Hindawi
International Journal of Intelligent Systems
Volume 2023, Article ID 6665534, 15 pages
https://doi.org/10.1155/2023/6665534

Research Article

A Hybrid Deep Learning Prediction Method of Remaining Useful Life for Rolling Bearings Using Multiscale Stacking Deep Residual Shrinkage Network

Xudong Song,[1] **Qi Zhang**,[2] **Rui Sun**,[3] **Rui Tian**,[2] **Jialiang Sun**,[2] **Changxian Li**,[4] **and Yunxian Cui**[5]

[1]*Department of Computer and Communication Engineering, Dalian Jiaotong University, Dalian 116000, China*
[2]*Department of Software, Dalian Jiaotong University, Dalian 116000, China*
[3]*Department of Stirling, Chengdu University, Chengdu 610000, China*
[4]*Department of Automation and Electrical Engineering, Dalian Jiaotong University, Dalian 116000, China*
[5]*Department of Mechanical Engineering, Dalian Jiaotong University, Dalian 116000, China*

Correspondence should be addressed to Qi Zhang; zqdeyx1218@163.com

Received 19 July 2023; Revised 28 September 2023; Accepted 6 November 2023; Published 17 November 2023

Academic Editor: Vasudevan Rajamohan

The vibration signal is easily interfered by noise due to the influence of environment and other factors, which can lead to the poor adaptability, low accuracy of remaining useful life (RUL) prediction, and other problems. To solve this problem, this paper proposes a novel RUL prediction method, which is based on multiscale stacking deep residual shrinkage network (MSDRSN). MSDRSN combines the ability of stacking in improving prediction accuracy and the advantages of deep residual shrinkage network (DRSN) in denoising. First, cumulative sum (CUSUM) from statistics is used to divide the full life cycle of the rolling bearings and discover the points of failure. Second, stacking is used for feature learning on the raw data, multiple convolutional kernels of different scales are selected as base-learners, and fully connected neural networks are selected as meta-learners for feature fusion and learning. Then, DRSN is used to do prediction, and the acquired results are fitted with Savitzky–Golay (SG) smoothing. Finally, the effectiveness of the proposed method is proved by the IEEE PHM 2012 data challenge dataset. Compared with the multiscale convolutional neural network with fully connected layer (MSCNN-FC) and the bidirectional long short-term memory (BiLSTM) for RUL prediction under the noise. Using the proposed method, the mean absolute error (MSE) of the best result is 0.002 and the mean square error (MSE) is 0.014; meanwhile, the coefficient of determination (R^2) of the best prediction result can reach 97.6%. It is also compared with other machine learning methods, and all the results prove the accuracy and effectiveness of the proposed method for RUL prediction applications.

1. Introduction

In the field of prognostics and health management (PHM), the accurate prediction of remaining useful life has always been a key and extremely challenging problem [1, 2]. Rolling bearing is one of the most common and the crucial parts of modern machinery, and almost all kinds of mechanical equipment needs it [3]. However, long-running and repetitive loads can cause wear and damage to bearings, which can lead to noise. Rolling bearings in normal operating conditions typically produce low levels of noise, but once the bearing is damaged or worn, the noise level increases significantly. Noise, which is a common indicator of rolling bearing failure, can provide engineers with critical information about the extent of bearing wear and damage [4]. Once a rolling bearing fails, it can affect the operation of the entire mechanical system and can cause serious consequences for the equipment and the enterprise [5].

Therefore, in order to avoid the premature failure of rolling bearings, the equipment will cause accidents and

even cause a huge economic loss to the company. We need to make an accurate prediction of the remaining useful life of the rolling bearing so that we can find the failure in advance and replace it and solve the problem in time. It can not only accurately reflect its health condition in operation but also provide an accurate theoretical basis for the development of the health management plan for the equipment [6].

In general, RUL prediction methods can be divided into three categories: methods based on physical model, methods based on data-driven models, and hybrid methods [7–19]. The method based on model-driven uses physical and mathematical models for modeling and then estimates the parameters of the model from the monitoring data to predict the degradation trend of rolling bearings over their full life cycle. For example, Ding et al. [20] proposed a method to extract time-domain features such as RMS and kurtosis from the vibration signals of rolling bearings, which were used to evaluate them by the proportional hazard model, and achieved a better result. Wang et al. [21] proposed a method to get the covariates in the Weibull proportional hazard model by KPCA, which achieved high accuracy in RUL prediction. However, this often requires numerous kinds of prior knowledge, resulting in difficulty in accurately establishing a degradation model under the complex system structure and working condition [22].

With the rapid development of sensor technology, computer science, and artificial intelligence theory, data-driven methods have become more and more attractive in prognostics and health management (PHM). The data-driven method uses machine learning and deep learning to learn autonomously from the data and then infer the degradation process of rolling bearings; it not only can save time and labor but also can accurately predict RUL [23, 24]. Zhang et al. [25] proposed an improved CNN to predict RUL by using CNN's ability of autonomous learning. Xin and Weitang [26] proposed a bearing remaining life prediction method based on multiscale convolutional neural network.

Although CNN-based deep learning algorithms have achieved a large number of excellent results in the field of bearing RUL prediction, most of these methods have accomplished performance validation on laboratory datasets. However, it is difficult to capture bearing vibration signals with high signal-to-noise ratio in industrial production because of various noise sources in the environment, which makes many RUL algorithms in the industrial field have serious problems of accuracy decline or even failure. Therefore, in order to realize the practical industrial production requirements and enable the methods to complete the life prediction task in noisy environments, many researchers have focused their research on improving the robustness of the model. They have studied extensively on fault diagnosis in noisy environments. For example, Zhicheng et al. [27] used empirical wavelet transform to reconstruct the signals, then used minimum entropy deconvolution CNN to reduce the noise of the signals after composition, and achieved better results. Zou et al. [28] automatically extracted features from the background noise with a 1DCNN by performing structure optimization. Su et al. [29] designed a class of hierarchical branching CNN structures and built a basic convolutional block with strong robustness by stacking one-dimensional small convolutional kernels, which improved the accuracy. Although there are a large number of research results on fault diagnosis in noisy environments by researchers, there are still relatively few reports on life prediction in noisy environments.

Inaccurate prediction results are obtained because the CNN-based prediction model usually has the problems of gradient disappearance and gradient explosion. Therefore, He et al. [30] proposed a residual network, which brought the shortcut connection into the network to improve the linear transformation ability, and it avoids the problems of gradient explosion and gradient disappearance, thus realizing the network to stack to a deeper layer. The proposal of residual network made it possible to have a deeper network, and then appeared the deep residual network, which can reduce the number of model parameters and shorten the time of model training by increasing the setting of residual connections. Yu et al. [31] proposed a ResNet model constructed by extracting features of 1D vibration signals.

When some data containing noise and complex data are used for feature learning, the results are often not very satisfactory. Zhao et al. [32] proposed DRSN, which is an improved network of DRN, brought soft-thresholding into DRSN as a nonlinear transformation layer, and achieved autonomous learning of thresholds by adding an attention mechanism, and thus it can extract degrading information effectively.

Although the hazard methods can make full use of the advantages of both methods, the process is more complicated; therefore, this kind of method has been rarely reported. Aiming at the problem of noisy signal data in rolling bearing remaining useful life prediction, this paper focuses on the ability of deep residual systolic network to handle noise-containing data and improve it.

The novelty in this paper can be summarized as follows:

(1) The deep residual shrinkage network is improved by bringing the idea of stacking integrated learning, which learns more features of the dataset by two layers of learners, so that the results are less partial and the prediction ability of the model is improved.

(2) Traditional DRSN models usually have a single scale of convolution. This paper proposes to use multiple scales of convolution kernels as the base-learner of DRSN, which can learn more features.

(3) Based on the above, this paper proposes a model based on the MSDRSN prediction method, and the experiment results show that the method can be accurate in RUL prediction under noisy data and complex data.

The remainder of the paper is organized as follows. In Related Theory, the related techniques are introduced, and the basics of the methods used in this paper are presented, such as the basic deep residual shrinkage network, stacking integrated learning, and SG smoothing algorithm. By combining these methods, the MSDRSN method is

constructed. In Materials and Methods, the detailed prediction process of MSDRSN is given, which includes the data preprocessing stage and the training process of the model. In Experimental Verification, the dataset used in this paper is introduced, the parameters of the experiments are determined, and the results obtained are discussed. Finally, this paper concludes the research.

2. Related Theory

2.1. DRSN. DRSN is a modified model of deep residual network, which consists of three parts: deep residual network, attention mechanism, and soft thresholding function, and is usually used to enhance deep neural networks so that more useful features can be extracted from noise-laden data and redundant information is eliminated better [33].

