Received: 30 October 2022 | Revised: 24 February 2023 | Accepted: 25 February 2023

DOI: 10.1002/qre.3314

SPECIAL ISSUE ARTICLE

WILEY

A CNN-BiLSTM-Bootstrap integrated method for remaining useful life prediction of rolling bearings

Junyu Guo[1,2,3] | **Jiang Wang**[1,2,3] | **Zhiyuan Wang**[1,2,3] | **Yu Gong**[1,2,3] | **Jinglang Qi**[1,2,3] | **Guoyang Wang**[4] | **Changping Tang**[5]

[1]Key Laboratory of Oil & Gas Equipment, Ministry of Education, Southwest Petroleum University, Chengdu, Sichuan, P. R. China

[2]School of Mechatronic Engineering, Southwest Petroleum University, Chengdu, Sichuan, P. R. China

[3]Oil and Gas Equipment Technology Sharing and Service Platform of Sichuan Province, Southwest Petroleum University, Chengdu, Sichuan, P. R. China

[4]School of Computer Science, Southwest Petroleum University, Chengdu, Sichuan, P. R. China

[5]Sichuan Honghua Petroleum Equipment Co., Ltd., Guanghan, Sichuan, P. R. China

Correspondence
Junyu Guo, Key Laboratory of Oil & Gas Equipment, Ministry of Education, Southwest Petroleum University, Chengdu, Sichuan 610500, P. R. China.
Email: junyguo@swpu.edu.cn

Funding information
Sichuan Science and Technology Program, Grant/Award Number: 2022YFQ0012; Joint Fund of Key Laboratory of Oil & Gas Equipment, Ministry of Education (Southwest Petroleum University); Honghua Group Co., Ltd, Grant/Award Number: OGEHH202005; Scientific Research Starting Project of SWPU, Grant/Award Number: 2019QHZ007

Abstract

Rolling bearings, an essential fundamental component in machinery and equipment, have been widely used. Predicting the remaining useful life (RUL) of rolling bearings helps maintain the reliability of mechanical systems. Accurate prediction of RUL requires extracting deep features in complex non-linear vibration signals, the prediction results often vary widely. This paper proposes a RUL prediction method based on convolutional neural network (CNN), bi-directional long-short term memory (BiLSTM), and bootstrap method (CNN-BiLSTM-Bootstrap) to model the forecasting uncertainty. The first step is to extract the first prediction time (FPT) of the degradation phase of rolling bearings using an adaptive method for the 3σ intervals of rolling bearing vibration signal kurtosis. The model extracts the spatio-temporal features through CNN and BiLSTM, and combines the bootstrap method to quantify the RUL prediction interval (PI) of rolling bearings. The comparison with existing models verified the effectiveness and generalization of the proposed model.

KEYWORDS

bidirectional long-short term memory network, bootstrap method, convolutional neural network, remaining useful life prediction, rolling bearings

1 | INTRODUCTION

Rolling bearings is a core component of rotating machinery and equipment, in the application of mechanical industrial products have an irreplaceable position, it can withstand the load and transfer kinetic energy, and its working condition directly determines the overall performance of the equipment.[1,2] However, rolling bearings are wearing parts, as one of the zero parts of rotating machinery that is highly susceptible to failure, when reaching their life limit, the working

accuracy of mechanical equipment will decline, the stability of the equipment will be sharply reduced, increasing the incidence of accidents and increasing maintenance costs, which will quickly lead to equipment failure.[3,4] Therefore, predicting the remaining useful life (RUL) of rolling bearings prevents accidents in machinery and equipment and provides support for safety and reliability decisions in the later stages of equipment maintenance planning. How to effectively predict the remaining service life of rolling bearings is of great importance for the health management of machinery and equipment.[5–7]

In recent years, the issue of equipment lifetime has become increasingly complex due to the highly non-linear and complex nature of mechanical systems.[8] Prognostics and health management (PHM) have received much attention as they can assess the system status by monitoring and analyzing data and evaluating its RUL,[9–12] thus significantly improving the efficiency of operation and maintenance. The main methods used to predict the RUL of rolling bearings are model-driven, data-driven and hybrid-based methods.[13] The model-driven approach uses models derived from damage rules and failure mechanism to describe the degradation process of rolling bearings and predict the RUL.[14] As model-driven prediction methods require complex mathematical models in the system based on the physical principles of bearing degradation and require a large amount of a priori knowledge, there are limitations to their application and dissemination.[15] The hybrid approach aims to address the limitations of data-driven and model-based approaches. Still, due to the complexity of actual engineering equipment, the variety of operating conditions, the fact that their degradation patterns are often difficult to model, and the high cost of obtaining a failure mechanism model, there needs to be more research available. However, data-driven approaches can infer the health of bearings by constructing a degradation assessment of current bearing performance based on data collected by sensors.[16] It does not depend on the specific degradation patterns and only needs sufficient historical data, which is more suitable for RUL prediction of complex modeling systems. Traditional machine learning methods to predict the RUL of rolling bearings by extracting shallow features of the data, such as using support vector machine (SVM),[17] artificial neural network (ANN)[18] and random forest regression (RFR)[19,20] to predict the RUL have also achieved some success. However, traditional machine learning methods rely on complex feature engineering techniques, with limited deep-seated data mining features.

Deep learning methods can learn representative features containing a large amount of information from complex original data, with powerful feature extraction[21] and non-linear fitting capabilities, widely used in RUL prediction.[22,23] For example, convolutional neural network (CNN)[24] and recurrent neural network (RNN)[25] have achieved better results in bearing RUL by partitioning the sensor data into a series of sliding windows and corresponding to a RUL label value independently of the other windows. However, Traditional CNN is not suitable for modeling time series problems as they cannot capture long-term dependent information to a large extent due to their convolutional kernel size limitations. And as the evolution of faults is a gradual process during equipment operation,[26] the ability of the constructed model to learn adequate temporal information is crucial for the prediction results.[27] The effectiveness of the RUL prediction method using long-short term memory (LSTM)[28,29] and bi-directional long-short term memory (BiLSTM)[30,31] has also been verified. Li[32] proposed a method that combines CNN and LSTM for concurrent feature extraction and RUL prediction. Shen[33] extracted signal characteristics, constructed non-linear degradation index, and integrated multi-head attention mechanism and BiLSTM network to predict the RUL of rolling bearings. In the above prediction tasks, most methods can only give deterministic RUL prediction results, no longer suitable for describing the uncertainties.[34] Based on the bootstrap method, rolling bearings' RUL prediction interval (PI) can be effectively quantified. Based on the integrated prediction method of bootstrap, Huang[35] embedded the network into the bootstrap framework and realized the quantification of RUL PI. Zhao[36] proposed a novel bootstrap ensemble learning convolutional simple recurrent unit method for RUL prediction. A single network model cannot achieve outstanding results between data feature extraction and information relevance. By overlaying multiple networks, the shortcomings of a single network can be overcome and have better generalization performance. To overcome these limitations, A CNN-BiLSTM network with a bootstrap method for the RUL prediction method is constructed in this paper. Contributions of this paper are summarized as follows:

1. Adaptive method based on the 3σ interval of the kurtosis of rolling bearing, dividing the normal operation phase and the degradation phase of rolling bearing and extracting the first prediction time (FPT) of the degradation phase of rolling bearing.
2. One-dimensional CNN extracted the spatial features and input the reduced-dimensional information sequence into BiLSTM for further time feature extraction, CNN combined with BiLSTM to achieve spatio-temporal feature fusion. The constructed CNN-BiLSTM combines the bootstrap method to quantify the RUL PI of rolling bearings, this method was verified to have better RUL prediction by comparison with the remaining commonly used models.

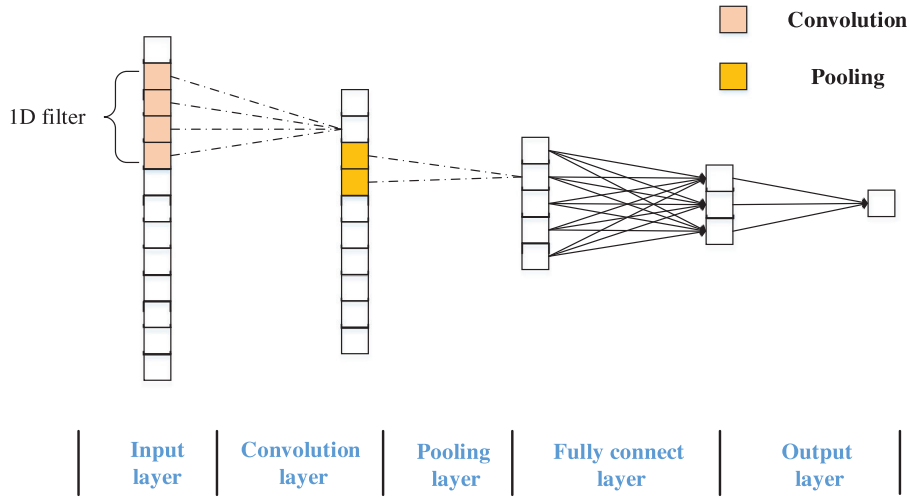

FIGURE 1 The structure of one-dimensional convolutional neural network (CNN).

