Artificial Intelligence Technologies and Applications
C. Chen (Ed.)

doi:10.3233/FAIA231340

A Rolling Bearing Degradation Index Construction Method Based on Deep Residual Networks

Yuanhan HU [a], Yuna PAN [a,1], Kun XIE [a], Xianghua MA [b]
[a]*School of Railway Transportation, Shanghai Institute of Technology, Shanghai 201418, China;*
[b]*School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai 201418, China*

Abstract. The evaluation of performance degradation in rolling bearings is a crucial technology for active maintenance. As the traditional construction method of rolling bearing degradation index is heavily dependent on manual screening features, this paper proposes an end-to-end construction method of degradation index based on Deep Residual Networks (DRN). In this method, two nodes are established in the *Softmax* output layer of DRN, which respectively represent the probability of a normal state and the probability of a failure state. The training sample is taken as the normalized amplitude spectrum of the normal state. The performance degradation index of the rolling bearing is constructed by taking into account the output probability of the normal node of the sample to be evaluated. The method's effectiveness and superiority are confirmed through the use of various experimental data sets, when compared to other indexes.

Keywords. rolling bearing; degradation index; deep residual networks(DRN)

1. Introduction

Rolling bearings are an important part of ensuring the safe and stable operation of rotating equipment, and according to statistics, their damage leads to the failure of about 45%-55% of equipment [1,2]. The accidental failure of bearings, often causes sudden failure of equipment and even systems, which brings great costs and negative benefits to enterprises. Therefore, it is crucial to carry out a comprehensive health evaluation and accurately grasp its deterioration status in all aspects to ensure the safe operation of machines and equipment.

When assessing the performance degradation of rolling bearings, establishing degradation indicators is crucial. To conduct real-time monitoring of it, it is necessary to reflect one or more features of its evolution from normal to fault, and use appropriate and sensitive feature extraction techniques to effectively identify it [3].

In recent years, because deep learning technology can extract important information from a large amount of raw data, it makes it a new type of fault diagnosis method and has been applied in a large number of applications on this basis [4-8]. Zhao Jingjiao et al established a new bearing fault identification method using a

[1] Corresponding author: Yuna PAN, School of Railway Transportation, Shanghai Institute of Technology, Shanghai 201418, China; E-mail: 1697539742@qq.com

one-dimensional convolutional neural network based on residual connection [9]. L.S. HOU et al proposed a new algorithm to automatically extract features from raw signals without pre-setting parameters, and DRN network was used as a fault classification to realize the effective identification of bearing vibration data [10]. Ren Shuang et al constructed a kind of Convolutional Neural Network (CNN) based on the Inception structure and residual structure, which can effectively diagnose the bearing faults [11]. In the above study, the DRN has good feature extraction performance, which improves the recognition accuracy of the model; at the same time, the DRN can also accelerate the learning speed of large-scale networks.

Based on the above analysis, this article proposes a DRN based method for constructing degradation indicators of rolling bearings. The core idea is to set two nodes at the output end of *Softmax* in DRN to obtain probabilities representing normal and failure states. Using normalized amplitude spectra of normal states as training samples, based on the output probability of the data to be evaluated at normal nodes, a degradation index (DI) for rolling bearing performance is constructed. Finally, the effectiveness and superiority of this method were verified through the application of rolling bearing full life cycle experimental datasets in different scenarios and comparison with other indicators.

2. Deep Residual Network

2.1 Network structure

The residual structure is utilized in ResNet to enhance the deep neural network's learning accuracy. The overall structure of ResNet, from input to output, comprises a convolution layer, a residual layer, a batch normalization layer, an activation function layer, a global mean pooling layer, and a fully connected output layer. The residual layer consists of a convolution layer, a batch standardization layer, an activation layer, among others. The residual module can be observed in Figure 1.