Compared with traditional convolutional neural networks, there are two major problems: first, when the network layers are deeper, it will be prone to gradient disappearance and explosion; second, the sample data are less adaptable under the strong noise environment. By using DRSN, it can improve the above two problems [34]. The main structure of DRSN is shown in Figure 1.

Residual building unit (RBU) is the basic component of ResNet. In Figure 1, where rectangle represents the feature map, C is channel, W is width, and 1 is height. A deep residual shrinkage network can contain two Batch Normalizations (BN), two Rectifier Linear Unit activation functions (ReLU), two Convolutional layers (Conv), a Identity shortcut, a Global Average Pooling (GAP) and some Fully Connected Layers (FC).

2.2. Stacking. Stacking is the process of stacking multiple models on the same layer to get the final prediction result. In the stacking method, there are two stages of models. The first stage model is a model based on the original training set, which is called the base-learner (level-0 model), and can choose multiple base models for training. The second stage model is the model in which the predictions of the base model on the original training set are used as the training set, and the predictions of the base model on the original test set are used as the test set, called the meta-learner (level-1 model) [35].

Stacking is attracting attention because the base-learner can get the prediction value with difference and more accuracy for the original data and then relearn it by using the new feature number, so the models of multiple base-learner in the integrating learning learn from each other to make the prediction result more accurate [35].

The stacking framework is shown in Figure 2.

2.3. Savitzky–Golay Algorithm. SG algorithm, published early by Savitzky and Golay in 1964, is widely used for data smoothing and denoising. It is an important method in the field of signal processing [36]. In the field of rolling bearing remaining useful life prediction, the test set will output a RUL prediction curve by the model. However, because the training of the model is difficult to achieve complete accuracy, some points in the predicted RUL curve may be predicted incorrectly, with sudden increases or decreases.

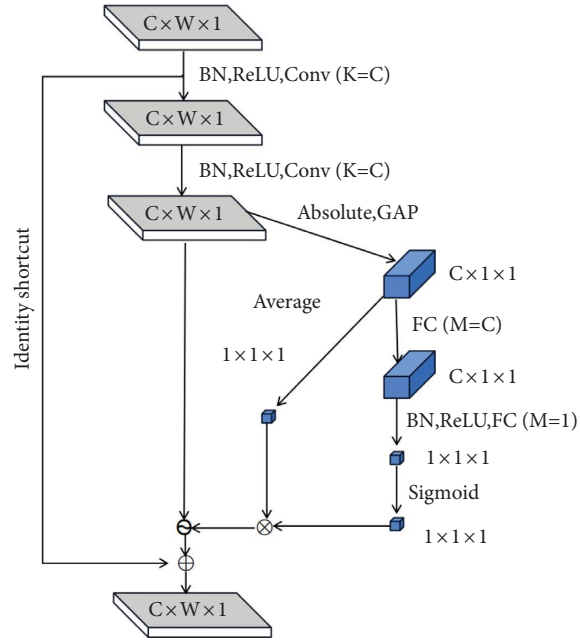

Figure 1: Model of DRSN.

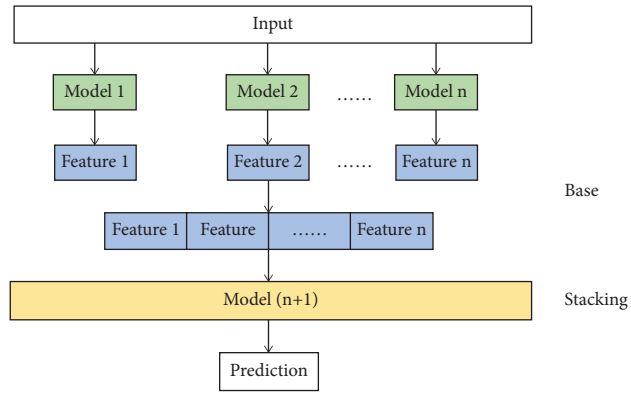

Figure 2: Framework of stacking.

Here, after smoothing the curve, these points can be modified with a local average trend, thereby having the effect of enhancing the prediction accuracy.

The signal after smoothing by the SG filter is shown in Figure 3.

The blue line is the original signal, and the red line is the signal after smoothing by the SG filter.

3. Materials and Methods

For the RUL prediction problem, this paper uses CUSUM to find degradation points. In addition, to improve model performance in learning degenerate features, this paper proposes a multiscale deep residual shrinkage network which combines the idea of stacking integrated learning. Applying integrated learning to DRSN enhances the ability to extract useful information. Using the residual connections of the residual network not only solves the gradient vanishing problem of the deep network but also improves the accuracy of the prediction results.

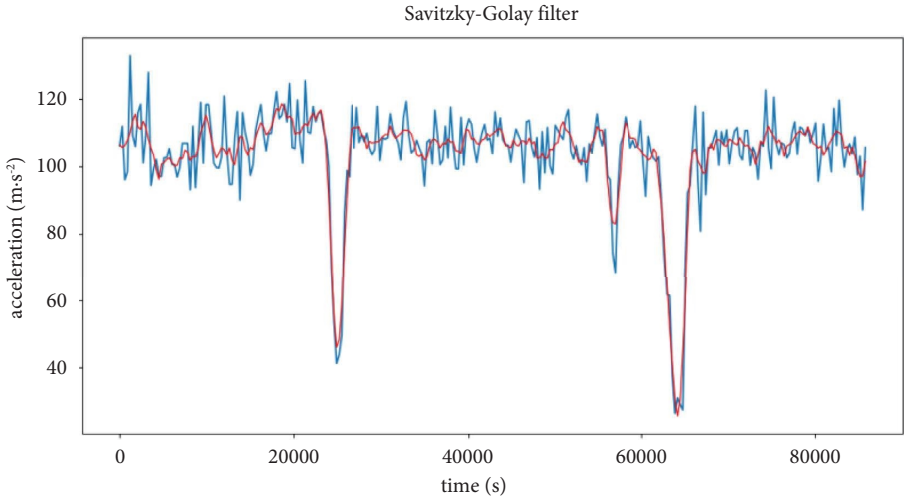

Figure 3: Smoothing comparison chart.

The complete RUL prediction process is shown in Figure 4.

3.1. CUSUM. CUSUM is a sequential analysis method that was first proposed by E. S. Page of Cambridge University, UK, in 1954. Its basic idea is to accumulate the small offset between the sample data and the target data during the process by accumulating the sample data, which plays the role of amplifying and enhancing the sensitivity of the small offset in the detection process, so as to detect the anomaly of the data, which is called CUSUM variable point detection [37, 38]. When applied to the bearing vibration signal, it can be used to find the variable point, which is the fault point of the rolling bearing.

The RUL of a rolling bearing is considered as a monotonically decreasing line. The health starts at 1. From the start of sampling until degradation decreases uniformly, it continues until the rolling bearing does not work, and at that point the health is 0.

In this paper, the marking method is different from the previous method. This paper divides the life cycle of rolling bearing into stable stage and declining stage. Generally, the rolling bearing works stably during the stable stage and the signal features basically do not change, but when it enters the declining stage, the signal features will change drastically until the rolling bearing breaks down. The advantage of this method is to keep the feature labels unchanged in the stable stage, and the labels start to decline gradually from the declining stage. It avoids the influence of the stable stage on the model training and improves the accuracy of the training results. By dividing the labels of the life cycle, the fault point of a rolling bearing can be found quickly, and the prediction result of RUL can be improved effectively.

The CUSUM algorithm is calculated in three steps, which are shown as follows.

Step 1: calculate the mean value of the series.

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}. \tag{1}$$

Step 2: calculate the cumulative sum.

$$T_i = \sum_{i=1}^{n} (x_i - \overline{x}). \tag{2}$$

Step 3: take the maximum value in the cumulative sum, and the corresponding horizontal coordinate is the fault point.

$$\left|T_{\max}\right| = \max\left(\left|T_i\right|\right). \tag{3}$$

This layer accepts the raw rolling bearing vibration signal and takes the absolute mean value of the vibration signal and then it is used for life division and labeling by CUSUM.

3.2. MSDRSN. The network structure of DRSN mainly consists of convolutional layer, residual shrinkage module, and pooling layer. MSDRSN is a new model that combines the DRSN model into stacking integrated learning and by using different scale DRSN as the base-learner and fully connected networks as the meta-learner. In RUL prediction, three residual shrinkage layers are stacked together as part of the model structure. Multiple residual shrinkage layers can better investigate the mapping connection between input and output and solve the problem of gradient disappearance. However, the full life data sequence of the bearings will contain some noise, and the residual network can only reduce the noise to a certain extent; its training effect is ordinary.

Therefore, this paper proposes an MSDRSN by considering the ability of stacking to learn features. It can not only solve the problem of gradient disappearance due to network in deepening but also enhance the training ability of the model and improve the accuracy of training, which are all contributions to rolling bearing RUL prediction.