3. The adaptability of the CNN-BiLSTM-Bootstrap model was verified by predicting the RUL of rolling bearings under different load conditions.

The rest of the paper is organized as follows. Section 2 presents some theoretical background, including CNN, BiLSTM and the bootstrap method. Section 3 describes the general implementation of the CNN-BiLSTM-Bootstrap model in detail. In Section 4, the effectiveness and generalization of the CNN-BiLSTM-Bootstrap model for life prediction are verified by experimenting with the bearing dataset of XJTU-SY and comparing the RUL prediction performance with other networks. Section 5 is the conclusion of this paper and the prospect of future work.

2 | THEORETICAL FOUNDATION

2.1 | Convolutional neural network

CNN is a feed-forward neural network with convolutional computation and deep structure.[37] CNN has tremendously succeeded in practical applications, such as image identification, image classification, automatic natural language processing, and sentence classification.[38] CNN's unique structure can extract local and low-level features from the raw data. Compared with the traditional neural network, CNN is inspired by local receptive fields and therefore has three characteristics: local connection, weight sharing, pooling operation and multi-level structure. These three characteristics are discussed separately below[39]:

1. Local connection: Neurons in the current layer of the CNN are only partially connected to neurons in the previous layer to learn local features. The size of the input area connected by each implicit unit is called the receptive field of the neuron. Although each output unit is only connected to a part of the input, the values are calculated unchanged, all being the dot product of the weights and the input, followed by a bias.
2. Weight sharing: The convolutional kernel in the convolutional layer is repeatedly applied to the entire receptive field, sharing the same parameters and generating a feature map. The sharing of weights dramatically reduces the number of free variables to be learned, allowing us to perform feature extraction more efficiently.
3. Pooling operation and multi-level structure: The Pooling operation is conducive to reducing the amount of data, speeding up operations and making some of the detected features more robust. The multi-level structure facilitates feature merging.

Due to the one-dimensional time-domain features of rolling bearings, we use one-dimensional CNN to process the data. The structure of one-dimensional CNN is shown in Figure 1.

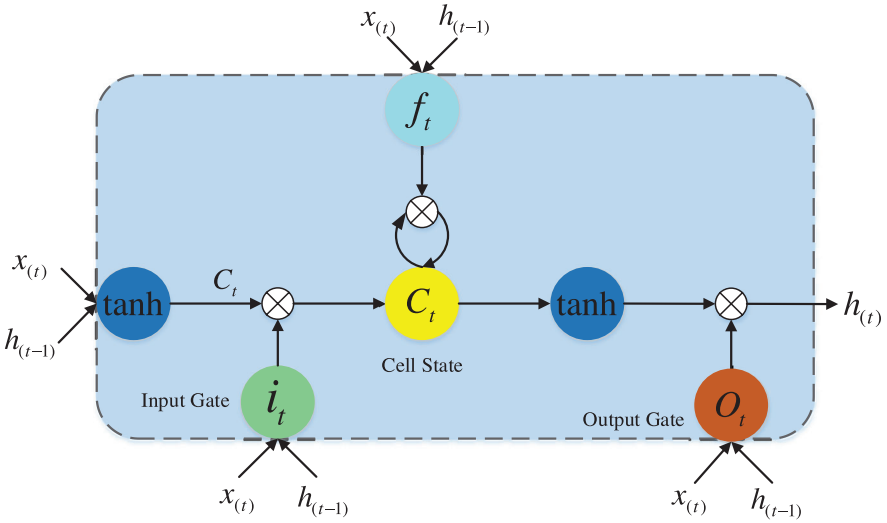

FIGURE 2 Structure of the long-short term memory (LSTM).

2.2 | Bi-directional long-short term memory

The LSTM network is a temporal RNN designed to solve the long-term dependency problem in traditional RNN. The LSTM network has a special talent for learning long-term dependence information by introducing the three gating concepts.[40] The three gating concepts of the LSTM network are input gate i_t, forget gate f_t, output gate O_t, respectively. As shown in Figure 2 is the framework of the LSTM network.

The LSTM networks can be expressed as follows:

The σ_s represents the gate-action function based on the sigmoid function, which compresses the inputs to the range [0,1]. If the output is 0 indicates all information is discarded, and 1 represents all information is input to the cell. Can be shown as:

$$\sigma_s = \frac{1}{1+e^{-t}} \tag{1}$$

The i_t is the input gate, it controls the transfer of which information should be imported in the cell, can be shown as:

$$i_t = \sigma_s(W_{xi}x_{(t)} + W_{hi}h_{(t-1)} + b_i) \tag{2}$$

The f_t is the forget gate that controls which previous state should be discarded, as shown in:

$$f_t = \sigma_s(W_{xf}x_{(t)} + W_{hf}h_{(t-1)} + b_f) \tag{3}$$

The C_t is the internal state, which delegates the temporary cell state, it can be calculated as:

$$C_t = \tanh(W_{xc}x_{(t)} + W_{hc}h_{(t-1)} + b_c) \tag{4}$$

$$C_t = f_t \odot C_{(t-1)} + i_t \odot C_t \tag{5}$$

The O_t is the output gate that in command of the output h_t of the LSTM cell, can be shown as:

$$O_t = \sigma_s(W_{xo}x_{(t)} + W_{ho}h_{(t-1)} + b_o) \tag{6}$$

$$h_t = O_t \odot \tanh(c_t) \tag{7}$$

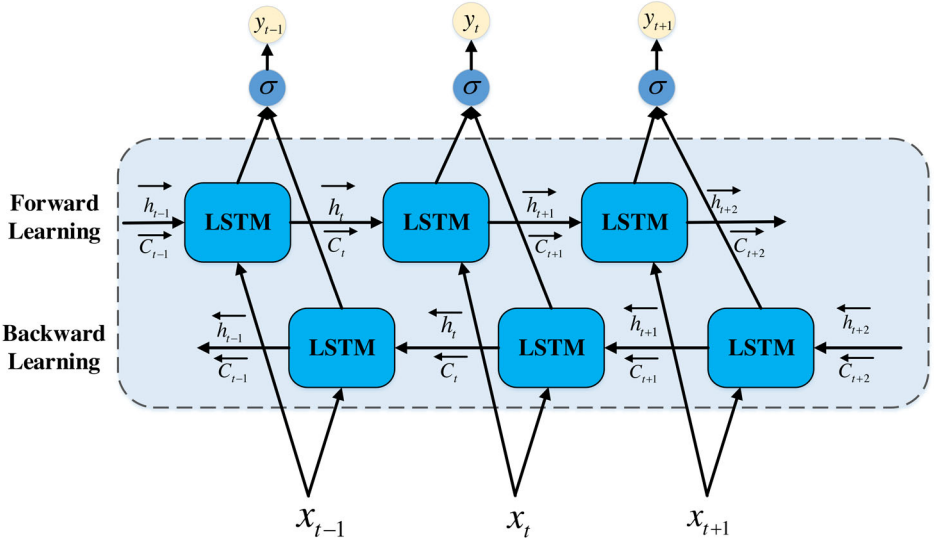

FIGURE 3 The architecture of the bi-directional long-short term memory (BiLSTM).

where $x_{(t)}$ and $h_{(t)}$ represents the input and hidden states at the time t respectively, W_{xi}, W_{xf}, W_{xc}, W_{xo} are the corresponding weight matrices associated with $x_{(t)}$, and W_{hi}, W_{hf}, W_{hc}, W_{ho} denotes the weight matrices associated with $h_{(t-1)}$; b_i, b_f, b_c, b_o are the bias vectors of the LSTM cell.

In order to exact more available information from the time-domain features, the BiLSTM is adopted in this paper. The main structure of a two-way network is composed of two one-way networks. At each moment, the input is supplied to both networks in opposite directions, while the output is determined jointly by the two unidirectional networks. As shown in Figure 3, the BiLSTM can extract contextual information through both forward and backward LSTM layers.[41]

The forward LSTM layer generates an output sequence while the backward LSTM layer exports $\overleftarrow{h_t}$. Finally, the BiLSTM layer generates an output vector which is called $Y = [y_1, y_2, y_3, \cdots, y_t]$,and each element is calculated as:

$$y_t = \sigma(\overrightarrow{h_t}, \overleftarrow{h_t}) \tag{8}$$

where the σ function is used to combine the two sequences $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$.

2.3 | Bootstrap method

The bootstrap method is a primary method for interval estimation of statistics in non-parametric statistics. The bootstrap method could resample the same size data from the original dataset by replacement.[42] When the above sampling process is repeated N times, N bootstrap samples can be generated. Then, N bootstrap samples were used as the input to N regression models for training, and the output of each regression model is averaged to minimize the RUL prediction error. The average output and model variance of N regression models can be calculated as follows:

$$\overline{y_t} = \frac{1}{N}\sum_{n=1}^{N} \hat{y}_t^n \tag{9}$$

$$\sigma_{y_t}^2 = \frac{1}{N-1}\sum_{n=1}^{n} (y_t - \overline{y_t})^2 \tag{10}$$

where $\hat{y}_t^n$ is the predicted RUL of the nth regression model, N is the number of the regression model, y_t is the actual RUL.

