Research Article

Rolling Bearing Fault Diagnosis Using Improved Deep Residual Shrinkage Networks

Zhijin Zhang¹, He Li¹, Lei Chen² and Ping Han¹

¹School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China

²Research Institute, Midea Group, Foshan 528311, China

Correspondence should be addressed to He Li; hli@mail.neu.edu.cn

Received 9 March 2021; Accepted 8 May 2021; Published 24 May 2021

Academic Editor: Huaitao Shi

Copyright © 2021 Zhijin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To improve feature learning ability and accurately diagnose the faults of rolling bearings under a strong background noise environment, we present a new shrinkage function named leaky thresholding to replace the soft thresholding in the deep residual shrinkage networks (DRSNs). In this work, we discover that such improved deep residual shrinkage networks (IDRSNs) can be realized by using a group searching method to optimize the slope value of leaky thresholding, and leaky thresholding in the IDRSNs can more effectively eliminate the noise of signal features. We highlight that our techniques can significantly improve the performance on various fundamental tasks. Experimental results show that IDRSNs achieve better fault diagnosis results on noised vibration signals compared with DRSNs. Moreover, we also provide a normalized processing to further improve the fault diagnosing accuracy of rolling bearing under a strong background noise environment.

1. Introduction

Rotating machines are integral to both industrial production and daily life. Once the rotating machines break down, it will not only reduce the production efficiency of the industrial production process and cause economic losses but also threaten the safety of human production and work. Rolling bearing is one of the key parts of rotating machinery [1, 2]. Incomplete statistics indicate that 30% of rotating machinery failures are caused by bearing failures [3, 4]. Therefore, it is highly significant to study the fault diagnosis methods of rolling bearing.

The existing methods mainly include the signal-processing-based methods and the machine-learning-based methods. Signal-processing-based methods implement fault diagnosis by detecting characteristic frequencies which are related to the faults. In general, individuals can diagnose the faults by signal-processing-based methods; consequently, a multitude of researchers have poured attention into signal-processing-based methods. For example, Li et al. [5] developed a variational-mode-decomposition-based bearing fault diagnosis method, which can effectively identify fault

frequencies. Wang et al. [6] proposed a wavelet-transform-based method, which can effectively detect the hidden fault features in rotating machinery. However, the signal-processing-based methods need to rely on professional knowledge. In addition, the original vibration signals of the faults are easy to be overwhelmed with the strong noise in an actual environment, especially in the stage of incipient faults [7]. Therefore, it is arduous for signal-processing-based methods to realize accurate fault diagnosis under an actual strong noise environment.

The machine-learning-based intelligent methods [8], on the other hand, can perform the fault diagnosis task without the fault-related characteristics frequencies and prior physical knowledge. It mainly learns the sensitive fault features of the original vibration signals as the training features of the diagnosis model to identify the health state of the machine, including artificial neural networks (ANNs) [9], support vector machine (SVM) [10], k-nearest neighbors (KNNs) [11], probabilistic graphical model (PGM) [12], and deep learning (DL) [13]. DL methods have been widely used in fault diagnosis task because of its strong ability of automatic learning features, including stacking autoencoders

(SAEs) [14–19], deep belief networks (DBNs) [20–24], and convolutional neural networks (CNNs) [25–29]. For example, Xiang et al. [18] used stacking autoencoders (SAEs) to diagnose rolling bearings without being affected by the speed and load. Mao et al. [19] applied the new deep autoencoders method to effectively improve the diagnosing accuracy. Shao et al. [23] proposed an adaptive deep belief networks (DBNs), which further improved the diagnosing accuracy and convergence speed. Chen et al. [25] effectively improved the diagnosing accuracy of rolling bearings by combining the preprocessing method based on cyclic spectral coherence with convolutional neural networks. Recently, deep CNNs have become the main-stream solution to many tasks. However, deep CNNs are more formidable to train than shallow neural networks, and thus, it would be arduous to diagnose accurately when the training process failed. Therefore, He et al. [30] developed deep residual networks (ResNets), which use identity shortcuts, in order to reduce the difficulty of the training process of deep CNNs. In recent years, ResNets have become the main-stream solution to fault diagnosis tasks [31–34]. As an example, Zhang et al. [31] developed a fault diagnosis method based on ResNets for rolling bearing. Peng et al. [34] integrated ResNets with CNNs for a wheelset bearing fault diagnosis.