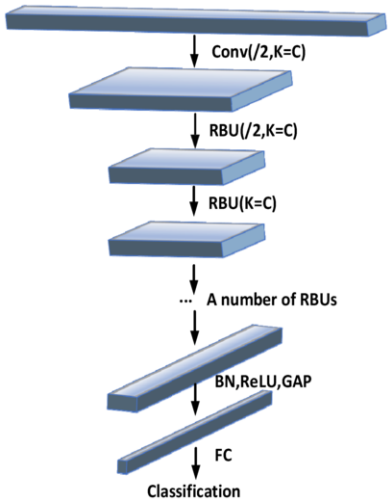


Figure.1 Structure of DRN

2.1.1. Convolutional Layer

In ResNet, the convolution operation is the most critical step, and its formula is as follows:

$$y^l = f(b_l + \sum_{i=1}^{k} F(x^l w_l)) \tag{1}$$

In the formula: l represents the current network; b_l represents the bias value; w_l is the weight between neurons; f represents the activation function; F represents the convolution operation; x^l is the output of the previous layer; y^l represents the output of the current layer; i represents the number of convolution operations. The Relu function is chosen as the activation function used in the model of this paper.

2.1.2. Batch Normalization Layer

This layer is used to re-adjust the distribution of the data, which is calculated as follows:

$$\partial_B^2 = \frac{1}{\mathrm{n}} \sum_{l=1}^{n} (xl - \frac{1}{\mathrm{n}} \sum_{l=1}^{n} x^l)^2 \tag{2}$$

$$y_l = \gamma_l \frac{x^l - \frac{1}{n} \sum_{i=1}^{n} x^l}{\sqrt{\partial_B^2 + \varepsilon}} + \beta_l \tag{3}$$

In the formula: ∂_B^2 represents the variance of the minimum batch processing; n represents the size of the batch processing data volume; x^l represents the lth sample data in the input of the batch layer; y_l represents the output of the final batch processing layer; ε is a constant that prevents the denominator from going to 0; γ and β represent the reconfiguration parameters that can be learned.

2.1.3. Residual Layer

ResNet mainly consists of multiple residual blocks, and the model in the paper contains two structures: the constant residual module and the convolutional constant residual module. The computational formula is:

$$yl = h(x^l) + p(x^l, \omega_l) \tag{4}$$

$$x^{l+1} = f(y^l) \tag{5}$$

In the formula: $h(x^l)$ stands for constant mapping or convolutional constant mapping; $p(x^l, \omega_l)$ stands for residual mapping. When $h(x^l)$ represents x^l, this signal can be transmitted directly from one cell to another cell, where the residual block expression after ignoring the activation operation is as follows:

$$x^{l+1} = x^l + p(x^l, \omega_l) \tag{6}$$

By iteration then the expression for the features of any deep cell can be reached as:

$$x_L = x^l + \sum_{i-l}^{L-1} p(x^i, \omega_i) \tag{7}$$

For the backpropagation phase in the backpropagation stage, according to the chain rule of backpropagation can be obtained:

$$\frac{\partial \varepsilon}{\partial x^l} = \frac{\partial \varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x^l} = \frac{\partial \varepsilon}{\partial x_L}(1 + \frac{\partial}{\partial x^l}\sum_{i=l}^{L-1} p(x^i, \omega_i)) \tag{8}$$

In this formula, ε represents the loss and the basic structure of the residual module is shown in Figure 2.

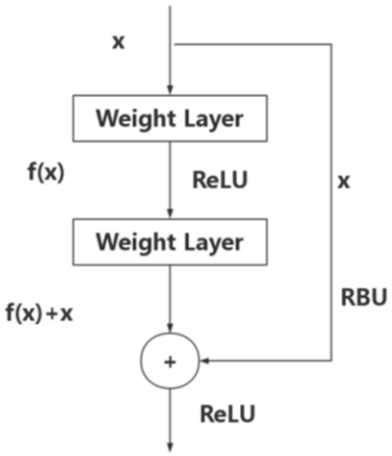


Figure.2 Residual module structure

2.1.4. Ponding Layer

Pooling operations reduce the complexity and computational effort of the network. Two of the most commonly used pooling methods are average pooling and maximum pooling.

2.1.5. Full Connectivity Layer

The fully connected layer synthesizes the extracted features and uses the corresponding system of linear equations to fit the inputs, which are modeled as follows:

$$y^l = b_l + f(\omega_l x^l) \tag{9}$$

In the formula: y^l is the current fully connected layer output.