The DRSN model construction process is shown as follows.

Step 1: convolutional layer—feature preextraction.

The calculation principle of the one-dimensional convolution layer is as follows:

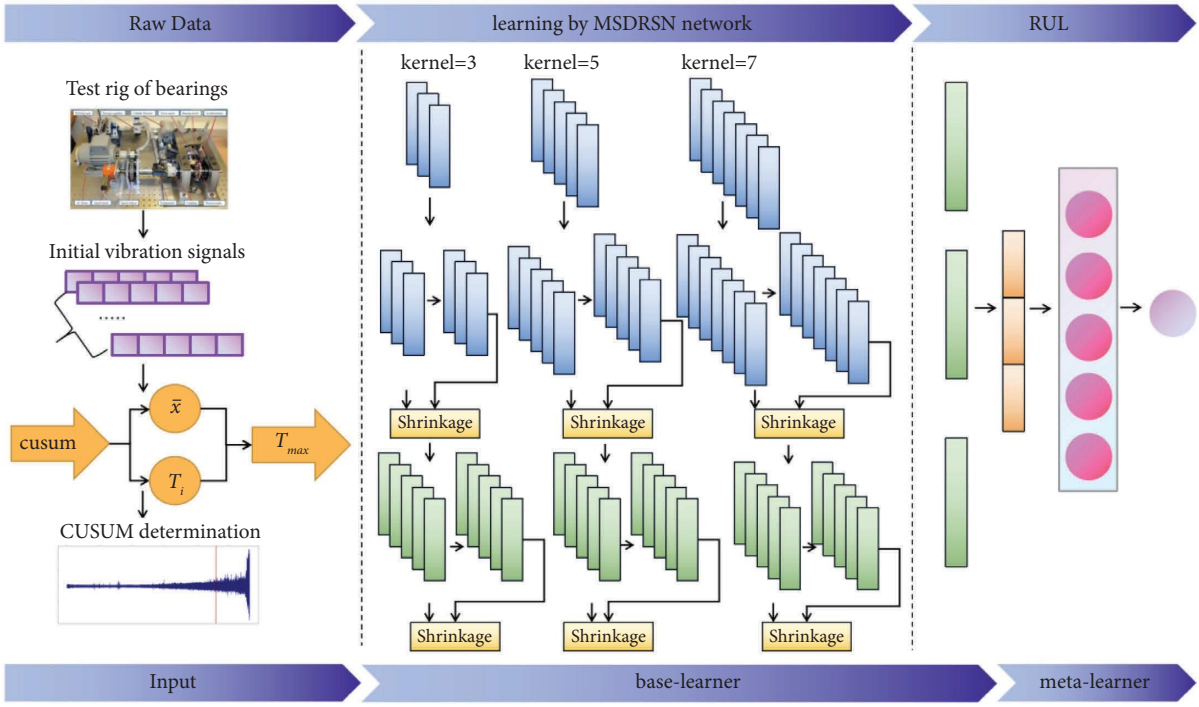

Figure 4: RUL prediction process based on MSDRSN.

$$y_i^k = f_{\text{cov}}(w_k \otimes x_i + b_k), \tag{4}$$

where x_i is the input, w_k is the weight of the k^{th} convolution kernel, $\otimes$ is the dot product operation, b_k is the bias of the k^{th} convolution kernel, f_{cov} is the activation function, and y_i^k is the output vector of the k^{th} convolution kernel. At this layer, the vibration signal is firstly processed by using a one-dimensional convolutional kernel to extract the shallow features of the signal, in order to provide a basis for the deep feature extraction in the next step and, at the same time, set the padding in this layer to avoid the loss of boundary features. This layer completes feature preextraction.

Step 2: residual shrinkage module—feature extraction and denoising.

Given that the input is x_i and the output is x_j, the soft thresholding formula can be shown:

$$x_t = \begin{cases} x_i - t \ (x_i > t), \\ 0 \ (-t \le x \le t), \\ x_i + t \ (x_i < -t), \end{cases} \tag{5}$$
$$x_j = x_i + x_t,$$

where x_t represents the output feature and t is the soft threshold. Finally, the sum of the soft thresholded feature map and the input feature map is the output feature map, and this step is the core of the identity function in the residual structure. In this layer, the network automatically extracts the features of the previous layer's data and automatically learns the threshold for denoising. After that, the data are transmitted to the next layer of the network.

Step 3: adaptive pooling layer—dimensionality reduction.

$$P_I^{l+1}(j) = \text{avg}_{(j-1)W+1 \le t \le jW} \left\{ q_i^l(t) \right\}, \tag{6}$$

where W is the width of the pooling domain, $q_i^l(t)$ is the t^{th} neuron of the i^{th} feature vector in layer l, and $P_I^{l+1}(j)$ is the value corresponding to the neuron in the $l+1$ layer. The network at this layer accepts the information output from the previous layer, automatically downsamples the data according to the set output size, and outputs the data to the next layer.

Step 4: splicing and fully connected layers—feature aggregation and relearning.

$$f(x) = W \times x + b, \tag{7}$$

where W is the weight of the fully connected layer, x is the input, and b is the bias. In this layer, it is similar to the meta-learner in the stacking. After the features are spliced in this layer, they are sent to the fully connected layer for feature training as the sample information, and finally the results are output.

Step 5: loss function.

We choose the loss function mean square error (MSE), which is most commonly used to solve the regression problem; the loss function is as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{8}$$

MSDRSN plays the role of feature extraction and initial output of RUL prediction results. The model accepts the data from the preprocessing of the previous layer, then does

feature self-extraction and output of lifetime, and puts the output results into the next layer for curve smoothing.

3.3. SG Filter. Because of the noise that inevitably appears in the training results, the SG filter was used to reduce the noise and smooth the curve to increase the accuracy and curve fit of the prediction results.

The SG filter is calculated in two steps, which are shown as follows.

Step 1: Let x[i] ($i = -\mathrm{m}, \ldots, 0, \ldots, \mathrm{m}$) be a set of consecutive integer values in the moving window; then, the width of the moving window is $M = 2\mathrm{m} + 1$. Using n^{th} degree polynomials ($n \le M$) within the filter window, the data are fitted locally using least squares.

$$y(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n, \quad (9)$$

where $y(x)$ are the output data after fitting; x are the data to be fitted; and a is the parameter to be solved. The curve is best fitted when the residuals of the n^{th} degree polynomial are zero.

Step 2: the procedure for finding the residual for the r^{th} point is as follows:

$$\begin{aligned} E &= \sum_{i=-m}^{m} [f(i) - x(i)]^2 = \left[\sum_{k=0}^{n} a_k i^k - x(i)\right]^2, \\ \frac{\partial E}{\partial a_r} &= \frac{\partial}{\partial a_r}\left\{\sum_{i=-m}^{m}\left[\sum_{k=0}^{n} a_k i^k - x(i)\right]^2\right\} = 0, \\ \frac{\partial E}{\partial a_r} &= 2\sum_{i=-m}^{m} i^r\left[\sum_{k=0}^{n} a_k i^k - x(i)\right] \\ &= \sum_{k=0}^{n} a_k \sum_{i=-m}^{m} i^{k+r} - \sum_{r=-m}^{m} x(i) \times i^r = 0, \end{aligned} \quad (10)$$

where $r = 0, 1, \ldots, n$, m is the number of single-sided points to be fitted, and $x(i)$ are the data to be fitted; then, we have $\sum_{r=-m}^{m} x(i) \times i^r$, carrying this coefficient into equations. The fitted polynomial is used to find the center point estimate within that window. The estimated value of the center point of the window is obtained by constantly moving the window and repeating the operation to obtain the estimated value of the center point of any window. Using SG filter to process the output data of MSDRSN, the noisy data can be effectively filtered out, so that the output prediction value can be more closely fitted to the real value, estimation accuracy can be improved, and model robustness can be enhanced.

3.4. Model Algorithms. The algorithm for the construction of the stacking-based prediction network process is shown in Algorithm 1.

4. Experimental Verification

To demonstrate the validity of the proposed model, this paper uses the rolling bearing accelerated degradation dataset published in the IEEE 2012 PHM Data Challenge [39]. The proposed RUL prediction method's feasibility and effectiveness are analyzed in detail by experiments and compared with other RUL prediction methods.

4.1. Data Description. The PHM 2012 challenge dataset was provided by the PRONOSTIA platform of the FEMTO-ST Institute. It is shown in Figure 5.