Meanwhile, the variance of RUL prediction uncertainty used to construct the PI[43] is:

$$\sigma_t^2 = \sigma_{y_t}^2 + \sigma_e^2 \tag{11}$$

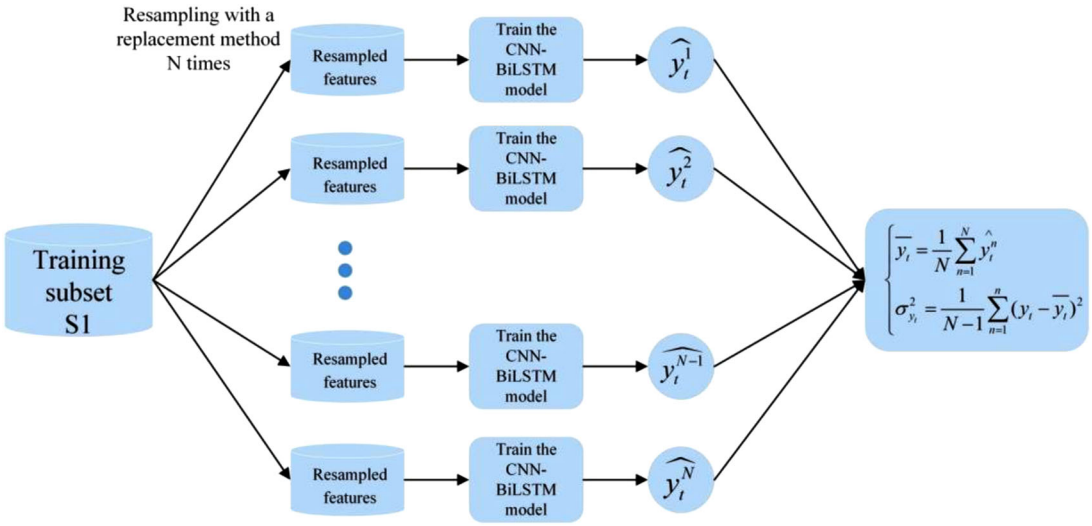

FIGURE 4 Flowchart of the bootstrap method.

where σ_t^2 denotes the variance between actual RUL y_t and predicted RUL $\hat{y}_t$; $\sigma_{y_t}^2$ denotes the model variance. σ_e^2 denotes the variance of the error term.

According to Equation (11), the range of PI can be listed as follows:

$$\begin{cases} L_{y_t} = y_t - t_d^{1-\frac{\alpha}{2}} \sqrt{\sigma_{y_t}^2 + \sigma_e^2} \\ U_{y_t} = y_t + t_d^{1-\frac{\alpha}{2}} \sqrt{\sigma_{y_t}^2 + \sigma_e^2} \end{cases} \tag{12}$$

where $t_d^{1-\frac{\alpha}{2}}$ represents the bootstrap distribution of a T-distribution function with d degrees of freedom. L_{y_t} and U_{y_t} represent the lower and upper bounds of PI

Figure 4 shows the flowchart of the bootstrap method. This paper has eight time-domain features, so the N is set to be equal to 8.

3 | CNN-BILSTM-BOOTSTRAP

The CNN-BiLSTM-Bootstrap model was constructed using deep learning techniques, Figure 5 is shown the structure of the CNN-BiLSTM-Bootstrap.

As shown in Figure 5, the original feature dataset is preprocessed using FPT recognition, mean filtering and normalization. The original feature dataset is divided into two training subsets S1 and S2. Then, training subset S1 is used to calculate the mean RUL result of the N CNN-BiLSTM. Meanwhile, the training subset S2 maps the input data and variance of the prediction error term. Finally, the models can obtain the RUL of rolling bearing and its uncertainty quantification interval.

The specific steps are as follows:

Step 1: The eight time-domain features are extracted from the raw vibration signal. The degradation stage of bearings during operation is determined by an adaptive method established based on the 3σ interval of bearing kurtosis. Then, the time-domain features are smoothed and normalized by the mean filtering method. The sliding window length L is 10.

Step 2: The original feature dataset is divided into two training subsets, S1 and S2. The training subset S1 is resampled with a replacement method N times. The resampled features are sent into the CNN-BiLSTM model for training.

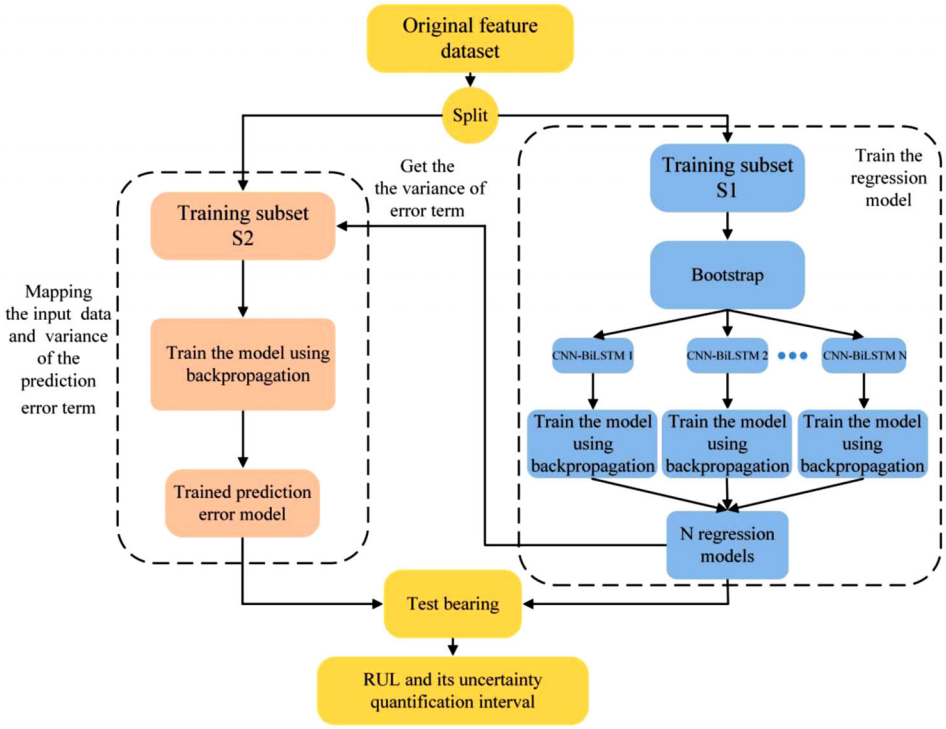

FIGURE 5 Structure of the convolutional neural network bi-directional long-short term memory and bootstrap method (CNN-BiLSTM-Bootstrap).

Step 3: After the N CNN-BiLSTM models are trained, training subset S2 is also input into the trained CNN-BiLSTM, and the RUL label of training subset S2 is known to calculate the prediction error term.

Step 4: The prediction error term and the training subset S2 are used to train the prediction error model.

Step 5: RUL and the uncertainty quantification interval of test bearings can be obtained by constructing the models mentioned above.

Figure 6 is shown the flowchart of the CNN-BiLSTM-Bootstrap method. The specific RUL prediction steps are as follows:

1. Data processing. This paper uses the XJTU-SY bearing dataset to verify the proposed method's effectiveness. An adaptive method based on the 3σ interval of bearing kurtosis is established to extract the FPT of the bearing degradation stage. Four time-domain features of each horizontal and vertical direction are extracted as the input.
2. RUL prediction using the CNN-BiLSTM-Bootstrap. The original feature dataset is divided into two training subsets. The training subset S1 is used to train the regression models, and the training subset S2 is utilized for training the prediction error model.
3. Prediction results. Test bearings are fed into the trained model to test the RUL prediction accuracy of the CNN-BiLSTM-Bootstrap model and compare it with other deep learning models to verify the benefits. Finally, bearing vibration signals collected under different working conditions were introduced to demonstrate the model's generalization ability.

4 | CASE STUDY AND ANALYSIS

4.1 | Dataset description

The experimental data are from the XJTU-SY bearing accelerated degradation experimental platform.[44] The Figure 7 shows that the platform mainly consists of a support shaft, an AC motor, two accelerometers, a hydraulic loading device and a test bearing.