2.1.6. Classifier Layer

The classifier layer is the last layer which uses the *Softmax* activation function and is able to convert the input neurons into probability distributions where the probability distributions sum to 1 to get the label distribution of the input data:

$$soft\max(z_i) = \frac{e^{z_i}}{\sum_{i=1}^{c} e^{z_i}} \tag{10}$$

In this formula, $z = [z_1, z_2, ..., z_c]$ represents the final output, and the value of the subscript is the number of neurons in the last layer, and also represents the number of classification labels of the classifier.

2.2 Network Training

In the residual structure, both steps including forward and backward propagation can be realized to update the parameters. In forward propagation, the output is represented as:

$$y_l = f_n(\cdots f_2(f_1(x^l Z^{(1)})Z^{(2)})\cdots Z^{(N)}) \tag{11}$$

In the above equation: Z represents the parameters of each layer in the network; x^l represents the input data; and f represents the activation function. For a single training sample x with label Y, define its loss function as:

$$C = \frac{1}{2}\left\|Y - y^l\right\|^2 \tag{12}$$

In the formula: C represents the loss function; Y represents the actual target value; y^l represents the prediction result. The formula for updating the weights and bias in the network is as follows:

$$\frac{\partial C}{\partial \omega_l} = \frac{\partial C}{\partial y_l}\frac{\partial y_l}{\partial \omega_l} = \delta^l x^l \tag{13}$$

$$\frac{\partial C}{\partial b_l} = \frac{\partial C}{\partial y_l}\frac{\partial y_l}{\partial b_l} = \partial^l \tag{14}$$

The update rule is obtained from the gradient descent method:

$$\omega_l^{'} = \omega_l - \alpha \wp^l x^l \tag{15}$$

$$b_l^{'} = b_l - \partial \wp^l \tag{16}$$

In the formula: $\wp^l$ and α represent the error and learning rate, respectively; y_l represents the output of the neuron in the first layer; l represents the input of the neuron in the previous layer. On this basis, this article utilizes the Adam algorithm to optimize the training of the model. The weights of the neural network can be iteratively updated based on the training data, enabling the network to converge faster and learn correctly, thereby minimizing the loss function.

3. Experimental Validation and Comparative Analysis

3.1 Experimental Data Presentation

To verify the applicability of the rolling bearing vibration signal evaluation method proposed in this article to bearing data in different scenarios, the following two sets of data were selected for experimental analysis and validation.

The first set of experimental data is the data published by the University of Cincinnati in the United States, as mentioned in reference [12]. In this experimental platform, 2700kg was applied to four bearings with a radial load of ZA-2115 type. During 163 hours of operation, one set of bearings experienced severe peeling failure until failure. The sampling frequency was 20kHz, and vibration signal data was collected every 10 minutes. A total of 984 sets of data were collected, recorded as Group A.

The second set of experimental data for the full life cycle of rolling bearings is sourced from reference [13]. The experimental platform used a 6307 bearing with a speed of 3000r/min and collected vibration signals. A radial load of 11.13kN was applied to the bearing with a sampling frequency of 25.6kHz, and a set of data was collected every 1 minute. A total of 2469 sets of full life data were collected, recorded as Group B.

3.2Model Parameter Setting and Selection

This paper is based on TensorFlow to build resnet-18 as the base model, the model selected Adam adaptive optimizer optimization parameters, set Batch Size = 64, epoch = 10, convolutional layers are used ReLU activation function, Dropout value of 0.5, that is, 50% of neurons are randomly selected to be deactivated in each training, to prevent the network structure overfitting, the structural parameters of the ResNet model are set as shown in Table 1.

Table 1. Model Structure Setting

Layer name	18-layer	Output size of group A	Output size of group B
Conv1	8×8,64,stride 2	32×32	32×32
Conv2_x	3×3 max pool,stride 2 $\begin{bmatrix} 3\times3,64 \\ 3\times3,64 \end{bmatrix}\times2$	13×13	13×13
Conv3_x	$\begin{bmatrix} 3\times3,128 \\ 3\times3,128 \end{bmatrix}\times2$	7×7	7×7
Conv4_x	$\begin{bmatrix} 3\times3,256 \\ 3\times3,256 \end{bmatrix}\times2$	3×3	3×3
conv5_x	$\begin{bmatrix} 3\times3,512 \\ 3\times3,512 \end{bmatrix}\times2$	2×2	2×2
	Average pool,Softmax	1×1	1×1