The PRONOSTIA platform consists of three main parts: the rotating part, the degradation part, and the measurement part [40]. The data provided by PRONOSTIA describe the degradation of ball bearings throughout their service life (until complete failure), and each degraded bearing contains almost all types of defects (inner and outer rings, balls, and cage). In the experiment, a radial load force is applied to accelerate the degradation of the bearings. The vibration signals in the X and Y directions are acquired with sampling frequency of 25600 Hz, 2560 data points are recorded at 0.1 s per sampling, and the recording interval is 10 s. When the vibration level of the measured data exceeds a certain threshold, the test is stopped [41]. The basic characteristics of bearings are listed in Table 1, and Table 2 gives a detailed description of the dataset.

Different working conditions will have some influence on the training of the model. This paper selects the datasets under three conditions for experiments. The reason for the choice is that the three datasets had long decline periods and more information could be learned. In total, three groups of tests were carried out, and the detailed arrangement is shown in Table 3 [41].

4.2. Evaluation Metrics. In this paper, three metrics are used to evaluate performance: mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2). The three evaluation metrics are defined as follows:

$$\begin{aligned} \mathrm{MSE} &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{RUL}_{\mathrm{true}} - \mathrm{RUL}_{\mathrm{predict}}\right)^2, \\ \mathrm{MAE} &= \frac{1}{n}\sum_{i=1}^{n}\left|\mathrm{RUL}_{\mathrm{true}} - \mathrm{RUL}_{\mathrm{predict}}\right|, \\ R^2 &= 1 - \frac{\sum_i\left(\mathrm{RUL}_{\mathrm{true}} - \mathrm{RUL}_{\mathrm{predict}}\right)^2}{\sum_i\left(\overline{\mathrm{RUL}_{\mathrm{true}}} - \mathrm{RUL}_{\mathrm{predict}}\right)^2}, \end{aligned} \quad (11)$$

where RULpredict and RULtrue represent the predicted RUL and the actual RUL, respectively, and n is the length of the testing data. The smaller the MSE and MAE, the better the prediction performance. The closer R^2 is to 1, the better the prediction performance is.

4.3. Data Preprocessing. In this paper, CUSUM is used to detect the change point of the operating state of the rolling bearing vibration signal and to divide the bearing life cycle.

Input: Training set $D = \{x_i, y_i\}_{i=1}^{m}$
Output: Predicted value after integration H
Process:
Step 1: Data preprocessing
for $i = 1$ to m do
abs (mean(x_i))
end for
p = do CUSUM on x_i
for $i = 1$ to m do
if $i < p$ set $y_i = 0$
else set $y_i = m - \mathrm{i}/m - p$
end for
normalization(x_i)
Step 2: Training base-learner
for $t = 1$ to T do
learn h_t based on D
end for
Step 3: Feature aggregation
for $i = 1$ to m do
$D_h = \{x_i', y_i\}$, where $\mathrm{x}_\mathrm{i}' = \{h_1(x_i), \ldots, h_T(x_i)\}$
end for
Step 4: Training meta-learner
learn H based on D_h
Step 5: Smoothing the curve
smooth H
return H

Algorithm 1: Stacking-based algorithm for prediction network process construction.

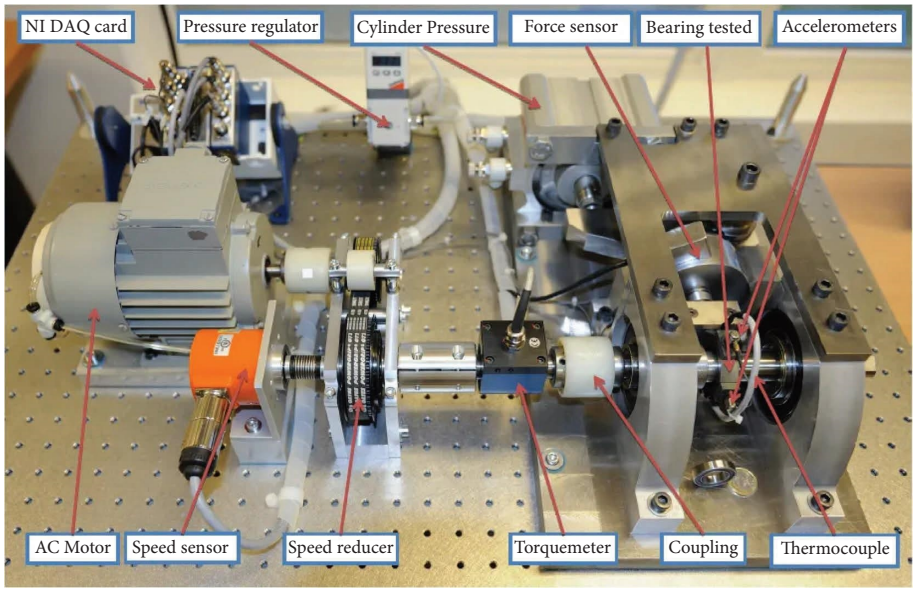

Figure 5: The PRONOSTIA platform.

Table 1: Characteristics of the bearings.

Parameter	Value
Diameter of rolling elements	3.5 mm
Number of rolling elements	13
Diameter of the outer race	29.1 mm
Diameter of inner race	22.1 mm
Bearing mean diameter	25.6 mm
Contact angle	0°

Table 2: Description of the dataset.

	Condition 1	Condition 2	Condition 3
Load/N	4000	4200	5000
Speed/(*r*·min-1)	1800	1650	1500
Bearings	Bearing 1_1	Bearing 2_1	Bearing 3_1
	Bearing 1_2	Bearing 2_2	Bearing 3_2
	Bearing 1_3	Bearing 2_3	Bearing 3_3
	Bearing 1_4	Bearing 2_4	Bearing 3_4
	Bearing 1_5	Bearing 2_5	Bearing 3_5
	Bearing 1_6	Bearing 2_6	Bearing 3_6
	Bearing 1_7	Bearing 2_7	Bearing 3_7

Table 3: RUL prediction tasks.

Tasks	Training bearings	Test bearings
A	Bearing 1_1, Bearing 1_2	Bearing 1_3
B	Bearing 2_1, Bearing 2_2	Bearing 2_3
C	Bearing 3_1, Bearing 3_2	Bearing 3_3

Taking Bearing 1_1 as an example, the steps of sample labeling are as follows:

Step 1: Rolling bearing vibration signal actually measures the acceleration of the rolling bearing in a certain direction, it is usually a vibration signal in the horizontal direction, it represents not the size but the direction of bearing vibration, and the direction of vibration of the bearing is not of concern in the signal analysis; therefore, the signal should be first to take the absolute value. Due to the 2560 vibration signals sampled at a single moment of the bearing, this paper uses the mean value to replace the vibration signal at a certain moment. After processing the data, the signal curve of the absolute mean vibration signal is shown in Figure 6.

Step 2: After the signal processing is completed, the signal curve is smoothed by SG for denoising, and then CUSUM is used to find the fault point. The former curve is regarded as the stable stage, and the RUL is 1. The latter curve is regarded as the declining stage, and the RUL is marked as decreasing from 1 to 0.

After all the data RUL are labeled, because the different bearing vibration signal values have different intervals, the bearing vibration signal samples also need to be normalized.

The degradation points of each bearing are given in Table 4, which demonstrates the validity of the method.

Step 3: To verify the superiority of MSDRSN model in handling noisy data, according to the SNR for noise

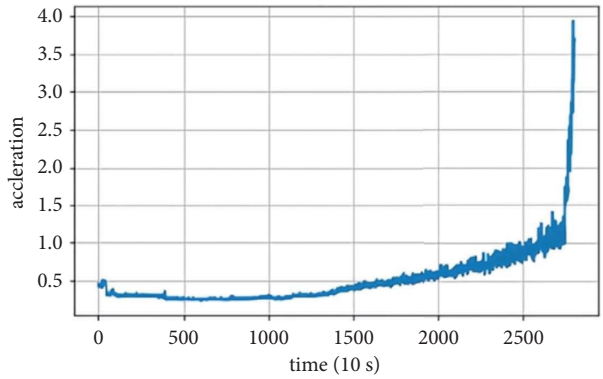

Figure 6: The absolute mean vibration signal.

addition process, this paper adds 2 db, 4 dB, 6 db, and 8 db of noise in order. SNR formula is shown as follows:

$$\mathrm{SNR}\,(\mathrm{dB}) = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right). \tag{12}$$

4.4. Model Building Experiments. The structural parameters of MSDRSN are shown in Table 5. The hyperparameters need to be preset, such as the number of residual blocks and the size of the learning rate. The hyperparameters are shown in Table 6.

There are some other parameters in this experiment such as the validity of CUSUM and the validity of SG smoothing. Meanwhile, the number of residual blocks and the learning rate size are also determined by the experiment.

(1) Effectiveness of CUSUM: a comparison test based on bearing 1_3, bearing 2_3, and bearing 3_3 at 8 db noise.

The SVR machine learning model is used to test the model, and the kernel function is RBF. The results of the comparison experiments are shown in Figure 7.