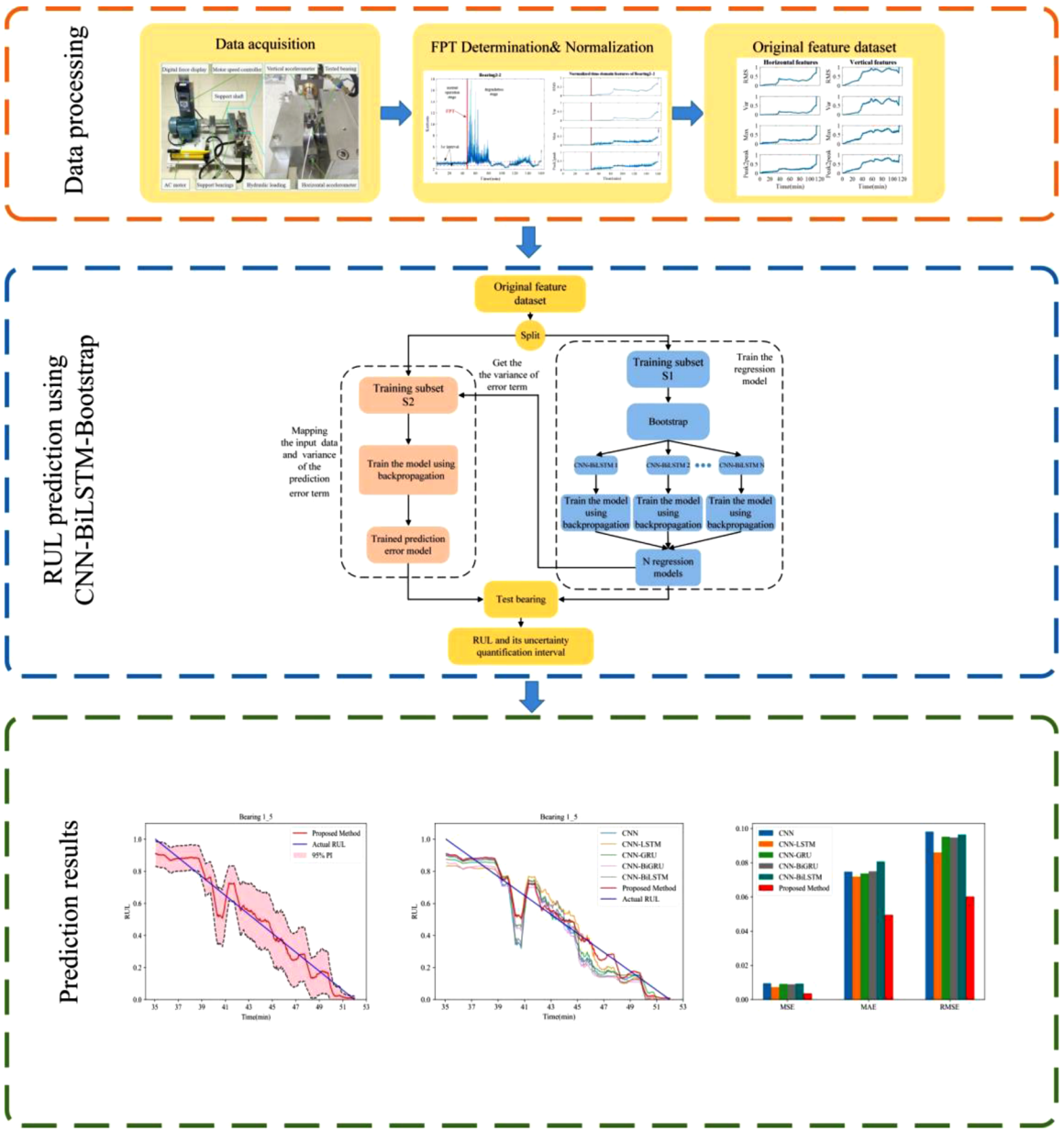

FIGURE 6 Flowchart of the proposed method.

The data used to validate the model is accurate vibration data that characterize the degradation of the bearing over its lifetime, collected at 12KN and 11KN loads, with motor speeds of 2100 and 2250 rpm, respectively and bearing type LDK UER204. The bearing vibration signal was recorded by PCB 352C33 acceleration sensor with a sampling frequency of 25.6 kHz, a sampling interval of 1 min, and a single sampling time of 1.28 s, every 32,768 data points were recorded in a single sampling. Meanwhile, the hydraulic loading device can provide a radial load to the test bearing. As the load is applied in the horizontal direction, the accelerometer placed in that direction captures more information about the degradation of the test bearing,[45] so the horizontal vibration signal is chosen to identify the FPT of each bearing.

4.2 | Data preprocessing

The kurtosis of the bearing is extremely sensitive to the early degradation stage.[46] In this paper, an adaptive method is established to select the accurate FPT based on the 3σ interval of bearing kurtosis. The implementation process of the method is as follows:

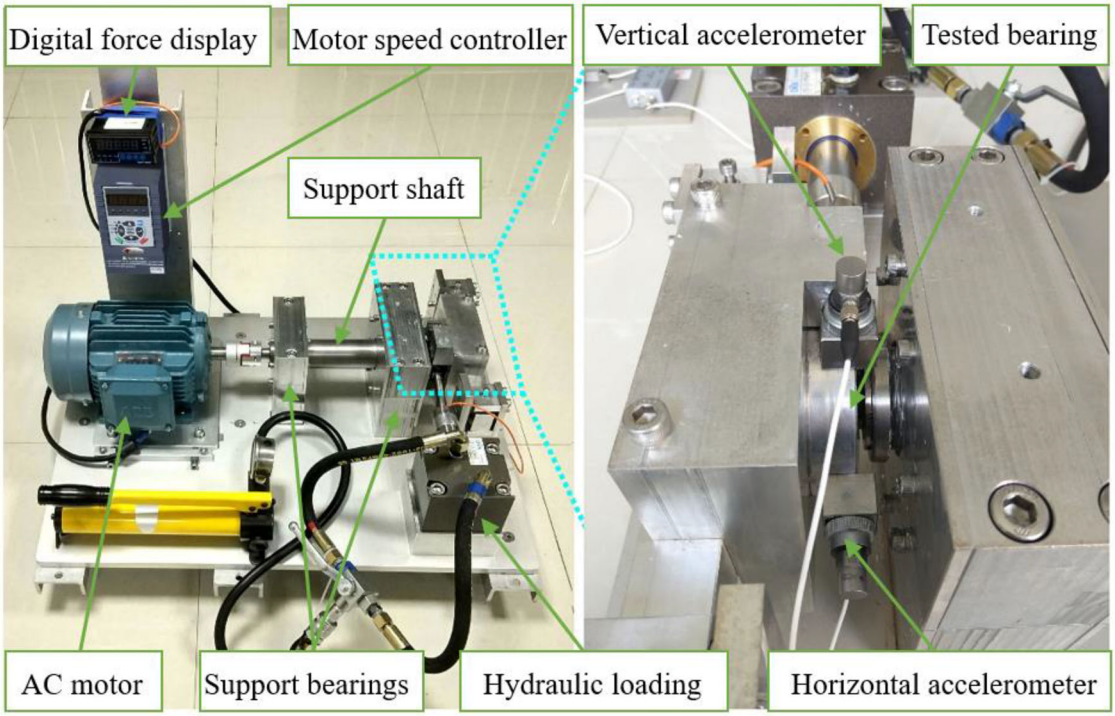

FIGURE 7 XJTU-SY experimental platform.

Step 1: The early normal operation stage is selected as the detection range based on the raw vibration data of the bearings.

Step 2: Calculate the kurtosis of the bearing, and then the mean value and the standard deviation are calculated within the detection range. Finally, the 3σ interval of the bearing kurtosis is established and noted as $[k_m - 3\sigma, k_m + 3\sigma]$.

Step 3: To restrict the interference of random noises from the raw vibration data, we define that the bearing enters the degradation stage when the kurtosis of two consecutive samples exceeds the 3σ interval. The representation is shown in Equation (13).

$$\{(k_{t-i} - k_m) > 3\sigma\}_{i=2} \tag{13}$$

where k_{t-i} represents the bearing kurtosis of the $(t-i)$th sample. k_m is the mean value of the detection range. t represents the FPT of the bearing.

To illustrate the accuracy of FPT selection, we use four time-domain features from a horizontal direction, namely RMS, Var, Max, and peak-to-peak. The identification process of FPT is shown in Figures 8 and 9.

The six rolling bearings and the FPT of each bearing is shown in Table 1. Once the FPT of each bearing is determined, the time domain features after FPT can be extracted. In this paper, four time-domain features of each horizontal and vertical directions are extracted as the input. The four time-domain features are RMS, Var, Max, and peak to peak. In dataset A, bearing 1_2 is used as the first training subset S1, and eight bootstrap samples are generated through resampling eight times. Bearing 1_1 is the second training subset S2 to establish the prediction error model. Meanwhile, bearing 1_5 is used as a test bearing. The domain adaptation capability of the model is verified by another working condition bearing dataset B. In dataset B, bearing 2_2 and bearing 2_3 is used as the training subset S1, and S2, respectively. And the test bearing is bearing 2_1.

