef]
14. Guo, L.; Lei, Y.G.; Li, N.P.; Yan, T.; Li, N.B. Machinery health indicator construction based on convolutional neural networks considering trend burr. *Neurocomputing* **2018**, *292*, 142–150. [CrossRef]
15. Wang, H.; Peng, M.J.; Miao, Z.; Liu, Y.K.; Ayodeji, A.; Hao, C. Remaining useful life prediction techniques for electric valves based on convolution auto encoder and long short term memory. *ISA Trans.* **2021**, *108*, 333–342. [CrossRef]
16. Zhang, J.M.; Lu, C.Q.; Wang, J.; Wang, L.; Yue, X.G. Concrete Cracks Detection Based on FCN with Dilated Convolution. *Appl. Sci.* **2019**, *9*, 2686. [CrossRef]
17. Wang, Y.J.; Wang, G.D.; Chen, C.L.Z.; Pan, Z.K. Multi-scale dilated convolution of convolutional neural network for image denoising. *Multimed. Tools Appl.* **2019**, *78*, 19945–19960. [CrossRef]
18. Guo, L.; Li, N.P.; Jia, F.; Lei, Y.G.; Lin, J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **2017**, *240*, 98–109. [CrossRef]
19. Li, F.; Xiang, W.; Wang, J.; Zhou, X.; Tang, B. Quantum weighted long short-term memory neural network and its application in state degradation trend prediction of rotating machinery. *Neural Netw.* **2018**, *106*, 237–248. [CrossRef]
20. Ma, M.; Mao, Z. Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1658–1667. [CrossRef]
21. Tian, F.; Wang, L.; Xia, M. Signals Recognition by CNN Based on Attention Mechanism. *Electronics* **2022**, *11*, 2100. [CrossRef]
22. Chen, Y.H.; Peng, G.L.; Zhu, Z.Y.; Li, S.J. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Appl. Soft Comput.* **2020**, *86*, 105919. [CrossRef]
23. Wen, L.; Li, X.Y.; Gao, L.; Zhang, Y.Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5990–5998. [CrossRef]
24. Yao, D.C.; Li, B.Y.; Liu, H.C.; Yang, J.W.; Jia, L.M. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. *Measurement* **2021**, *175*, 109166. [CrossRef]
25. Zhang, Z.; Wang, X.; Jung, C. DCSR: Dilated Convolutions for Single Image Super-Resolution. *IEEE Trans. Image Process.* **2019**, *28*, 1625–1635. [CrossRef]
26. Spuhler, K.; Serrano-Sosa, M.; Cattell, R.; DeLorenzo, C.; Huang, C. Full-count PET recovery from low-count image using a dilated convolutional neural network. *Med. Phys.* **2020**, *47*, 4928–4938. [CrossRef]
27. Wang, F.T.; Liu, X.F.; Deng, G.; Yu, X.G.; Li, H.K.; Han, Q.K. Remaining Life Prediction Method for Rolling Bearing Based on the Long Short-Term Memory Network. *Neural Process. Lett.* **2019**, *50*, 2437–2454. [CrossRef]
28. Guo, R.X.; Wang, Y.; Zhang, H.C.; Zhanga, G.L. Remaining Useful Life Prediction for Rolling Bearings Using EMD-RISI-LSTM. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
29. Choi, H.; Cho, K.; Bengio, Y. Fine-grained attention mechanism for neural machine translation. *Neurocomputing* **2018**, *284*, 171–176. [CrossRef]

30. Xu, X.W.; Wang, J.W.; Zhong, B.F.; Ming, W.W.; Chen, M. Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism. *Measurement* **2021**, *177*, 109254. [CrossRef]
31. Singleton, R.K.; Strangas, E.G.; Aviyente, S. Extended Kalman Filtering for Remaining-Useful-Life Estimation of Bearings. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1781–1790. [CrossRef]