However, in the actual application environment of rolling bearing, the vibration signals collected often contain a large amount of noise. Therefore, it is necessary to develop a method which is suitable for highly noised vibration signals. In the recent paper, Shao et al. [35] developed a new deep residual shrinkage network (DRSN), which integrated ResNets with soft thresholding. Thus, the DRSN is an evolution body of the ResNet. As mentioned in [35], DRSN can improve the fault diagnosing accuracy of vibration signals containing strong noise, so that DRSN is suitable for fault diagnosis under strong background noise. However, the soft thresholding in the DRSN may eliminate the effective signal features except noise in the process of feature learning, so that DRSNs fail to implement accurate fault diagnosis due to the loss of valid features. Therefore, it is essential to develop a new shrinkage function in the DRSN for vibration signals containing strong noise.

In this paper, inspired by the latest work, DRSNs are improved to address this issue of soft thresholding in the DRSNs under strong background noise and further improve the fault diagnosing accuracy of vibration signals containing strong noise, with the final objective of yielding high fault diagnosing accuracy of rolling bearing under strong background noise. The main contributions are summarized as follows:

- (1) In this article, we firstly develop a new shrinkage function named leaky thresholding and replace the soft thresholding with leaky thresholding in the deep residual shrinkage networks. As mentioned in [35], the soft thresholding may eliminate the effective signal features except noise. Our leaky thresholding

further improves the fault diagnosing accuracy of vibration signals containing strong noise.

- (2) Secondly, we provide a group searching method to determine the slope value of leaky thresholding under different input signals. The most optimal slope value can be selected by using the group searching method such that the whole network can achieve the better diagnosing accuracy to improve the diagnosis effect of the deep residual shrinkage networks for vibration signals containing noise.
- (3) Finally, we discover that the normalized original vibration signals can further improve the fault diagnosing accuracy of rolling bearings under strong background noise.

The rest of this article is organized as follows. The basic concepts of deep residual shrinkage networks and a detailed elaboration on leaky-thresholding are illustrated in Section 2. The experimental comparisons are in Section 3, in order to illustrate the effectiveness of the improved deep residual shrinkage networks proposed in this paper, and conclusions are summarized in Section 4.

2. Theory of IDRSN

2.1. DRSN Model

2.1.1. Basic Components. From fundamental principles, DRSN is an improved variant of the ResNet so that DRSN's basic components are the same as the ResNet, including the convolutional layer, batch normalization (BN), activation function, global average pooling (GAP), and loss function. The detailed description is given as follows.

The aim of the convolutional layer is to extract different features of the input, which can reduce parameters in network training, avoid the occurrence of overfitting, and improve the network model accuracy. The mapping relationship between the input features of the convolutional layer and the convolutional kernel can be expressed as follows:

$$y_j = \sum_{i \in M_j} x_i * k_{ij} + b_j, \quad (1)$$

where x and y are input feature map and output feature map, respectively. b is the bias, k is the convolutional kernel, i and j are indicators of the channels, and M_j is a channel collection for calculating y_j .