From the experiments, it can be shown that the MSE of bearing 1_3, bearing 2_3, and bearing 3_3 in which CUSUM is applied is lower than that of those in which CUSUM is not applied, and thus it can be concluded that CUSUM significantly improves the prediction capability of the model.

(2) Validity of SG smoothing: a comparison test based on bearing 1_3 at 2 db, 4 db, 6 db, and 8 db noise.

The number of network layers in this experiment is 2; the pooling layer size is 80, and the comparison item is whether SG smoothing is applied or not. The results of the comparison experiment are shown in Figure 8.

From the experiments, it can be concluded that the training results are better when the model is SG smoothed. The training results are worse when SG smoothing is not set, so the training ability of the model is greatly enhanced by SG smoothing.

Table 4: The degradation points of rolling bearing.

	Bearing	Degradation moment
Condition 1	Bearing 1-1	1325
	Bearing 1-2	696
Condition 2	Bearing 2-1	850
	Bearing 2-2	745
Condition 3	Bearing 3-1	458
	Bearing 3-2	1570

Table 5: Structural parameters of MSDRSN.

Module	Layers	Output size and activation
Convolution layers	Conv 1	(None, 3, 6), Activation: ReLU
	Conv 2	(None, 3, 6), Activation: ReLU
	Conv 3	(None, 3, 6), Activation: ReLU-sigmoid
DRSN layers	Residual building unit 1	(None, 6, 3), Activation: ReLU-sigmoid
	Residual building unit 2	(None, 6, 3), Activation: ReLU-sigmoid
	Residual building unit 3	(None, 6, 3), Activation: ReLU-sigmoid
	Pooling	(None, 6, 1), Activation: ReLU-sigmoid
	Fully connected	(None, 1), Activation: ReLU-sigmoid

Table 6: The crucial hyperparameters.

Hyperparameters	Value
Input dimensions	2560
Output dimensions	1
Residual blocks	2
Learning rate	0.000001
Batch size	128

(3) Number of residual blocks: a comparison test based on bearing 1_3 at 8 db noise.

The most influential factor for the MSDRSN network is how many residual blocks are available. The validation part of the model structure mainly compares the training effects of MSDRSN with different numbers of basic residual blocks, and it determines the optimal number of network layers for this experiment. The experiment sets the convolutional kernel size of convolutional layer as 3, 5, and 7; the basic residual block convolutional kernel size as 3, 5, and 7; the output size of mean pooling layer as 18; and the aggregation layer as 2 layers with the number of neurons in each layer as 18 and 1. The number of basic residual blocks of the comparison items is 1, 2, 3, 4, and 5. The results of the comparison experiments are shown in Figure 9.

(4) The size of the learning rate: a comparison test based on bearing 1_3 at 8 db noise.

The setting of the learning rate has a significant impact on the MSDRSN network. It is often that too low learning rate will lead to slow convergence of the model, while too high learning rate will lead to failure of the model to converge, so this experiment compares the optimal values of different learning rate losses and determines the optimal learning rate. The results of the comparison experiment are shown in Figure 10:

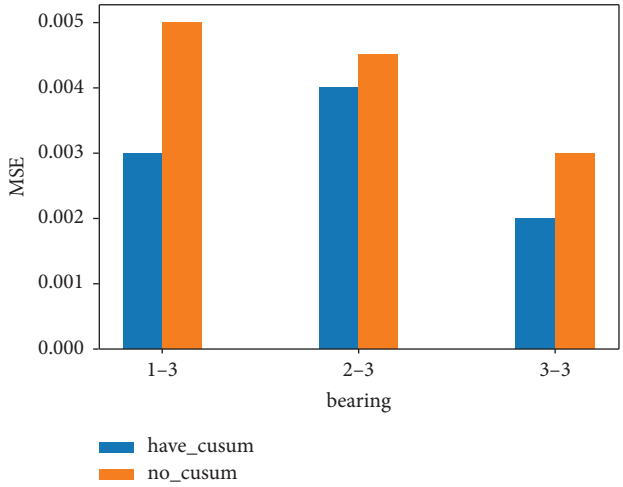

Figure 7: Comparison of CUSUM.

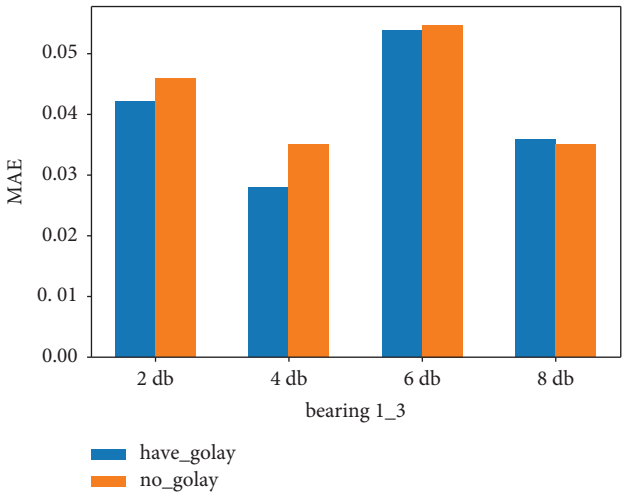

Figure 8: Comparison of Savitzky-Golay smoothing.

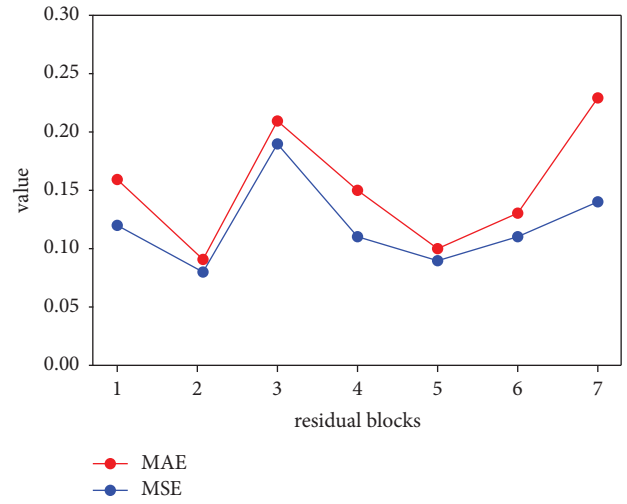

Figure 9: Comparison of the number of residual blocks.

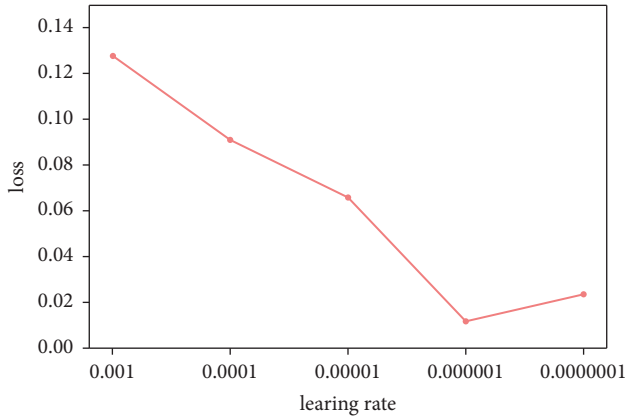

Figure 10: Comparison of learning rate.

From the experiment, it can be concluded that when the learning rate is set as 0.000001, the loss value is the smallest and the model training result is the best, so the optimal learning rate of MSDRSN network is set as 0.000001.

5. RUL Prediction Results

The programming software used for the experiments is Python 3.6, and the central processing unit (CPU) used in the workstation is Intel i7-11800H.

During the experiment, the MSDRSN method was compared with the common methods, selecting two machine learning models, which are Random forest and SVR, and two deep learning models, which are BiLSTM [42] and MSCNN-FC [43]. BiLSTM is a stack of two LSTM layers. It can effectively utilize the input forward and backward feature information. MSCNN-FC can be used for prediction problems, which is similar to their algorithm for classification problems, but the last layer is usually a fully connected layer with only one neuron.

Table 7 shows the comparison table of evaluation metrics of MSDRSN method, BiLSTM, and MSCNN-FC in different noise environments.

Among them, MSDRSN has the smallest MSE and MAE, indicating that it has the smallest prediction error, and R^2 is the largest, indicating that its model has the best fitting prediction results. Besides, the prediction accuracy of tasks 2_3 is low, and the prediction results of both condition 1 and condition 3 are better than those of other methods. The feasibility and superiority of MSDRSN are demonstrated by comparing with BiLSTM and MSCNN-FC. The results of the experiments conducted under different noises also demonstrate that the MSDRSN network is more suitable for prediction tasks with noisy data. Taking bearings 1_3 as an example, the RUL prediction curves under different noises are shown in Figure 11.

It can be shown that the prediction results of the method have a similar trend to the actual RUL values, and the prediction accuracy is significantly higher than that of BiLSTM and MSCNN-FC, and thereby it verifies the effectiveness of the model in rolling bearing RUL prediction.