3.3. Construction Process of Rolling Bearing Degradation Index Based on DRN

DRN is commonly applied as supervised learning for fault classification and diagnosis of rolling bearings, and the performance degradation process is a gradual change, so it is not possible to divide the degradation process and treat it as a simple classification. The key of the proposed method in this paper is to set two nodes at the output of *Softmax* of DRN to get the probabilities representing the normal and failure states, respectively, and take the normalized amplitude spectrum of the normal state as the training sample, and take the data to be evaluated at the normal node as the basis of the output probability, and obtain the proposed index after formula processing. The flow chart of DRN-based rolling bearing degradation index construction is shown in Figure 3, and the specific implementation steps are shown below:

Step 1: Collect and organize the vibration signal data of the whole life cycle of rolling bearings, take a large number of raw signal data under normal working condition and process them by Fast Fourier Transform (FFT).

Step 2: The data obtained after normalization of the data obtained in Step 1 is used as a training sample, and the label is set to [1,0] to complete the construction of the DRN evaluation model.

Step 3: For the test bearing vibration signal, it is subjected to the FFT and normalization methods in Steps 1 and 2, and the obtained samples are then subjected sequentially to the DRN assessment model constructed and completed in Step 2, to obtain the probability P that the output of the test sample output of the *Softmax* output belongs to the normal state.

Step 4: The probability of *Softmax* output in DRN output layer is $P=\{p_1, p_2,...p_i..., p_n\}=\{P_d,...p_i..., p_n\}$, where n is the total number of samples, i is the ordinal number of samples, and p_d is the probability of normal state sample output. After processing P, the degradation index(DI) can be obtained. The purpose of the processing is to increase the range of probability values and facilitate the observation of the obtained results. The processing formula is:

$$f_{DI} = \frac{(p_i - \min(p_d)) \times 100}{\max(p_d) - \min(p_d)} \tag{17}$$

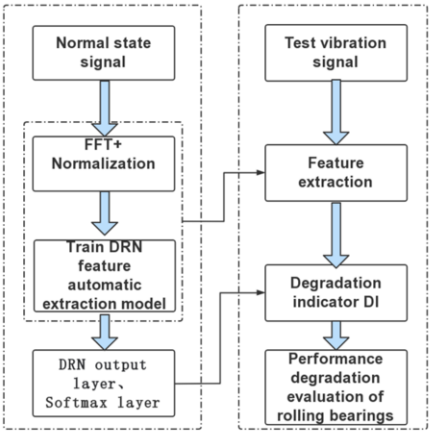
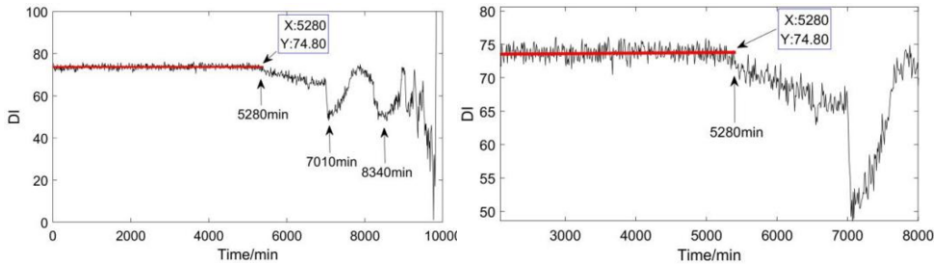