BN is a technique to normalize input features during deep learning networks training. BN is mainly divided into two stages: in the first stage, BN adjusts features to the standard normal distribution; in the second stage, BN adjusts features to the distribution of ideal. As a result, BN can not only reduce internal covariant shift but also prevent the vanishing gradient problem and improve the learning convergence speed, so that BN can improve the training

speed of the model. The formula of the batch normalization is expressed by

$$\mu = \frac{1}{N_{\text{batch}}} \sum_{n=1}^{N_{\text{batch}}} x_n, \quad (2)$$

$$\sigma^2 = \frac{1}{N_{\text{batch}}} \sum_{n=1}^{N_{\text{batch}}} (x_n - \mu)^2, \quad (3)$$

$$\hat{x}_n = \frac{x_n - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \quad (4)$$

$$y_n = \gamma \hat{x}_n + \beta, \quad (5)$$

where x_n represents the input of the n th observation and y_n represents the output of the n th observation. N_{batch} represents the size of minibatch. ε is a constant value which is close to zero. γ is a parameter value which is used to scale the distribution, and β is a parameter value which is used to move the distribution. Both γ and β can be obtained by training.

Activation function is an important component of the neural network, which is mainly used for nonlinear transformation. The rectifier linear unit (ReLU) has been widely used in the neural network because it can effectively prevent gradient vanishing. A rectifier linear unit is expressed by

$$y = \begin{cases} x, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (6)$$

where x and y are the input feature and output feature, respectively.

GAP can carry out average value calculation from each channel before the FC output layer. The advantages of GAP include two aspects. GAP can not only reduce the number of neural network parameters but also reduce the probability of the fitting phenomenon encountered in the neural network training. Furthermore, GAP can avoid the shift variant problem and reduce the effect of fault pulse position for training learning characteristics.

In the multicategory classification task, the cross-entropy error is often adopted as loss function because it can improve the training efficiency of the neural network. It is necessary to apply softmax function before calculating the cross-entropy error. The softmax function is expressed by

$$y_j = \frac{e^{x_j}}{\sum_{i=1}^{N_{\text{class}}} e^{x_i}}, \quad (7)$$

where x_j and y_j are the input and output of j th neuron in the function, respectively. x_i is an input of i th neuron, and N_{class} represents the number of neurons. The cross-entropy function is expressed by

$$E = - \sum_{j=1}^{N_{\text{class}}} t_j \log(y_j), \quad (8)$$

where t_i is the actual probability of j th neuron at the output layer.

2.1.2. Architecture of the RSBUs. The DRSN is an evolution body of the ResNet. The soft thresholding in the DRSN can effectively get rid of noise. Soft thresholding has been often used as a denoising method in the field of signal processing. In general, the soft thresholding realizes denoising of the original signal by converting the near-zero feature of original signal to zero. The function of soft thresholding is expressed by

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau < x < \tau \\ x + \tau & x < -\tau \end{cases}, \quad (9)$$

where x and y are the input feature and output feature, respectively. τ is the threshold, and it is a positive number. Its derivative form is as follows:

$$\frac{\partial y}{\partial x} = \begin{cases} 1 & x > \tau \\ 0 & -\tau < x < \tau \\ 1 & x < -\tau \end{cases}. \quad (10)$$

It can be observed from the above function that the result of the function is either one or zero. Therefore, the soft thresholding can not only remove the noise of signals but also prevent the problem of gradient vanishing.

The most important module of the DRSN is the residual shrinkage building unit (RSBU), as shown in Figure 1. C is the size of channels, W is the size of width, and the number 1 is the size of height. An RSBU includes two BNs, two ReLUs, two convolutional layers, a threshold module, a soft thresholding module, and an identity shortcut. Threshold module is composed of an absolute, a GAP, a BN, a ReLU, and two output FC layers. Threshold module can automatically determine the threshold in the process of neural network training. The detailed process is that the feature map x is propagated into an absolute operation, a GAP, a BN, a ReLU, and two output FC layers, the number of channels in the second FC layer is the same as the feature map x . The scaling parameter is obtained by the formula as follows:

$$\alpha_c = \frac{1}{1 + e^{-z_c}}, \quad (11)$$

where z and α are the feature of neuron and the scaling parameter, respectively. c is the index. Furthermore, the threshold calculation formula is expressed as follows:

$$\tau_c = \alpha_c \cdot \text{average} |x_{i,j,c}|, \quad (12)$$

where τ is the threshold of the feature map x and i, j , and c are the indexes of width, height, and channel of the feature map x , respectively.