In order to verify the superiority of the model, this paper chooses two traditional machine learning methods to compare with it, Random forest and SVR. Taking bearings 1_3 as an example, the RUL prediction curves of the three models under different noises are shown in Figure 12.

It is obvious from the prediction results that the model has a higher prediction accuracy compared with other traditional machine learning methods, and the curves are most compatible with the actual values, and there is no obvious fluctuation trend.

In summary, the method proposed in the paper can more effectively predict the remaining life of rolling bearings. The method can more accurately capture the early degradation characteristics of the bearing. At the same time, the method makes a significant improvement in the prediction of the later bearing life.

6. Discussion of RUL Prediction

(1) The MSDRSN method convolves operations with multiple convolution kernels of different scales to achieve the effect of extracting detailed features and increasing the prediction accuracy of the model. Stacking integrated learning can observe the features of the data from multiple perspectives and learn them with the base-learner and relearn them after fusing the obtained features so as to improve the accuracy of the prediction results. Then, this paper uses the powerful noise reduction capability of DRSN, which enables it to get more useful features of the dataset, which improves the feature learning capability of the model in complex datasets and improves the feature extraction capability of the model.

(2) The model utilizes the "skip connection" structure of the residual network, which illustrates the good feature extraction performance of the deep network and avoids the problem of gradient disappearance when the deep neural network trains.

(3) Compared with the four models including BiLSTM, MSCNN-FC, SVR, and Random forest, *the*

Table 7: Comparison of evaluation metric results.

Testing bearing		MSDRSN			BiLSTM			MSCNN-FC		
		MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
1-3	2 db	**0.005**	**0.042**	**0.954**	0.102	0.243	0.418	0.047	0.186	0.537
	4 db	**0.003**	**0.028**	**0.975**	0.078	0.141	0.568	0.050	0.126	0.503
	6 db	**0.004**	**0.054**	**0.956**	0.115	0.548	0.432	0.053	0.122	0.477
	8 db	**0.004**	**0.036**	**0.962**	0.256	0.106	0.566	0.035	0.068	0.659
2-3	2 db	**0.004**	**0.012**	**0.625**	0.015	0.275	0.578	0.007	0.720	0.551
	4 db	**0.005**	**0.016**	**0.503**	0.032	0.349	0.498	0.008	0.135	0.256
	6 db	**0.003**	**0.014**	**0.653**	0.048	0.217	0.651	0.008	0.057	0.146
	8 db	**0.003**	**0.014**	**0.657**	0.041	0.136	0.557	0.009	0.060	0.093
3-3	2 db	**0.003**	**0.026**	**0.971**	0.033	0.159	0.667	0.015	0.076	0.851
	4 db	**0.004**	**0.054**	**0.964**	0.006	0.047	0.941	0.011	0.063	0.892
	6 db	**0.002**	**0.043**	**0.976**	0.046	0.189	0.537	0.025	0.144	0.744
	8 db	**0.004**	**0.054**	**0.963**	0.024	0.102	0.759	0.018	0.077	0.811

The bold values in the table are the values of the proposed method, which is convenient and clearer to compare with the values of other methods.

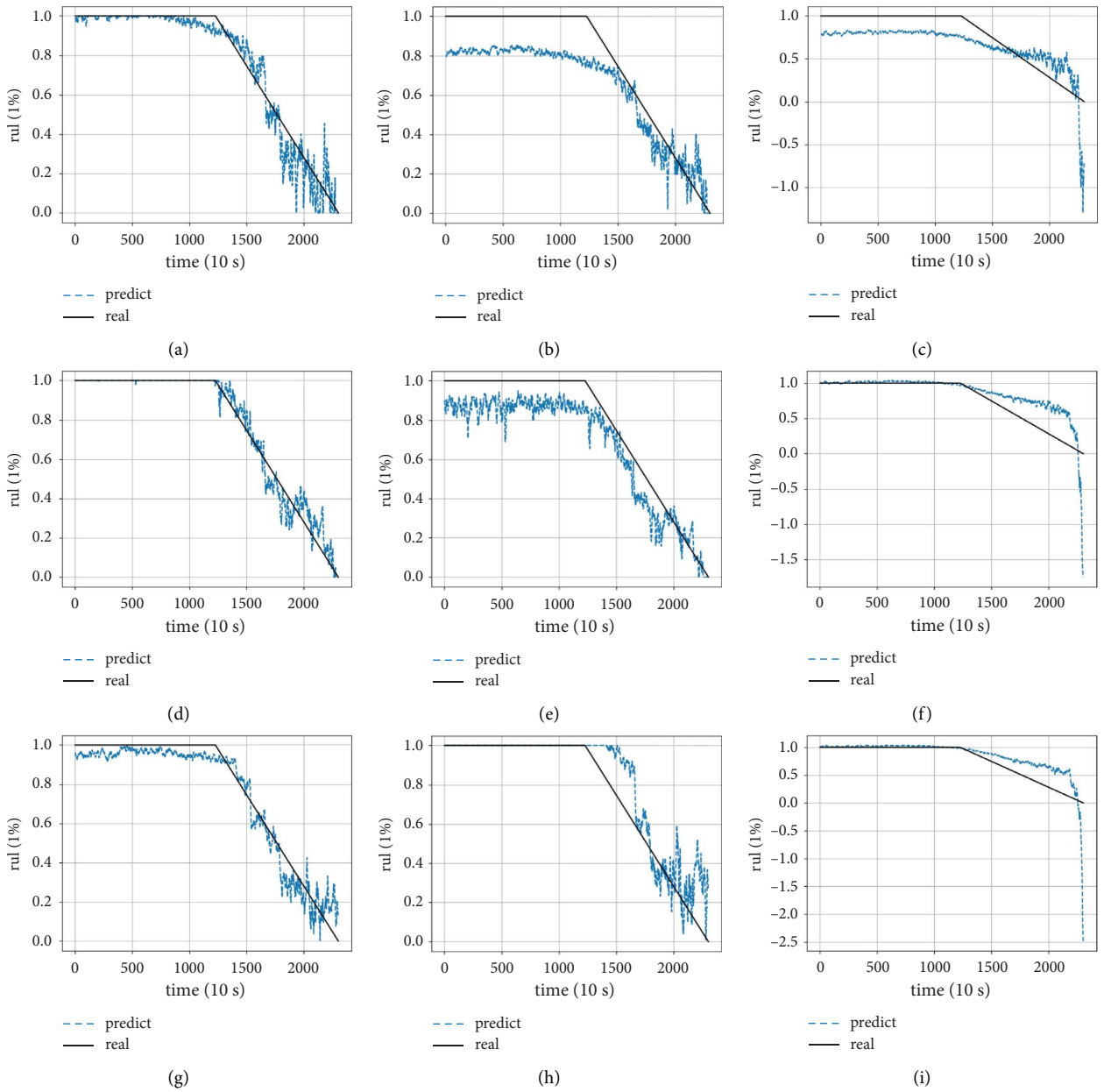

Figure 11: Continued.

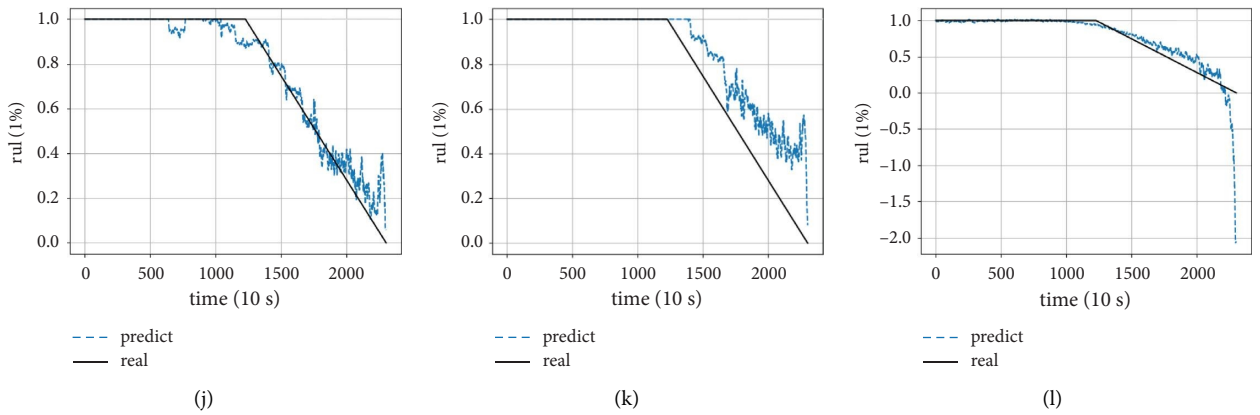

Figure 11: (a) 2 db MSDRSN. (b) 2 db BiLSTM. (c) 2 db MSCNN-FC. (d) 4 db MSDRSN. (e) 4 db BiLSTM. (f) 4 db MSCNN-FC. (g) 6 db MSDRSN. (h) 6 db BiLSTM. (i) 6 db MSCNN-FC. (j) 8 db MSDRSN. (k) 8 db BiLSTM. (l) 8 db MSCNN-FC.