Figure.3 Evaluation model establishment process of DRN

3.4 Experimental Result

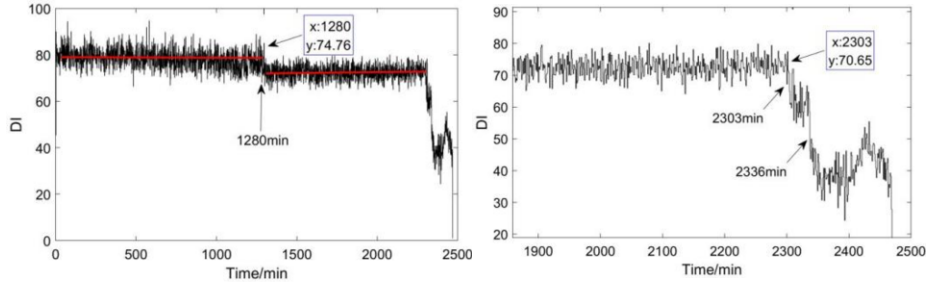
Firstly, analyze the data of Group A, which selected the first 300 groups as training samples. The degradation index DI under the entire life cycle is shown in Figure 4(a), and the locally enlarged image is shown in Figure 4(b). During the 0-5280 minutes of operation of the rolling bearing, the DI value of the sample is relatively stable, and this interval is considered as its normal working state; During the period of 5280min to 7010min, the DI value slowly decreases, indicating that the bearing has experienced early weak faults starting from 5280min, and as time increases, the characteristic trend of early weak faults will become more apparent; A sudden increase in sample value at 7010min indicates a significant weak fault; During the operation period of 7010min to 8340min, the overall trend of the DI value of the vibration signal is increasing first and then decreasing, which can be seen as the process of bearing failure and grinding to further deterioration [14], indicating that the bearing failure intensifies during this period; The DI value starting from 8340min shows irregular fluctuations, indicating that the bearing has failed at 8340min.



(a) Full life cycle **(b)**Partial enlarged drawing

Figure.4 Life cycle degradation index DI of group A

Furthermore, the first 400 sets of data from Group B were selected as training samples for analysis. The DI values of the degradation index for the entire life cycle are shown in Figure 5(a), and the locally enlarged image is shown in Figure 5(b). During the period of 0-1280min, the DI value is basically in a stable state, indicating that the bearing operates normally during this period; The DI value significantly decreased from 1280min, indicating that the bearing entered the early stage of weak failure at 1280min; During the period of 1280 to 2303 minutes, the trend of sample DI values did not change significantly, indicating that the bearing continued to maintain normal operation during this stage; The DI value suddenly decreases at 2303 minutes, and the trend of the sample DI value between 2303 and 2336 minutes is slightly increasing and slightly decreasing, which can also be seen as the process of fault smoothing and re deterioration. The DI value starting from 2336 minutes shows irregular fluctuations, indicating that the bearing has failed at 2336 minutes.



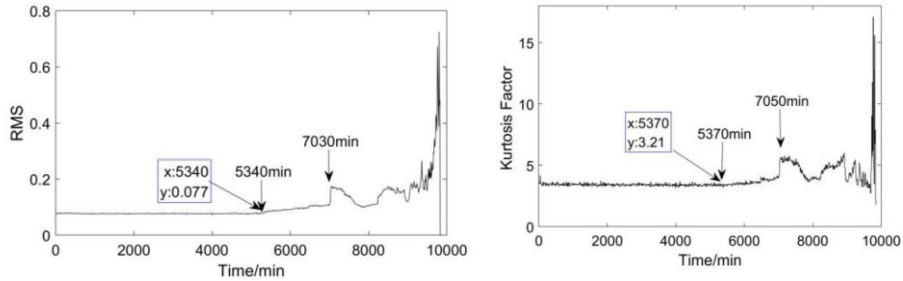
(a) Full life cycle **(b)**Partial enlarged drawing

Figure.5 Life cycle degradation index DI of group B

3.5 Comparison and Analysis

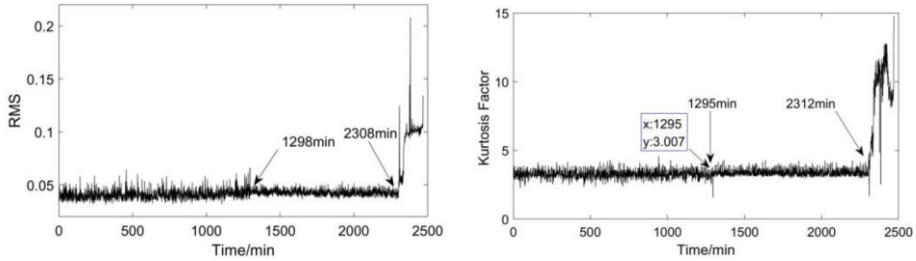
Two indicators, RMS and Kurtosis Factor, are selected to verify the validity of the method in this paper. RMS and Kurtosis Factor, as commonly used indicators in the degradation assessment, have better performance realizations compared with other traditional time-domain indicators. The changes of RMS and Kurtosis Factor indicators for the whole life cycle of group A data are shown in Figs. 6 and 7.