2.1.3. The DRSN's Structure. The structure of a DRSN is shown in Figure 2, which consists of an input layer, a convolutional layer, one or more RSBUs, a BN, a ReLU, a

GAP layer, and an output FC layer. Finally, the fault classification results can be obtained.

2.2. Leaky Thresholding. In general, the near-zero features of highly noised vibration signals are unimportant. It is necessary for accurate fault diagnosis to eliminate the unimportant features. As mentioned in [35], the soft thresholding in the DRSN can effectively eliminate the unimportant features. However, the soft thresholding may eliminate the effective signal features except noise, leading to the reduction of fault diagnosing accuracy. Therefore, in order to address this issue, we firstly develop a new shrinkage function named leaky thresholding. The leaky thresholding can retain the effective features of the vibration signals containing noise as much as possible, and the function of leaky thresholding is expressed by

$$y = \begin{cases} x + (\alpha - 1)\tau & x > \tau \\ \alpha \cdot x & -\tau < x < \tau \\ x + (1 - \alpha)\tau & x < -\tau \end{cases}, \quad (13)$$

where x and y are the input feature and output feature, respectively. τ and α are the threshold and the slope of the shrinkage function, respectively. And, τ and α are positive numbers.

The leaky-thresholding process is shown in Figure 3(a). As shown in Figure 3(b), the result of the leaky-thresholding function derivative is either 1 or α , which can also prevent gradient vanishing and explosion problems during model training. The derivative of the leaky-thresholding function is expressed by

$$\frac{\partial y}{\partial x} = \begin{cases} 1 & x > \tau \\ \alpha & -\tau < x < \tau \\ 1 & x < -\tau \end{cases}. \quad (14)$$

2.3. Improved DRSN Model (IDRSN). In order to address the issue that soft thresholding in the DRSN model may eliminate the effective signal features except noise, in this paper, we replace the soft thresholding with leaky thresholding in the DRSN. The architecture of the improved residual shrinkage building unit (IRSBU) is shown in Figure 4.

The basic architecture of IDRSN is similar to the DRSN, as shown in Figure 5. The improved residual shrinkage building unit (IRSBU) proposed in this paper needs to be replaced with the RSBU unit in Figure 2 to form an improved deep residual shrinkage network (IDRSN).

3. Experiments and Results

The developed IDRSN was implemented in the TensorFlow environment using Python. In this paper, we implemented experiments on a computer with a NVIDIA GeForce RTX 2060 MAX-Q graphics processing unit and an i7-10875H CPU. The experimental results have been summarized in this section.

3.1. Description of Datasets. As shown in Figure 6, we adopt the vibration signal datasets which are obtained from the Case Western Reserve University (CWRU), in order to prove the IDRSNs' ability of rolling bearing fault diagnosis when dealing with highly noised vibration signals. In this paper, the bearing conditions include a healthy condition (H) and three fault conditions. The fault conditions include inner race fault (IF), outer race fault (OF), and ball fault (BF), and the fault diameter at each fault condition is 7 mils, 14 mils, and 21 mils, respectively. Therefore, the datasets adopted in this paper include ten bearing conditions, and the datasets are under a load of 0hp. Each observation includes 400 data points, and each bearing condition has 3000 observations. The training sets consist of 2800 observations, and the test sets consist of 200 observations. The detailed description of the datasets is presented in Table 1.

3.2. Noise Preparation. To verify the IDRSNs' ability of fault diagnosis under a strong background noise environment, Gaussian white noise was performed on the original vibration signal to simulate the real environment. Signal to noise ratio (SNR) stands for noise intensity, and its expression is as follows:

$$\text{SNR} = 10 \log \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \quad (16)$$

where P_{signal} and P_{noise} are signal energy and noise energy, respectively. SNR stands for noise intensity in dB.