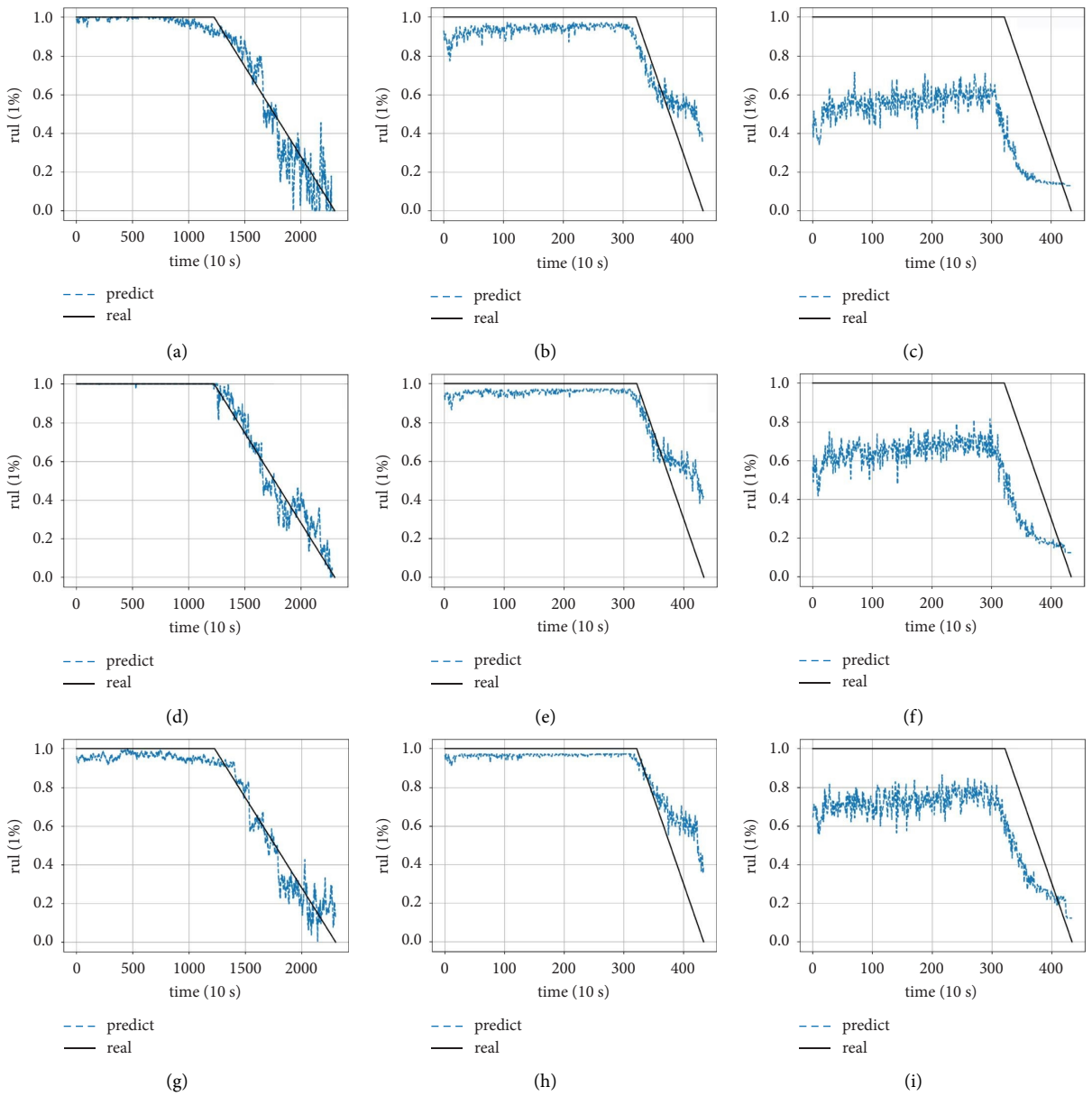

Figure 12: Continued.

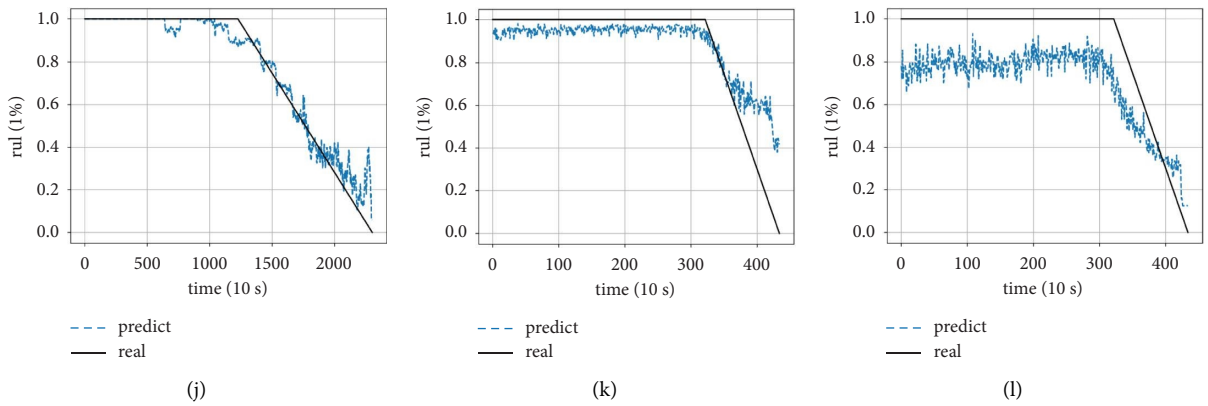

Figure 12: (a) 2 db MSDRSN. (b) 2 db Random forest. (c) 2 db SVR. (d) 4 db MSDRSN. (e) 4 db Random forest. (f) 4 db SVR. (g) 6 db MSDRSN. (h) 6 db Random forest. (i) 6 db SVR. (j) 8 db MSDRSN. (k) 8 db Random forest. (l) 8 db SVR.

MSDRSN method can get better results, and the R^2under 4 db noise in bearings 1_3 is 0.975, and the loss values MAE and MSE under 6 db noise in bearings 3_3 are 0.002 and 0.043; the prediction is almost close to the actual life. Thus, it is verified that the MSDRSN method has a high accuracy of RUL prediction.

7. Conclusions

This paper proposes a novel rolling bearing RUL prediction method using MSDRSN. In the preprocessing stage of the dataset, this paper finds the fault points in the bearing with CUSUM. In the RUL prediction network, the idea of stacking integrated learning is merged with the traditional deep residual shrinkage network to improve the prediction ability of the model with noise. We have conducted experiments to validate the method with the IEEE PHM 2012 data challenge dataset, and the results show that the proposed MSDRSN method improves the accuracy of RUL prediction. Comparing with different methods for predicting the remaining useful life, it is demonstrated that the feasibility and superiority of the method.

The future directions of the article can be improved as follows:

(1) MSDRSN has a powerful denoising capability, but the calculations take a very long time and the model parameters are too many, making it difficult to debug. In the future, people can consider using distributed technology to speed up the training of the model and create an algorithm that can automatically adjust the parameters of the model to find the optimal solution.

(2) Stacking fusion technique can effectively improve the model training results, but the selection of base-learner is a difficult problem, which requires a lot of time to select. In the future, people can consider proposing an evaluation system to predict the effect of the base-learner model fusion and reduce the difficulty of model selection.

Data Availability

Previously reported data were used to support this study and are available at https://hal.science/hal-00719503. These prior studies (and datasets) are cited at relevant places within the text as references [35].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant no. 52175379 and Liaoning Provincial Science and Technology Department under grant no. 2022JH2/101300268.

References

[1] S. Wan, X. Li, Y. Zhang, S. Liu, J. Hong, and D. Wang, "Bearing remaining useful life prediction with convolutional long short-term memory fusion networks," *Reliability Engineering & System Safety*, vol. 224, Article ID 108528, 2022.

[2] W. Yu, D. Pi, L. Xie, and Y. Luo, "Multiscale attentional residual neural network framework for remaining useful life prediction of bearings," *Measurement*, vol. 177, Article ID 109310, 2021.

[3] S. Xiang, Y. Qin, F. Liu, and K. Gryllias, "Automatic multi-differential deep learning and its application to machine remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 223, Article ID 108531, 2022.