From Fig. 6(a) and Fig. 6(b), it can be seen that the early weak emergence moments are at 5340min and 5370min respectively, which indicates that the RMS indicator is more sensitive to the data of group A. However, the indicator proposed in this paper has already found the weak failure trend at 5290min, which is 60min ahead of the RMS indicator, which verifies the validity of this paper's indicator for group A data. From Fig. 7(a) and Fig. 7(b), it can be seen that the early weak emergence of the moment in 1298min and 1295min, respectively, indicating that the Kurtosis Factor indicator is more sensitive to the group B data, but this paper's proposed indicator has found a weak failure trend at 1280min, compared with the Kurtosis Factor indicator 15min ahead of the time, which verifies the validity of this paper's indicators for the group B data. From this, we can see that the traditional indicators can not be sensitive to all the data, and the indicators proposed in this paper are sensitive and feasible for both A and B data, and can also clearly describe the various stages of bearing degradation.



(a) Change of RMS of full life cycle in group A **(b)**Change of kurtosis factor of full life cycle in group A

Figure.6 RMS and kurtosis factor changes of the whole life cycle of group A



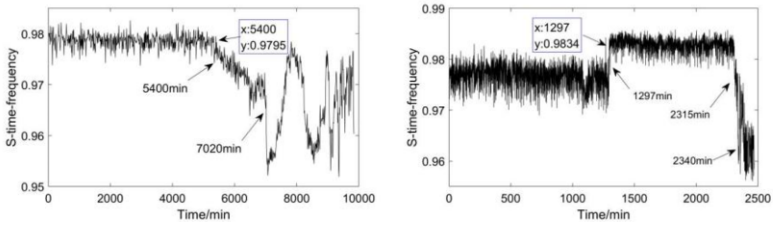
(a) Change of RMS of full life cycle in group B **(b)** Change of kurtosis facto of full life cycle in group B

Figure.7 RMS and kurtosis factor changes of the whole life cycle of group B

To verify the superiority of the proposed indicators, the S-time-frequency entropy proposed in reference [15] was selected for comparison. The S-time-frequency entropy changes over the entire life cycle of group A and B data are shown in Figure 8. From the time-frequency entropy change of Group A's entire life cycle S shown in Figure 8 (a), it can be seen that the rolling bearing operates normally during the 0-5400min period and has entered an early weak fault state since 5400min; Starting from 7020min, the bearing failure intensifies until failure; The time for detecting early weak faults using the indicators proposed in this article is 5280 minutes, which is 120 minutes earlier than before, indicating the superiority of the proposed indicators for Group A data;

From the time-frequency entropy changes of Group B throughout its entire life cycle S shown in Figure 8 (b), it can be seen that the rolling bearings during the 0-1297 minute period belong to normal working conditions. The bearings entered an early weak fault state after 1297-2315 minutes of operation, and the bearing failure began to intensify after 2315-2340 minutes. After 2340 minutes, the bearing failed. The indicators proposed in this article detect early weak faults and the time of fault exacerbation is 1280 minutes and 2303 minutes, which is 17 minutes and 12 minutes earlier than before, indicating the superiority of the proposed indicators for Group B data;

The S-time-frequency entropy failed to describe the occurrence of bearing failure stages in Group A data, and the description of each degradation stage in Group B data lacked monotonicity, failing to make the changes in indicators increase or decrease solely with the severity of the fault. However, the indicators proposed in this article effectively avoided these problems.



(a) Group A **(b)** Group B

Figure.8 S-time-frequency entropy change in the whole life cycle of group A and group B

In order to further verify the superiority of the indicators proposed in this article in sensitivity to early weak fault detection, the results of the occurrence time of early weak faults in groups A and B were compared with the methods proposed in references [16] and [17], respectively. Group A was about 60 minutes earlier than reference [16], and Group B was 15 minutes earlier than reference [17], verifying that this method has a certain sensitivity to early weak fault detection.

4. Conclusion

This paper proposes a rolling bearing degradation index construction method based on DRN, and through experimental verification and comparative analysis, it shows that the method has the following advantages:

(1) The feature self-extraction advantage of DRN solves the high dependence on feature extraction of traditional assessment methods.