In this paper, Gaussian noise was performed on the original vibration signal with noise intensity ranging from -4 dB to 6 dB. The vibration signals containing noise were used as training datasets and test datasets to simulate the real strong background noise environment. The training datasets and test datasets were generated randomly each time, in order to reduce the randomness of the experimental results.

3.3. Experimental Comparison with the DRSN. The superiority of DRSNs compared to ResNets and deep CNNs under strong background noise has been validated in [35]. Thus, this article takes DRSNs as a benchmark to be further compared. And, the architecture-related hyperparameters selected in this article are based on the popular recommendations [35], as presented in Table 2, including the number of layers, the number of convolutional kernels, and the size of convolutional kernels. The first numbers in the bracket of the third and fourth column are the number of convolutional kernels, and the second numbers in the bracket of the third and fourth column are the width of convolutional kernels. The meaning of “/2” in the bracket is that the step size of convolution kernels is 2.

In the training process, we define some optimization-related hyperparameters, the Adam optimizer is adopted, and the learning rate is set to 0.001 in the 50 epochs. The coefficients of L2 regularization are set to 0.0001, and minibatch size is set to 16.

The accuracy results of the DRSN and the IDRSN for fault diagnosis under the influence of noise of different

FIGURE 1: Residual shrinkage building unit (RSBU).

FIGURE 2: Structure of the DRSN.

intensity are listed in Table 3. The IDRSN adopts the leaky thresholding, which is proposed in this paper, and the slope value parameter α of the leaky thresholding adopts the group searching method which is provided in this paper. The process of the group searching method for the leaky thresholding developed in this article is shown in Figure 7. The slope value that makes the diagnosis accuracy reach the optimum can be selected as the model training parameter value. In this paper, the group of slope value consisted of 0.001, 0.003, 0.01, 0.05, 0.075, 0.1, and 0.2. The fault diagnosis results of different slope values in the group are listed in Table 3. The accuracy of test samples in the table is the average values of the five experimental results as the final experimental results.

In order to prove the superiority of the IDRSN over the DRSN, the IDRSN is compared with the DRSN under different noise intensity, and the comparison results of fault diagnosing accuracy of the IDRSN proposed in this article and the DRSN are shown in Figure 8. The accuracy result under optimal slope is selected as the IDRSN's actual fault diagnosing accuracy. As indicated in Figure 8, the developed IDRSN for rolling bearing fault diagnosis under strong background noise is more advantageous than the DRSN, and the experimental results show that the developed IDRSN yields improvements of 1.88% in terms of average diagnosis accuracy under different noise intensity, when compared with the DRSN. This paper demonstrates that the proposed leaky thresholding effectively to address the issue of soft

FIGURE 3: (a) Leaky-thresholding illustration and (b) its derivative illustration.

FIGURE 4: Improved residual shrinkage building unit (IRSBU).

FIGURE 5: Architecture of the IDRSN.

FIGURE 6: Test stand used by CWRU.

TABLE 1: The description of the experimental datasets.

Bearing condition	H		OF		IF		BF			
Label	0	1	2	3	4	5	6	7	8	9
Damage diameter	0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021

TABLE 2: Experiment-related hyperparameters of DRSN and IDRSN.