[4] D. Chen, Y. Qin, Y. Wang, and J. Zhou, "Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction," *ISA Transactions*, vol. 114, pp. 44–56, 2021.

[5] T. Han, L. Zhang, Z. Yin, and A. C. Tan, "Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine," *Measurement*, vol. 177, Article ID 109022, 2021.

[6] K. Yu, T. R. Lin, and J. Tan, "A bearing fault and severity diagnostic technique using adaptive deep belief networks and

Dempster–Shafer theory," *Structural Health Monitoring*, vol. 19, no. 1, pp. 240–261, 2020.

[7] L. Yuan, W. Zhuojian, L. Zhe, and J. Zihan, "Research on fault prognosis methods based on data-driven: a survey," *IOP Conference Series: Materials Science and Engineering*, vol. 1043, no. 4, Article ID 042008, 2021.

[8] J. Kou, S. Le Clainche, and E. Ferrer, "Data-driven eigensolution analysis based on a spatio-temporal Koopman decomposition, with applications to high-order methods," *Journal of Computational Physics*, vol. 449, Article ID 110798, 2022.

[9] Y. Wang, M. H. Yang, H. Y. Zhang, X. Wu, and W. Hu, "Data-driven prediction method for characteristics of voltage sag based on fuzzy time series," *International Journal of Electrical Power & Energy Systems*, vol. 134, Article ID 107394, 2022.

[10] X. Sun, B. Bahmani, N. N. Vlassis, W. C. Sun, and Y. Xu, "Data-driven discovery of interpretable causal relations for deep learning material laws with uncertainty propagation," *Granular Matter*, vol. 24, pp. 1–32, 2022.

[11] S. Jha, S. Bandyopadhyay, S. Mehta et al., "Data-driven predictive electrochemical behavior of lignin-based supercapacitors via machine learning," *Energy & Fuels*, vol. 36, no. 2, pp. 1052–1062, 2021.

[12] J. Pulpeiro González, K. Ankobea-Ansah, Q. Peng, and C. M. Hall, "On the integration of physics-based and data-driven models for the prediction of gas exchange processes on a modern diesel engine," *Proceedings of the Institution of Mechanical Engineers-Part D: Journal of Automobile Engineering*, vol. 236, no. 5, pp. 857–871, 2022.

[13] H. Huang, X. Huang, W. Ding, M. Yang, D. Fan, and J. Pang, "Uncertainty optimization of pure electric vehicle interior tire/road noise comfort based on data-driven," *Mechanical Systems and Signal Processing*, vol. 165, Article ID 108300, 2022.

[14] F. Mei, X. Li, J. Zheng, H. Sha, and D. Li, "A data-driven approach to state assessment of the converter valve based on oversampling and Shapley additive explanations," *IET Generation, Transmission & Distribution*, vol. 16, no. 8, pp. 1607–1619, 2022.

[15] D. Tellez-Castro, C. Garcia-Tenorio, E. Mojica-Nava, J. Sofrony, and A. Vande Wouwer, "Data-driven predictive control of interconnected systems using the koopman operator," *Actuators*, vol. 11, no. 6, p. 151, 2022.

[16] B. Li and S. Rangarajan, "A conceptual study of transfer learning with linear models for data-driven property prediction," *Computers & Chemical Engineering*, vol. 157, Article ID 107599, 2022.

[17] Y. Fan, H. Wang, H. Gemmeke, T. Hopp, and J. Hesser, "Model-data-driven image reconstruction with neural networks for ultrasound computed tomography breast imaging," *Neurocomputing*, vol. 467, pp. 10–21, 2022.

[18] S. Z. Feng, X. Han, Z. Li, and A. Incecik, "Ensemble learning for remaining fatigue life prediction of structures with stochastic parameters: a data-driven approach," *Applied Mathematical Modelling*, vol. 101, pp. 420–431, 2022.

[19] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: a systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.

[20] F. Ding, H. Zhengjia, Z. I. Yanyang, C. H. E. N. Xuefeng, C. A. O. Hongrui, and T. A. N. Jiyong, "Reliability assessment based on equipment condition vibration feature using proportional hazards model," *Journal of Mechanical Engineering*, vol. 45, no. 12, pp. 89–94, 2009.

[21] F. Wang, X. Chen, C. Liu, D. Yan, Q. Han, and H. Li, "Reliability assessment of rolling bearing based on principal component analysis and Weibull proportional hazard model," in *Proceedings of the 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2017.

[22] Y. Qin, D. Chen, S. Xiang, and C. Zhu, "Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6438–6447, 2021.

[23] B. L. Lu, Z. H. Liu, H. L. Wei, L. Chen, H. Zhang, and X. H. Li, "A deep adversarial learning prognostics model for remaining useful life prediction of rolling bearing," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 4, pp. 329–340, 2021.

[24] X. Su, H. Liu, L. Tao, C. Lu, and M. Suo, "An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive transformer model," *Computers & Industrial Engineering*, vol. 161, Article ID 107531, 2021.

[25] J. Zhang, Y. Zou, J. Deng, and X. Zhang, "Bearing remaining life prediction based on full convolutional layer neural networks," *China Mechanical Engineering*, vol. 30, no. 18, 2019.

[26] S. Xin and S. Weitang, "Research on remaining life prediction of bearing based on multi-scale convolution neural network," *Modular Machine Tool & Automatic Manufacturing Technique*, vol. 10, pp. 168–171, 2020.

[27] Q. Zhicheng, L. Yongqiang, and L. Yingying, "Application of improved wavelet transform and minimum entropy deconvolution in railway bearing fault diagnosis," *Journal of Vibration and Shock*, vol. 40, no. 2, pp. 81–90, 2021.

[28] F. Zou, H. Zhang, S. Sang et al., "An anti-noise one-dimension convolutional neural network learning model applying on bearing fault diagnosis," *Measurement*, vol. 186, Article ID 110236, 2021.

[29] K. Su, J. Liu, and H. Xiong, "Hierarchical diagnosis of bearing faults using branch convolutional neural network considering noise interference and variable working conditions," *Knowledge-Based Systems*, vol. 230, Article ID 107386, 2021.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, June 2016.

[31] Z. Yu, H. Huang, Y. Tang, D. Wu, and Y. Bao, "Remaining useful life prediction via multi-neural network integration," in *Proceedings of the 2021 China Automation Congress (CAC)*, October 2021.

[32] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4681–4690, 2020.

[33] J. Tong, S. Tang, Y. Wu, H. Pan, and J. Zheng, "A fault diagnosis method of rolling bearing based on improved deep residual shrinkage networks," *Measurement*, vol. 206, Article ID 112282, 2023.

[34] W. Zhao and S. Hu, "End-to-End auto-encoder system for deep residual shrinkage network for AWGN channels," *Journal of Computer and Communications*, vol. 11, no. 5, pp. 161–176, 2023.

[35] S. Cui, Y. Yin, D. Wang, Z. Li, and Y. Wang, "A stacking-based ensemble learning method for earthquake casualty prediction," *Applied Soft Computing*, vol. 101, Article ID 107038, 2021.

[36] X. Yang, A. Zou, J. Cao, Y. Lai, and J. Zhang, "Systemic risk prediction based on Savitzky-Golay smoothing and temporal

convolutional networks," *Electronic Research Archive*, vol. 31, no. 5, pp. 2667–2688, 2023.

[37] G. Zhang, H. Hao, Y. Wang et al., "Optimized adaptive Savitzky-Golay filtering algorithm based on deep learning network for absorption spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 263, Article ID 120187, 2021.

[38] J. S. Mohammed and J. A. Abdulhady, "Rolling bearing fault detection based on vibration signal analysis and cumulative sum control chart," *FME Transactions*, vol. 49, no. 3, pp. 684–695, 2021.

[39] P. Nectoux, R. Gouriveau, K. Medjaher et al., "PRONOSTIA: an experimental platform for bearings accelerated degradation tests," in *Proceedings of the IEEE International Conference on Prognostics and Health Management, PHM'12*, Denver, CO, USA, June 2012.

[40] Y. Qian and R. Yan, "Remaining useful life prediction of rolling bearings using an enhanced particle filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 10, pp. 2696–2707, 2015.

[41] Y. Cao, Y. Ding, M. Jia, and R. Tian, "A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings," *Reliability Engineering & System Safety*, vol. 215, Article ID 107813, 2021.

[42] M. F. Ahammed, A. A. Molla, R. Kadir, and M. I. Kadir, "Deep bidirectional LSTM for the signal detection of universal filtered multicarrier systems," *Machine Learning with Applications*, vol. 10, Article ID 100425, 2022.

[43] C. Hu, C. Hu, and X. Si, "Remaing useful lifetime prediction method of rolling bearing based on MSCNN-LSTM," *China Measurement&Test*, vol. 46, no. 9, pp. 103–110, 2020.