(2) Utilizing the advantage that the output of *Softmax* layer in DRN can be two-dimensional data and can represent the two probabilities of normal and failure respectively, a new rolling bearing degradation index is obtained by setting the sample label as [1,0], and its effectiveness and superiority are verified.

(3) The evaluation results of this method have good consistency in the whole life cycle and have good sensitivity to the early weak degradation.

(4) The construction of the model can be realized with the data in normal state, which overcomes the problem of difficulty obtaining the fault samples in the operation of the actual bearing equipment, and it is of good significance for the monitoring of the bearing equipment performance.

Acknowledgment

(1) National key research and Development Plan "High-end bearing condition Monitoring and health Management Technology for Big Data"(2020YFB2007700)

(2) Shanghai Science and Technology plan project (21210750300)

(3) Shanghai Institute of Technology interdisciplinary graduate team (GN203006020-B20)

References

[1] Akhand Rai, S. H. Upadhyay. Bearing performance degradation assessment based on a combination of

empirical mode decomposition and k-medoids clustering[J]. Mec-hanical Systems and Signal Processing,2017,93.DOI：10.1016/j.ymssp.2017.02.003

[2] Li H B. The engineering design and development of intelligent bearing fault diagnosis[D]. Shenyang Ligong Uni-versity, 2009. (in Chinese)

[3] ZHOU J M, YIN W H, YOU T, et al. Review on performance degradation assessment of rolling bearings based ondata-driven [J]. Modern Manufacturing Engineering, 2021(05):146-153+160.

[4] JIA F, LEI Y, GUO L, et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines[J]. Neurocomputing, 2018, 272: 619-628.

[5] GUO L, LEI Y G, LI N P, et al. Deep convolution feature learning for health indicator construction of bearings[C]//2017 Prognostics and System Health Management Conference (PHM-Harbin). IEEE, 2017: 1-6.

[6] Diwan B, Yanda N, Suneetha M, et al. MSRFNet for skin lesion segmentation and deep learning with hybrid optimization for skin cancer detection[J]. The Imaging Science Journal,2023,71(7).

[7] Karthi S, L. R. S, M. K N. Water cycle tunicate swarm algorithm based deep residual network for virus detection with gene expression data[J]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,2023,11(5).

[8] Ridhi A, Balasubramanian R. BUS-Net: Breast Tumour Detection Network for Ultrasound Images Using Bi-directional ConvLSTM and Dense Residual Connections. [J]. Journal of digital imaging, 2022, 36(2).

[9] ZHAO J J, ZHAO Z H, YANG S P. Rolling bearing fault diagnosis based on residual connection and 1D-CNN[J]. Journal of Vibration and Shock,2021,40(10):1-6.

[10] Hou Liangsheng, Jiang Ruizheng, Tan Yanghui, Zhang Jundong. Input Feature Mappings-Based Deep Residual Networks for Fault Diagnosis of Rolling Element Bearing With Complicated Dataset[J]. IEEE ACCESS,2020,8.

[11] REN S, LIN G H, TIAN Z C, et al. Bearing fault diagnosis based on STFT-Inception-residual network[J]. Journal of Jilin University(Information Science Edition),2022,40(04):621-627.

[12] QIU H, LEE J, LIN J, et al. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics[J]. Journal of sound and vibration, 2006, 289(4-5): 1066-1090.

[13] PAN Y N. Study on feature extraction and assessment method of rolling element bearing performance degradation[D]. Shanghai: Shanghai Jiao Tong University, 2011. (in Chinese)

[14] ZHANG L, WU R Z. Performance degradation assessment of rolling bearings based on the entropy energy ratio of adaptive frequency band[J]. Journal of Vibration and Shock,2021, 40(02): 63-71+110.

[15] QIU H, LEE J, LIN J, et al. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics[J]. Journal of sound and vibration, 2006, 289(4-5): 1066-1090.

[16] PAN Y N, WEI T T, CHENG D L. Assessment of rolling bearing performance degradation using DBN and CHMM[J]. Mechanical Science and Technology for Aerospace Engineering:1-7[2022-11-14].

[17] CHENG D L, WEI T T, PAN Y N, et al. Performance degradation assessment method for rolling bearings based on DBN-SVDD[J]. Bearing, 2021(10):41-46.