The number of units	The size of output	DRSN	IDRSN
1	$1 \times 400 \times 1$	Input	Input
1	$4 \times 200 \times 1$	Conv (4, 3/2)	Conv (4, 3/2)
1	$4 \times 100 \times 1$	RSBU (4, 3/2)	IRSBUs (4, 3/2)
1	$4 \times 100 \times 1$	RSBU (4, 3)	IRSBUs (4, 3)
1	$8 \times 50 \times 1$	RSBU (8, 3/2)	IRSBUs (8, 3/2)
1	$8 \times 50 \times 1$	RSBU (8, 3)	IRSBUs (8, 3)
1	$16 \times 25 \times 1$	RSBU (16, 3/2)	IRSBUs (16, 3/2)
1	$16 \times 25 \times 1$	RSBU (16, 3)	IRSBUs (16, 3)
1	16	BN	BN
1	16	ReLU	ReLU
1	16	GAP	GAP
1	10	FC	FC

TABLE 3: Fault diagnosis results of the DRSN and IDRSN.

SNR (dB)	DRSN (%)	IDRSN (%) (0.001)	IDRSN (%) (0.003)	IDRSN (%) (0.01)	IDRSN (%) (0.05)	IDRSN (%) (0.075)	IDRSN (%) (0.1)	IDRSN (%) (0.2)
-4	85.8	85.5	82.9	83.5	82.6	86.0	84.3	86.7
-2	86.8	88.8	87.5	85.7	90.2	88.5	90.5	90.0
0	91.3	90.9	92.1	90.6	91.1	89.5	91.1	90.5
2	93.1	92.7	91.0	91.7	92.9	94.5	93.4	93.1
4	93.2	95.5	95.3	92.4	92.7	93.2	94.8	93.4
6	94.3	95.9	94.3	96.5	93.8	94.8	95.8	94.8

FIGURE 7: Flowchart of the proposed group searching method.

thresholding may get rid of useful characteristics in the process of signal denoising.

To further improve the average test accuracy, the input signal is preprocessed and the normalization method is adopted. The normalization process is expressed by

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (17)$$

where x and y are the input signal and output signal, respectively. $x_{\max}$ stands for the maximum of input signal and $x_{\min}$ is the minimum of input signal.

The results of fault diagnosing accuracy of the normalized DRSN and normalized IDRSN under different noise intensity are summarized in Table 4.

As shown in Figure 9, the accuracy of the normalized IDRSN and normalized DRSN is significantly improved compared with that before normalization, and the accuracy of the normalized IDRSN is higher than that before normalization. It can be observed that the normalized IDRSN and normalized DRSN yields improvements of 4.53% and 4.4% in terms of average diagnosis accuracy under different noise intensity, when compared with the IDRSN and DRSN, respectively.

FIGURE 8: Fault diagnosing accuracy of IDRSN and DRSN under different noise intensities.

TABLE 4: Fault diagnosing accuracy results of the normalized DRSN and normalized IDRSN.

SNR (dB)	DRSN (%)	IDRSN (%) (0.001)	IDRSN (%) (0.003)	IDRSN (%) (0.01)	IDRSN (%) (0.05)	IDRSN (%) (0.075)	IDRSN (%) (0.1)	IDRSN (%) (0.2)
-4	87.6	92.0	90.8	90.4	93.4	84.3	88.5	90.4
-2	93.8	93.3	92.2	87.6	94.9	93.5	92.6	91.0
0	92.6	96.8	95.4	91.4	95.5	95.1	93.5	94.5
2	98.1	95.7	98.1	98.4	98.1	96.5	96.2	97.3
4	99.3	98.9	97.4	99.7	97.4	99.4	98.3	99.3
6	99.5	98.1	99.6	99.3	99.2	99.8	99.8	99.1

FIGURE 9: Comparison of accuracy results of normalized IDRSN, normalized DRSN, IDRSN, and DRSN under different noise intensities.

The computational time of the DRSN, IDRSN, normalized DRSN, and normalized IDRSN is summarized in Table 5. It can be observed that our IDRSN proposed in

this paper improves the accuracy of fault diagnosis without increasing the computational complexity of the model.

TABLE 5: Computational time of the considered methods.