4.3 | The architecture of the CNN-BiLSTM-Bootstrap

The nonlinear activation functions relu and tanh are applied to the CNN and BiLSTM. To reduce overfitting, L2 regularization and dropout is used in this paper. Mean squared error (MSE) is utilized as the loss function, the Adam optimizer

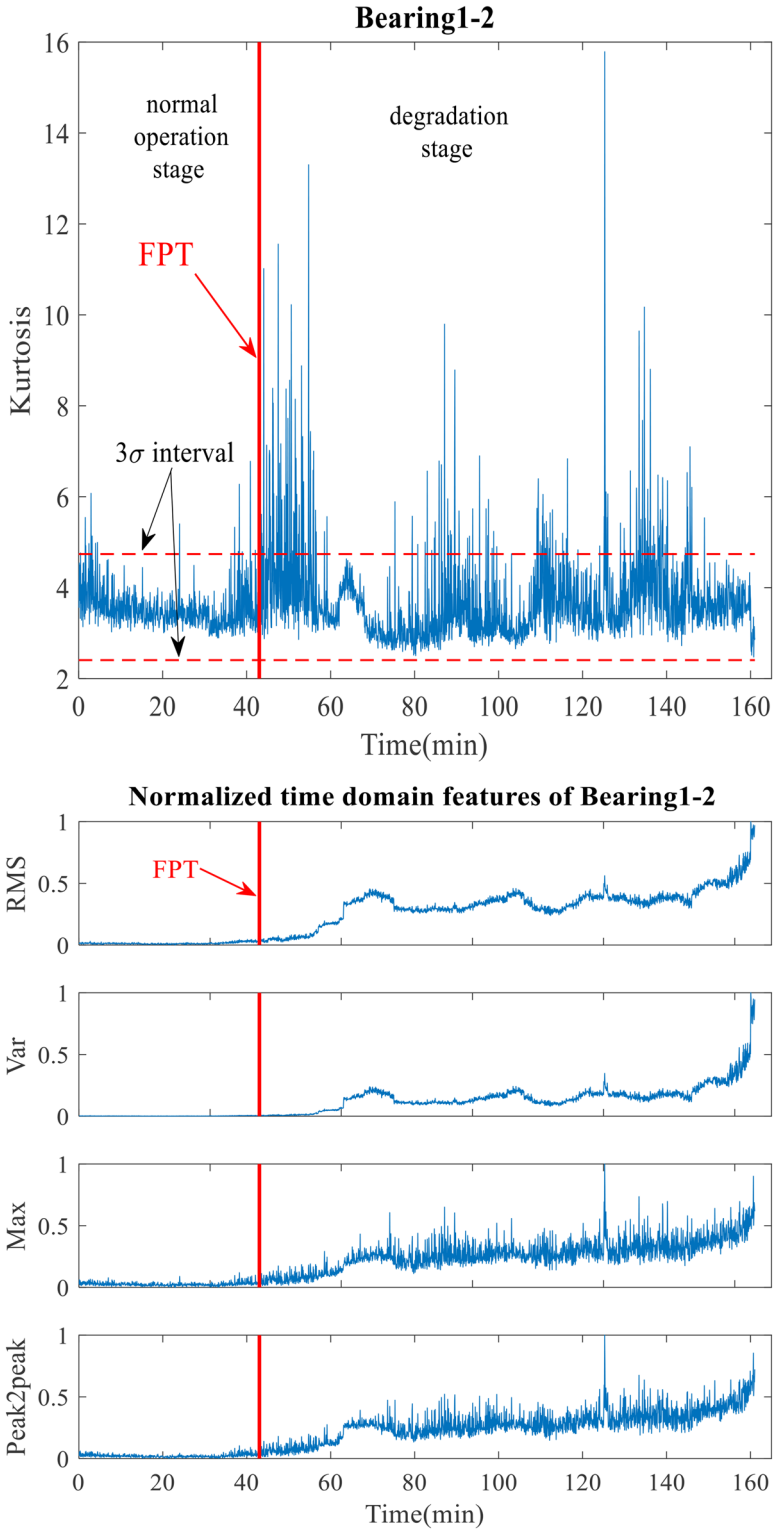

FIGURE 8 The identification process of first prediction time (FPT) for bearing 1—2.

is used to update the model parameters. The learning rate, batch size and epoch are 0.001, 64, and 20, respectively. During the training of networks, the back-propagation is used to update the model's parameters.

During the RUL prediction of bearings, extracting the features with obvious degradation process from the raw vibration signals is vital. To obtain more information about the features, the raw vibration signals are divided into some signals of 2048 data points. Then, each of these signals was processed to extract the time-domain features. Meanwhile, the structure parameters of the B regression models are shown in Table 2, Table 3 shows another structure parameters of the prediction error model.

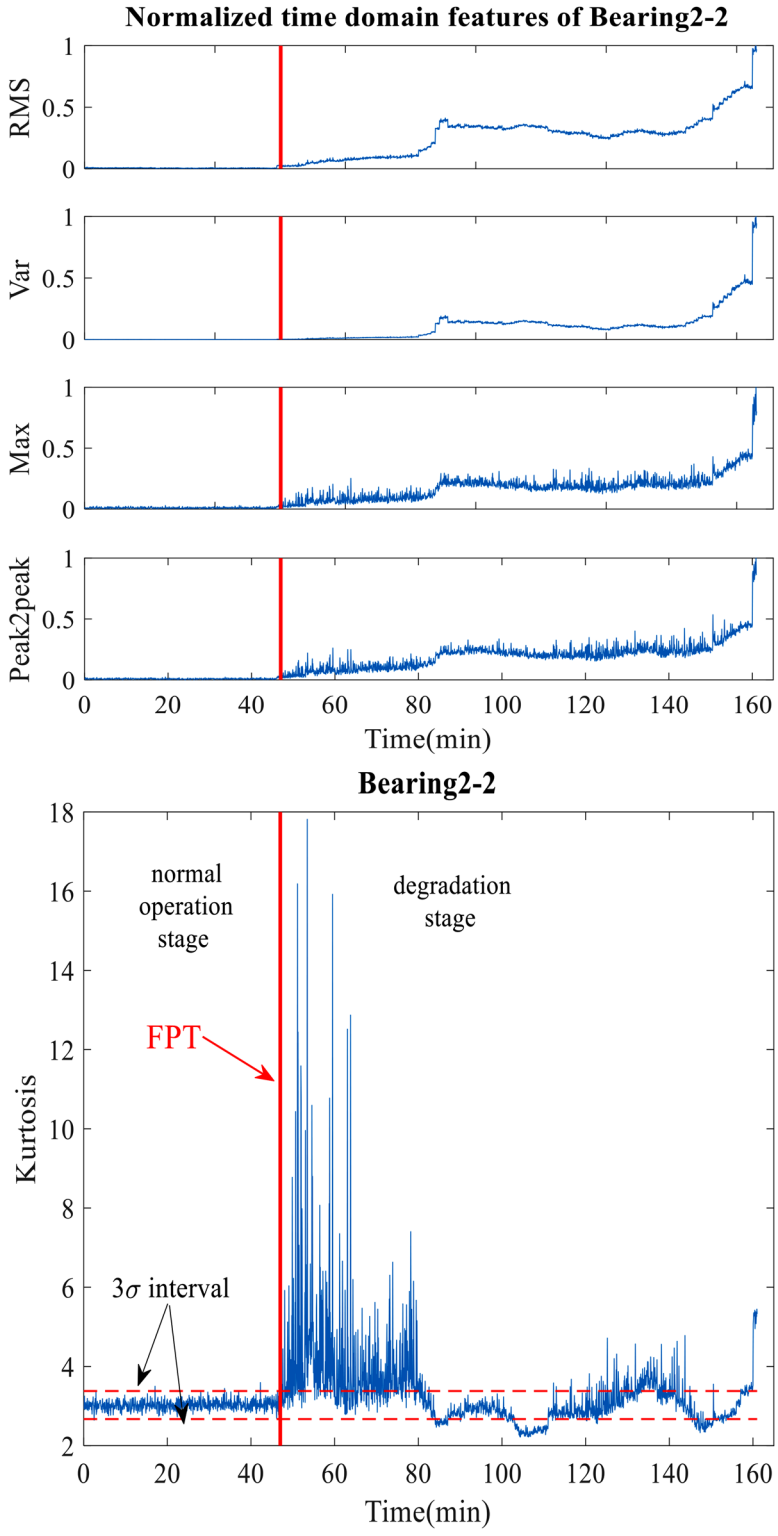

FIGURE 9 The identification process of first prediction time (FPT) for bearing 2-2.

4.4 | Model evaluation and metrics method

In this paper, MSE, mean absolute error (MAE), and root mean squared error (RMSE) are employed as performance metrics to test the validity of the model. MSE is a more straightforward representation of the model's prediction error and is often used as the model's loss function. MAE is more robust to outliers than MSE and is used to measure the mean of the residuals in the dataset. RMSE is also utilized to evaluate the RUL prediction performance comprehensively. MSE,

TABLE 1 XJTU-SY bearing dataset and the FPT of each bearing.

Dataset	Working condition	Bearing number	Run-to-failure time	Number of sampling points	FPT	Number of RUL samples
A	12KN/2100 rpm	Bearing 1_1	123 min	2048	75 min	768
		Bearing 1_2	161 min	2048	43 min	1888
		Bearing 1_5	52 min	2048	35 min	272
B	11KN/2250 rpm	Bearing 2_1	491 min	2048	454 min	592
		Bearing 2_2	161 min	2048	47 min	1824
		Bearing 2_3	533 min	2048	326 min	3312

TABLE 2 Structure parameters of the regression model.

Layer	Network structure parameter
Conv1D_1	Filters = 16, kernel = 2, stride = 2, activation function = relu
Maxpooling	pool size = 2, stride = 2
Dropout	rate = 0.2
Conv1D_2	filters = 32, kernel = 2, stride = 1, activation function = relu
Maxpooling	pool size = 2, stride = 1
Dropout	rate = 0.2
BiLSTM	32, activation function = tanh
BiLSTM	16, activation function = tanh
Dropout	rate = 0.3
Dense_1	16, activation function = relu
Dropout	rate = 0.3
Dense_2	1, activation function = relu

TABLE 3 Structure parameters of the prediction error model.

Layer	Network structure parameter
LSTM	32, activation function = relu
Dense_1	32, activation function = relu
Dense_2	16, activation function = relu
Dropout	rate = 0.3
Dense_3	1, activation function = relu

MAE, and RMSE are defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{14}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{16}$$

where N is the bearing time samples; y_i is the true value, $\hat{y}_i$ is the prediction value and $\bar{y}_i$ is the mean value. When the values of MAE, MSE and RMSE are smaller, the model's prediction performance is better.

TABLE 4 Comparison of three metrics of different methods on bearing1_5.

	CNN	CNN-LSTM	CNN-GRU	CNN-BiGRU	CNN-BiLSTM	CNN-BiLSTM-Bootstrap
MAE	0.0748	0.0719	0.0737	0.0750	0.0807	**0.0495**
MSE	0.0095	0.0073	0.0091	0.0089	0.0093	**0.0035**
RMSE	0.0982	0.0860	0.0952	0.0948	0.0965	**0.0602**

The bold values show the performance of the proposed model in this paper.

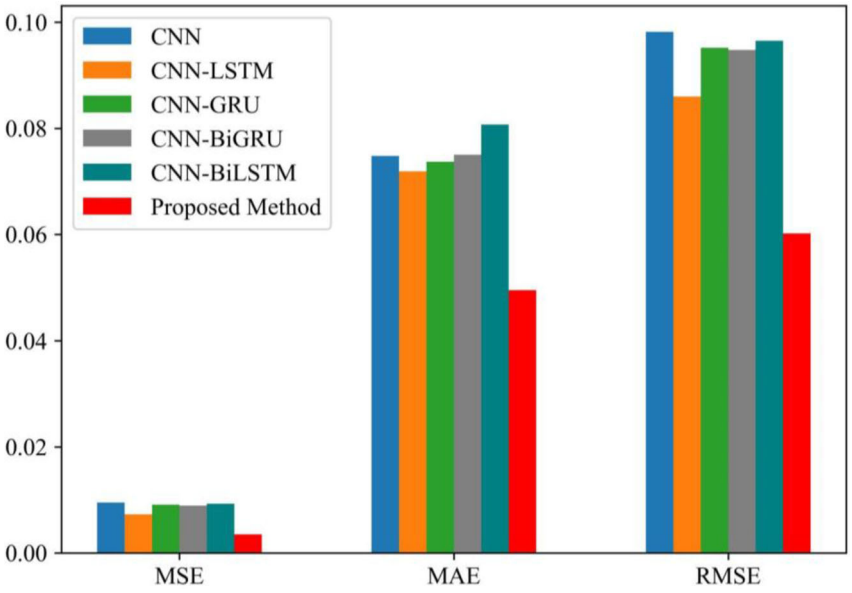

FIGURE 10 Visualization results of three metrics of different models on bearing 1_5.

4.5 | Experiment results and analysis

To verify the benefits of the CNN-BiLSTM-Bootstrap method, five kinds of deep learning models are introduced for comparison, including CNN, CNN-LSTM, CNN-GRU, CNN-BiLSTM, and CNN-BiGRU. Furthermore, as the initial weight and bias settings impacted the results of the runs, three independent experiments are conducted with the same parameter settings to reduce the effect of random initialization of parameters on prediction results. Then, the average of three metrics is taken as the prediction performance. Table 4 shows that the proposed CNN-BiLSTM-Bootstrap method outperforms the other deep learning models, the MAE, MSE, and RMSE are significantly improved compared with CNN-BiLSTM. The MAE, MSE, and RMSE decreased by up to 38.66%, 62.37%, and 37.62%, respectively. It strongly demonstrates that the bootstrap method reduces the model's error. Among the six deep learning models, the MAE, MSE, and RMSE of the CNN-BiLSTM are the worst. However, with the bootstrap method, it dramatically improves the capacity of the model and reduces the prediction error, the effectiveness of the bootstrap method is also demonstrated.

To express the methods' more intuitively, three metrics are visualized in Figure 10, which shows that the MAE, MSE and RMSE of the CNN-BiLSTM-Bootstrap method are lower than the other methods. Furthermore, the prediction results of the six methods on the RUL of bearing1_5 are shown in Figure 11. Among the six deep learning methods, it is obvious from Figure 11 that the CNN-BiLSTM-Bootstrap method is the most consistent with the actual RUL trend. In addition, Figure 12 shows the 95% PI of the CNN-BiLSTM-Bootstrap method. It can be observed that it almost contains the actual RUL, which illustrates that the uncertainty quantification interval calculated by the CNN-BiLSTM-Bootstrap method is effective. In summary, the proposed method can predict the RUL of rolling bearings accurately and precisely calculate the uncertainty quantification interval.

4.6 | Exploring the domain adaptation capability of the model

Though the performance of the CNN-BiLSTM-Bootstrap method has been validated with the other five deep learning methods, the additional experiments with another working condition are also utilized to explore the domain adaptation

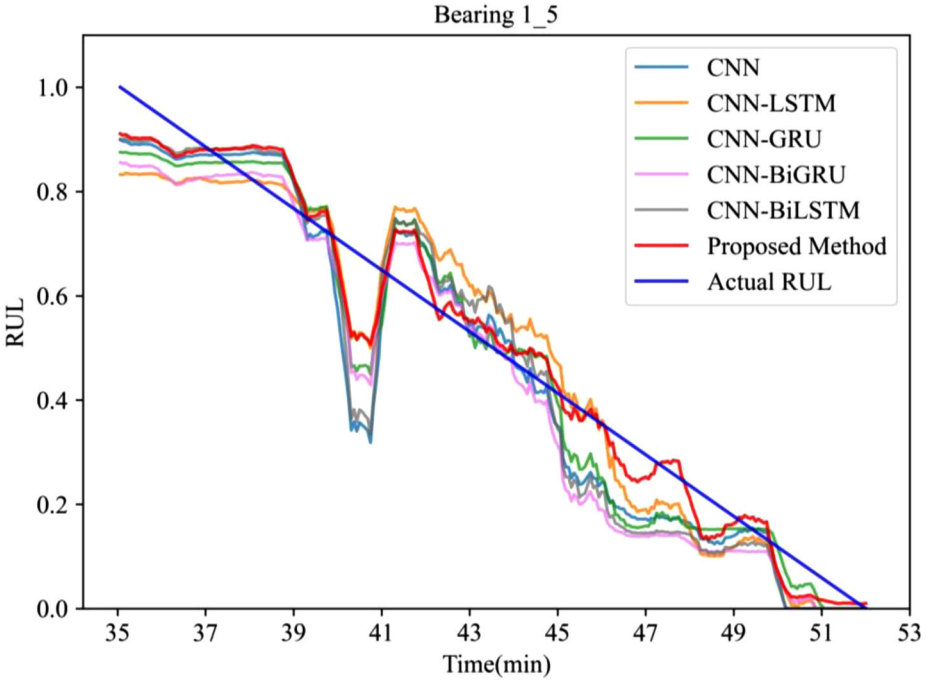

FIGURE 11 Remaining useful life (RUL) prediction results of different models for bearing1_5.

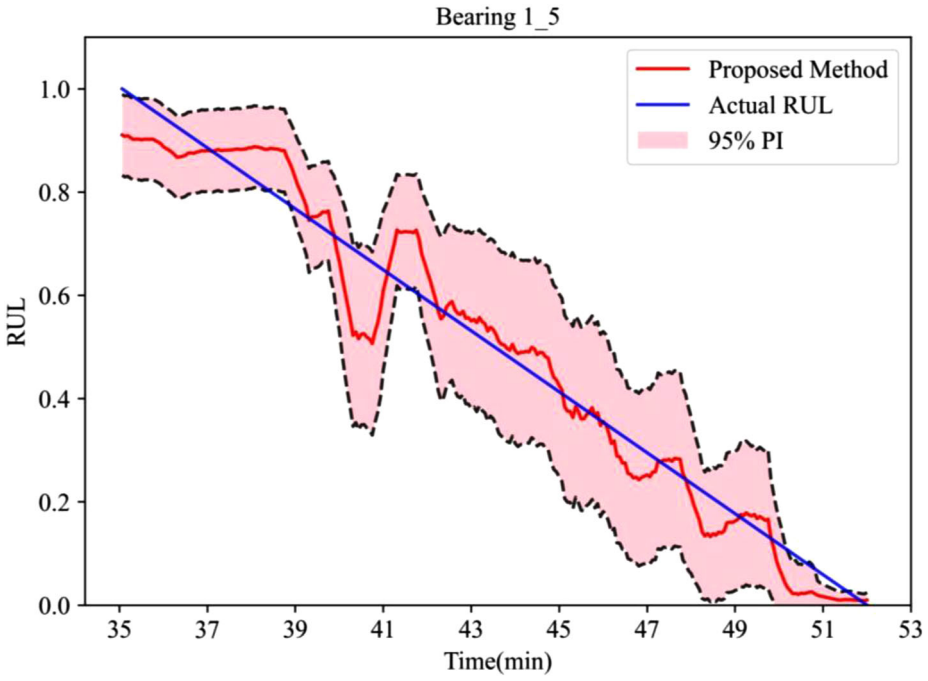

FIGURE 12 Remaining useful life (RUL) prediction results of the proposed method for bearing 1_5.

capability of the model, bearing 2_1 in dataset B corresponds to a motor load of 11KN and a speed of 2250 rpm under the working condition of the bearing. The same comparison method shown in Table 4 is used during the experiment. The comparison results are shown in Table 5, smallest MAE, MSE, and RMSE can obtain with the CNN-BiLSTM-Bootstrap method. Compared with the CNN-BiLSTM, The MAE, MSE, and RMSE decreased by up to 33.31%, 58.58%, and 35.58%, respectively. Moreover, among the other five deep learning methods, the MAE, MSE, and RMSE of CNN-BiLSTM are smaller than those of the other four models. Comparing the other five deep learning methods, the domain adaptation capability of CNN-BiLSTM-Bootstrap is verified. The results in Table 5 are visualized in Figure 13.