Method	Time (s)
DRSN	98.84
IDRSN	103.25
DRSN_normalization	99.42
IDRSN_normalization	104.10

4. Conclusions

In this paper, inspired by the latest work [35], we proposed an improved deep residual shrinkage network (IDRSN) to improve the fault diagnosing accuracy of rolling bearing under strong background noise environment. As described by the authors in [35], soft thresholding may eliminate the effective signal features except noise. We develop a new shrinkage function named leaky thresholding to replace the soft thresholding with leaky thresholding in the DRSN. The leaky thresholding can retain the effective features of the vibration signal containing noise as much as possible. The slope value of leaky thresholding can be determined by using group searching method, where the method can select the most optimal slope as the model training parameter value.

Compared with the original DRSN in [35], our IDRSN can achieve better simulation results of rolling bearing fault diagnosis when vibration signals contain noise. The superiority of the IDRSN compared with the DRSN in improving fault diagnosing accuracy of rolling bearing has been verified in the experiment. The IDRSN outperformed the DRSN by yielding improvements of 1.88%. In this paper, we also provide a normalized processing to further improve the diagnosis accuracy. The normalized IDRSN and normalized DRSN outperformed the IDRSN and DRSN by yielding improvements of 4.53% and 4.4%, respectively.

Therefore, our IDRSN in this paper has better fault diagnosis effect on noised vibration signals compared with the DRSN, and the normalized processing can further improve the fault diagnosis effect of the IDRSN and DRSN. In summary, the IDRSN developed in this paper can effectively improve the fault diagnosing accuracy of rolling bearing under strong background noise environment.

Data Availability

The vibration signal datasets of the Case Western Reserve University (CWRU) can be downloaded from website: <https://csegroups.case.edu/bearingdatacenter/home>.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 51675091).

References

- [1] H. T. Shi and X. T. Bai, "Model-based uneven loading condition monitoring of full ceramic ball bearings in starved lubrication," *Mechanical Systems and Signal Processing*, vol. 139, 2020.
- [2] H. M. Zhao, H. D. Liu, Y. Jin, X. J. Dang, and W. Deng, "Feature extraction for data-driven remaining useful life prediction of rolling bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021.
- [3] Y. Wang, B. Tang, Y. Qin, and T. Huang, "Rolling bearing fault detection of civil aircraft engine based on adaptive estimation of instantaneous angular speed," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4938–4948, 2020.
- [4] H. Li, T. Liu, X. Wu, and Q. Chen, "Research on bearing fault feature extraction based on singular value decomposition and optimized frequency band entropy," *Mechanical Systems and Signal Processing*, vol. 118, pp. 477–502, 2019.
- [5] G. Li, G. Tang, G. Luo, and H. Wang, "Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition," *Mechanical Systems and Signal Processing*, vol. 120, pp. 83–97, 2019.
- [6] D. Wang and K.-L. Tsui, "Dynamic bayesian wavelet transform: new methodology for extraction of repetitive transients," *Mechanical Systems and Signal Processing*, vol. 88, pp. 137–144, 2017.
- [7] M. X. Hou and H. T. Shi, "Stator-winding incipient shorted-turn fault detection for motor system in motorized spindle using modified interval observers," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2020.
- [8] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.
- [9] T. Tang, L. Bo, X. Liu, B. Sun, and D. Wei, "Variable predictive model class discrimination using novel predictive models and adaptive feature selection for bearing fault identification," *Journal of Sound and Vibration*, vol. 425, pp. 137–148, 2018.
- [10] Y. Li, B. Miao, W. Zhang, P. Chen, J. Liu, and X. Jiang, "Refined composite multiscale fuzzy entropy: localized defect detection of rolling element bearing," *Journal of Mechanical Science and Technology*, vol. 33, no. 1, pp. 109–120, 2019.
- [11] J. Ma, F. Xu, K. Huang, and R. Huang, "GNAR-GARCH model and its application in feature extraction for rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 93, pp. 175–203, 2017.
- [12] M. Y. Asr, M. M. Ettefagh, R. Hassannejad, and S. N. Razavi, "Diagnosis of combined faults in rotary machinery by non-naive bayesian approach," *Mechanical Systems and Signal Processing*, vol. 85, pp. 56–70, 2017.
- [13] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning

- towards mechanical big data,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [14] C. Li, W. Zhang, G. Peng, and S. Liu, “Bearing fault diagnosis using fully-connected winner-take-all autoencoder,” *IEEE Access*, vol. 6, pp. 6103–6115, 2018.
 - [15] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, “Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features,” *Mechanical Systems and Signal Processing*, vol. 99, pp. 459–477, 2018.
 - [16] Z. Meng, X. Zhan, J. Li, and Z. Pan, “An enhancement denoising autoencoder for rolling bearing fault diagnosis,” *Measurement*, vol. 130, pp. 448–454, 2018.
 - [17] J. Sun, C. Yan, and J. Wen, “Intelligent bearing fault Diagnosis method combining compressed data acquisition and deep learning,” *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 185–195, 2018.
 - [18] Z. Xiang, X. Zhang, W. Zhang, and X. Xia, “Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO spectrum and stacking auto-encoder,” *Measurement*, vol. 138, pp. 162–174, 2019.
 - [19] W. T. Mao, W. S. Feng, Y. M. Liu, D. Zhang, and X. H. Liang, “A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 150, 2021.
 - [20] S. Tang, C. Shen, D. Wang, S. Li, W. Huang, and Z. Zhu, “Adaptive deep feature learning network with nesterov momentum and its application to rotating machinery fault diagnosis,” *Neurocomputing*, vol. 305, pp. 1–14, 2018.
 - [21] H. Shao, H. Jiang, F. Wang, and Y. Wang, “Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet,” *ISA Transactions*, vol. 69, pp. 187–201, 2017.
 - [22] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, “Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing,” *Mechanical Systems and Signal Processing*, vol. 100, pp. 743–765, 2018.
 - [23] H. Shao, H. Jiang, H. Zhang, and T. Liang, “Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, 2018.
 - [24] Z. Gao, C. Ma, D. Song, and Y. Liu, “Deep quantum inspired neural network with application to aircraft fuel system fault diagnosis,” *Neurocomputing*, vol. 238, pp. 13–23, 2017.
 - [25] Z. Y. Chen, A. Mauricio, W. H. Li, and K. Gryllias, “A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks,” *Mechanical Systems and Signal Processing*, vol. 140, 2020.
 - [26] L. Eren, T. Ince, and S. Kiranyaz, “A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier,” *Journal of Signal Processing Systems*, vol. 91, no. 2, pp. 179–189, 2019.
 - [27] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, and S. Liu, “ACDIN: bridging the gap between artificial and real bearing damages for bearing fault diagnosis,” *Neurocomputing*, vol. 294, pp. 61–71, 2018.
 - [28] F. Jia, Y. Lei, N. Lu, and S. Xing, “Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization,” *Mechanical Systems and Signal Processing*, vol. 110, pp. 349–367, 2018.
 - [29] D. K. Appana, A. Prosvirin, and J.-M. Kim, “Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks,” *Soft Computing*, vol. 22, no. 20, pp. 6719–6729, 2018.
 - [30] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Seattle, WA, USA, March 2016.
 - [31] W. Zhang, X. Li, and Q. Ding, “Deep residual learning-based fault diagnosis method for rotating machinery,” *ISA Transactions*, vol. 95, pp. 295–305, 2019.
 - [32] M. Zhao, M. Kang, B. Tang, and M. Pecht, “Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4290–4300, 2018.
 - [33] S. Ma, F. Chu, and Q. Han, “Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions,” *Mechanical Systems and Signal Processing*, vol. 127, pp. 190–201, 2019.
 - [34] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, “A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains,” *IEEE Access*, vol. 7, pp. 10278–10293, 2019.
 - [35] 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.