TABLE 5 Comparison of three metrics of different methods on bearing 2_1.

	CNN	CNN-LSTM	CNN-GRU	CNN-BiGRU	CNN-BiLSTM	CNN-BiLSTM-Bootstrap
MAE	0.1634	0.1450	0.1474	0.1656	0.1366	**0.0911**
MSE	0.0371	0.0309	0.0323	0.0403	0.0268	**0.0111**
RMSE	0.1928	0.1754	0.1785	0.2001	0.1630	**0.1050**

The bold values show the performance of the proposed model in this paper.

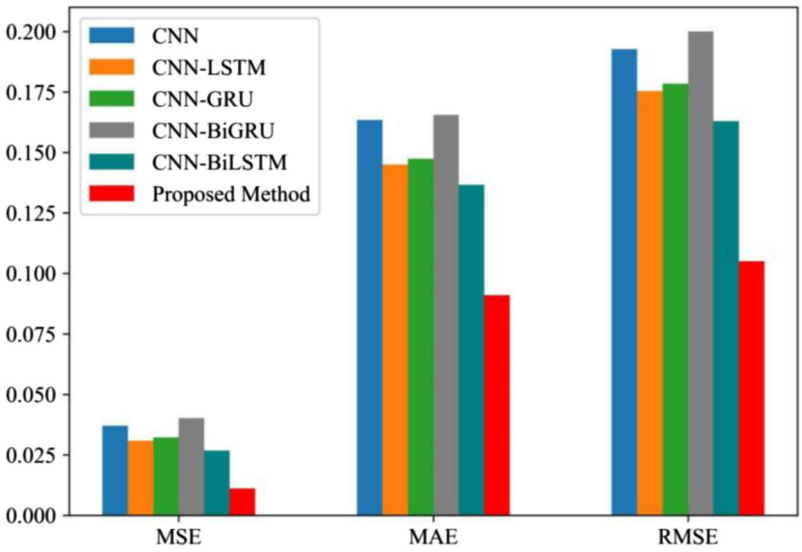

FIGURE 13 Visualization results of three metrics of different models on bearing 2_1.

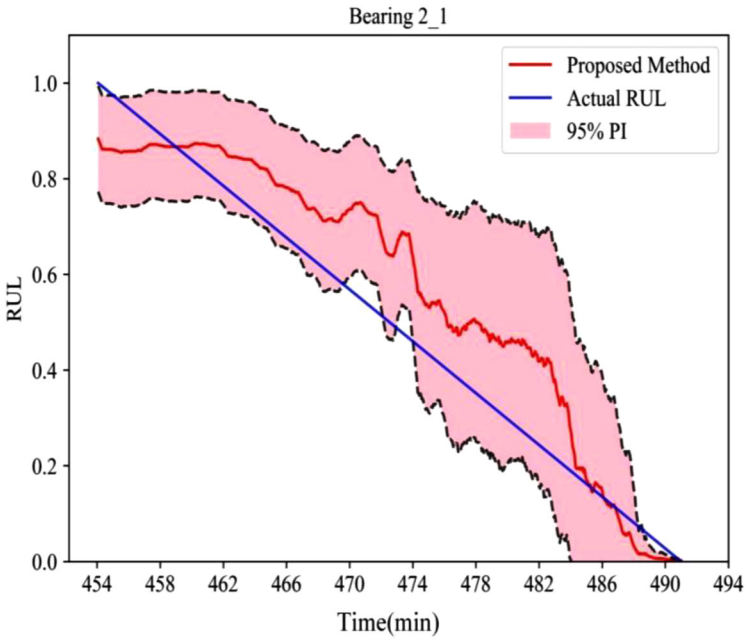

FIGURE 14 Remaining useful life (RUL) prediction results of the proposed method for bearing 2_1.

Furthermore, the prediction results of bearing 2_1 with the proposed method is shown in Figure 14. As shown in Figure 14, the RUL prediction curve does not correspond to the actual RUL curve at some points, but the curve shows the degradation trend of the rolling bearing. The main reason for this RUL prediction result may be caused by the features extracted from the raw vibration signal. While the raw vibration signal contains some noise, the feature series extracted from it are noisy and not robust. So, the RUL prediction of bearing 2_1 needs to be more accurate than the above bearing. However, the 95% PI covers the actual RUL curve and precisely calculate the uncertainty quantification interval of bearing 2_1. In summary, even though the proposed method does not achieve a particularly accurate RUL prediction of

bearing under different working condition, it precisely calculates the uncertainty quantification interval and is promising for industrial application.

5 | CONCLUSION

A CNN, BiLSTM network with bootstrap method (CNN-BiLSTM-Bootstrap) for RUL prediction method is constructed in this paper. The first prediction time (FPT) of the degradation phase of rolling bearings is extracted using an adaptive method for the 3σ interval of the kurtosis of rolling bearings. The model extracts the spatial features of the eight resampled features through a one-dimensional CNN, and feeds the reduced-dimensional information sequence into a BiLSTM for further temporal feature extraction. At last, it uses bootstrap method to calculate the average output of the eight CNN-BiLSTM models and quantifies the RUL PI. After testing the XJTU-SY bearing experiment data to demonstrate the effectiveness and availability of the CNN-BiLSTM-Bootstap model, it is verified that the model proposed in this paper has the lowest MAE, MSE, and root mean squared error (RMSE) values than the other five deep learning methods. Furthermore, the proposed method can quantify the uncertainty of RUL prediction by constructing PI, and the experiment shows that the CNN-BiLSTM-Bootstap model can precisely calculate the uncertainty quantification interval.

The CNN-BiLSTM model combines the bootstrap method to achieve high RUL prediction accuracy and precisely calculate the uncertainty quantification interval. Therefore, much effort can still be done into feature extraction and bootstrap method to promote the applicability and stability of the bearing RUL prediction under various operation conditions. In the future, the relationship between channel attention and spatial attention will be considered simultaneously, along with the extension of the model for application to RUL prediction for other equipment.

ACKNOWLEDGMENTS

This research was supported by the Sichuan Science and Technology Program (grant number 2022YFQ0012), Joint Fund of Key Laboratory of Oil & Gas Equipment, Ministry of Education (Southwest Petroleum University) and Honghua Group Co., Ltd (grant number OGEHH202005), and Scientific Research Starting Project of SWPU (grant number 2019QHZ007).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in IEEE at https://ieeexplore.ieee.org/abstract/document/8576668, reference number 44.

ORCID

Junyu Guo https://orcid.org/0000-0001-9462-9501

REFERENCES

1. Zhao HM, Liu HD, Jin Y, Dang XJ, Deng W. Feature extraction for data-driven remaining useful life prediction of rolling bearings. *IEEE Trans Instrum Meas*. 2021;70:1-10.
2. Huang HZ, Zuo MJ, Sun ZQ. Bayesian reliability analysis for fuzzy lifetime data. *Fuzzy Sets Syst*. 2006;157:1674-1686.
3. Li H, Soares CG. Assessment of failure rates and reliability of floating offshore wind turbines. *Reliab Eng Syst Saf*. 2022;228:108777.
4. Li H, Díaz H, Guedes Soares C. A failure analysis of floating offshore wind turbines using AHP-FMEA methodology. *Ocean Eng*. 2021;234:109261.
5. Liu Y, Huang HZ. Optimal selective maintenance strategy for multi-state systems under imperfect maintenance. *IEEE Trans Reliab*. 2010;59:356-367.
6. Li H, Yazdi M. *Advanced Decision-Making Methods and Applications in System Safety and Reliability Problems: Approaches, Case Studies, Multi-Criteria Decision-Making, Multi-objective Decision-Making, Fuzzy Risk-Based Models*. Springer International Publishing.
7. Huang HZ, Liu ZJ, Murthy DNP. Optimal reliability, warranty and price for new products. *IIE Trans*. 2007;39:819-827.
8. Mi J, Li YF, Peng W, Huang HZ. Reliability analysis of complex multi-state system with common cause failure based on evidential networks. *Reliab Eng Syst Saf*. 2018;174:71-81.
9. Li H, Huang CG, Guedes Soares C. A real-time inspection and opportunistic maintenance strategies for floating offshore wind turbines. *Ocean Eng*. 2022;256:111433.
10. Feng K, Ji JC, Zhang YC, Ni Q, Liu Z, Beer M. Digital twin-driven intelligent assessment of gear surface degradation. *Mech Syst Sig Process*. 2023;186:109896.
11. Li H, Guedes Soares C, Huang HZ. Reliability analysis of a floating offshore wind turbine using Bayesian Networks. *Ocean Eng*. 2020;217:107827.

12. Feng K, Ji JC, Ni Q, Beer M. A review of vibration-based gear wear monitoring and prediction techniques. *Mech Syst Sig Process*. 2023;182:109605.
13. Zhu J, Chen N, Shen C. A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. *Mech Syst Sig Process*. 2020;139:106602.
14. Wu PB, Guo JY, Wu H, Wei J. Influence of DC-link voltage pulsation of transmission systems on mechanical structure vibration and fatigue in high-speed trains. *Eng Fail Anal*. 2021;130:105772.
15. Pan ZZ, Meng Z, Chen ZJ, Gao WQ, Shi Y. A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings. *Mech Syst Sig Process*. 2020;144:106899.
16. Gu LD, Zheng R, Zhou YF, Zhang ZS, Zhao KK. Remaining useful life prediction using composite health index and hybrid LSTM-SVR model. *Qual Reliab Eng Int*. 2022;38:1-3.
17. Chen LGH, Zhang Y, Zheng Y, Li XS, Zhang XJ. Remaining useful life prediction of lithium-ion battery with optimal input sequence selection and error compensation. *Neurocomputing*. 2020;414:245-254.
18. Wu B, Tian Z, Chen M. Condition-based maintenance optimization using neural network-based health condition prediction. *Qual Reliab Eng Int*. 2013;29:1151-1163.
19. Wang FK, Mamo T. Hybrid approach for remaining useful life prediction of ball bearings. *Qual Reliab Eng Int*. 2019;35:2494-2505.
20. Wang L, Mao ZH, Xuan H, et al. Status diagnosis and feature tracing of the natural gas pipeline weld based on improved random forest model. *Int J Press Vessels Pip*. 2022;200:104821.
21. Gao DW, Zhu YS, Yan K, et al. Joint learning system based on semi-pseudo-label reliability assessment for weak-fault diagnosis with few labels. *Mech Syst Sig Process*. 2023;189:110089.
22. Zhu R, Peng WW, Wang D, Huang CG. Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data. *Mech Syst Sig Process*. 2023;183:109628.
23. Yang K, Wang YJ, Yao YN, Fan SD. Remaining useful life prediction via long-short time memory neural network with novel partial least squares and genetic algorithm. *Qual Reliab Eng Int*. 2021;37:1080-1098.
24. Shang YJ, Tang XL, Zhao GQ, Jiang PG, Ran LT. A remaining life prediction of rolling element bearings based on a bidirectional gate recurrent unit and convolution neural network. *Measurement*. 2022;202:111893.
25. Yu WN, Kim IY, Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. *Reliab Eng Syst Saf*. 2020;199:106926.
26. Chen Z, Li T, Xue XW, Zhou Y, Jing S. Fatigue reliability analysis and optimization of vibrator baseplate based on fuzzy comprehensive evaluation method. *Eng Fail Anal*. 2021;127:105357.
27. Ye R, Dai Q. A novel transfer learning framework for time series forecasting. *Knowl Based Syst*. 2018;156:74-99.
28. Su C, Li L, Wen ZJ. Remaining useful life prediction via a variational autoencoder and a time-window-based sequence neural network. *Qual Reliab Eng Int*. 2020;36:1639-1656.
29. Chang L, Lin YH, Zio E. Remaining useful life prediction for complex systems considering varying future operational conditions. *Qual Reliab Eng Int*. 2021;38:516-531.
30. Huang CG, Yin XH, Huang HZ, Li YF. An enhanced deep learning-based fusion prognostic method for RUL prediction. *IEEE Trans Reliab*. 2019;69:1097-1109.
31. Singh RM, Harsha SP. Rolling bearing prognostic analysis for domain adaptation under different operating conditions. *Eng Fail Anal*. 2022;139:106414.
32. Li J, Li X, He D. A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction. *IEEE Access*. 2019;7:75464-75475.
33. Shen YZ, Tang BP, Li B, Tan Q, Wu YL. Remaining useful life prediction of rolling bearing based on multi-head attention embedded Bi-LSTM network. *Measurement*. 2022;202:111803.
34. Li YF, Huang HZ, Mi JH, Peng WW, Han XM. Reliability analysis of multi-state systems with common cause failures based on Bayesian network and fuzzy probability. *Ann Oper Res*. 2022;311:195-209.
35. Huang CG, Huang HZ, Li YF, Peng W. A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing. *J Manuf Syst*. 2021;61:757-772.
36. Zhao CY, Huang XZ, Liu HZ, Gao TH, Shi JS. A novel bootstrap ensemble learning convolutional simple recurrent unit method for remaining useful life interval prediction of turbofan engines. *Meas Sci Technol*. 2022;33.
37. Le Cun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1989;1:541-551.
38. Chen HG, He XH, Yang H, Wu YY, Qing LB, Sheriff RE. Self-supervised cycle-consistent learning for scale-arbitrary real-world single image super-resolution. *Expert Syst Appl*. 2023;212:118657.
39. Liu M, Lu Y, Long S, Bai J, Lian W. An attention-based CNN-BiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification. *Expert Syst Appl*. 2021;186:115714.
40. Wang F, Liu X, Deng G, Yu X, Li H, Han Q. Remaining life prediction method for rolling bearing based on the long short-term memory network. *Neural Processing Lett*. 2019;50:2437-2454.
41. Shan L, Liu Y, Tang M, Bai X. CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. *J Pet Sci Eng*. 2021;205:108838.
42. Huang CG, Huang HZ, Li YF, Peng WW, Fault prognosis using deep convolutional neural network and bootstrap-based method. 2020 IEEE 18th International Conference on Industrial Informatics (INDIN). IEEE. 1(2020):742-749..

43. Li K, Wang R, Lei H, Zhang T, Liu Y, Zhang X. Interval prediction of solar power using an Improved Bootstrap method. *Sol Energy*. 2018;159:97-112.
44. Wang B, Lei YG, Li NP, Li NB. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans Reliab*. 2020;69:401-412.
45. Wang P, Long ZQ, Wang G. A hybrid prognostics approach for estimating remaining useful life of wind turbine bearings. *Energy Rep*. 2020;6:173-182.
46. Li NP, Lei YG, Lin J, Ding SX. An improved exponential model for predicting remaining useful life of rolling element bearings. *IEEE Trans Ind Electron*. 2015;62:7762-7773.

How to cite this article: Guo J, Wang J, Wang Z, et al. A CNN-BiLSTM-Bootstrap integrated method for remaining useful life prediction of rolling bearings. *Qual Reliab Eng Int*. 2023;39:1796–1813. https://doi.org/10.1002/qre.3314

AUTHOR BIOGRAPHIES

Junyu Guo is currently a lecturer in the School of Mechanical Engineering at Southwest Petroleum University. He received his PhD degree in Mechanical Engineering from the University of Electronic Science and Technology of China. He has published over 20 peer-reviewed papers in international journals and conferences. His research interests include reliability modeling and analysis, residual life prediction of mechanical equipment, and intelligent fault diagnosis.

Jiang Wang is currently a postgraduate in the School of Mechanical Engineering at Southwest Petroleum University. He received the BE degree from Chongqing University of Science and Technology. His research interest covers reliability modeling and analysis, intelligent fault diagnosis of mechanical equipment.

Zhiyuan Wang received the BE degree in Vehicle Engineering from Sichuan University of Science & Engineering, Yibin, China, in 2022. He is currently a postgraduate in the School of Mechanical Engineering at Southwest Petroleum University, Chengdu, China. His research interests include remaining useful life prediction of rotating machinery and uncertainty quantification.

Yu Gong is currently an undergraduate student in the School of Mechanical Engineering at Southwest Petroleum University. His research interests include remaining useful life prediction and intelligent fault diagnosis of rotating machinery.

Jinglang Qi is currently an undergraduate student in the School of Mechanical Engineering at Southwest Petroleum University. His research interests include remaining useful life prediction and intelligent fault diagnosis of rotating machinery.

Guoyang Wang is currently an undergraduate student in the School of Computer Science at Southwest Petroleum University. His research interests include deep learning method and fault feature extraction.

Changping Tang received the BE degree in Electronic Information Engineering from Sichuan Normal University, Chengdu, China, in 2008. He is currently working in the R&D department of Sichuan Honghua Petroleum Equipment Co., Ltd. His research interests include drilling pump electrical control system and algorithm design.