

## Research Article

# Fault Early Warning Based on Improved Deep Neural Network of Auto-Encoder

Huijuan Hao <sup>1</sup>, Huimiao Yuan <sup>1</sup>, Yongwei Tang <sup>1,2</sup>, Yu Zhang <sup>1</sup>, Yuanyuan Zhao <sup>1</sup>, and Qingxuan Wei <sup>3</sup>

<sup>1</sup>Qilu University of Technology (Shandong Academy of Sciences),  
Shandong Computer Science Center (National Supercomputer Center in Jinan),  
Shandong Key Laboratory of Computer Networks, Jinan 250014, China

<sup>2</sup>School of Mechanical Engineering, Shandong University, Key Laboratory of High Efficiency and Clean Mechanical Manufacture,  
Jinan 250100, China

<sup>3</sup>Beijing Institute of Petrochemical Technology, School of Information Engineering,  
Information Technology Teaching and Experiment Center, Beijing 102617, China

Correspondence should be addressed to Yongwei Tang; [tangyw@sdas.org](mailto:tangyw@sdas.org)

Received 20 June 2022; Accepted 25 July 2022; Published 9 September 2022

Academic Editor: Shudong Li

Copyright © 2022 Huijuan Hao 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.

In order to realize rapid fault detection and early warning, a fault detection method based on normal operation data is proposed. Firstly, the fault detection model is constructed based on the improved deep neural network of the auto-encoder. Secondly, the unsupervised pretraining and supervised fine-tuning of the network are finished through the operation data in a normal state to solve the contradiction between the small fault sample and the large training sample required by the deep network model. The adaptive threshold of reconstruction error is used as the evaluation index of the fault state to reduce the influence of environmental factors. Experimental results show that the proposed method can detect faults effectively.

## 1. Introduction

Major accidents caused by mechanical equipment failures always alert people to ensure the safe and reliable operation of equipment. Especially, the equipment failure of the key production line will bring significant shutdown losses. It will not only cause huge economic losses but also endanger personal safety in serious cases. The fault diagnosis and prediction of mechanical equipment play an important role in improving equipment operation reliability and optimizing operation [1]. It is the premise to ensure the safe and stable operation of mechanical equipment, and it is very important for the maintenance of mechanical equipment.

Due to the large number of measuring points, monitoring parameters, and high sampling frequency of mechanical equipment, the complex condition of monitoring big data has been formed, which makes the fault diagnosis of mechanical equipment enter the era of big data [2–6]. The

big data bring new opportunities and challenges to the health monitoring and fault diagnosis of mechanical equipment [7]. In recent years, deep learning has made a breakthrough in big data analysis in the fields of speech recognition and image recognition, and deep learning theory has been applied in the fields of mechanical fault diagnosis and health monitoring [8–11], but the diagnosis and identification of faults are still in the preliminary exploration stage [12–29]. A deep neural network of auto-encoder (DAE) can extract fault features from noise signals and can be well combined with a sample enhancement method to deal with small sample problems [30–37]. Fault diagnosis methods based on deep learning are often realized through supervised learning of a large number of fault data, which requires a large number of fault marking data. However, in practical industrial applications, the monitoring data are a large amount of normal operation data, and the fault data under abnormal conditions are difficult to collect from the mechanical

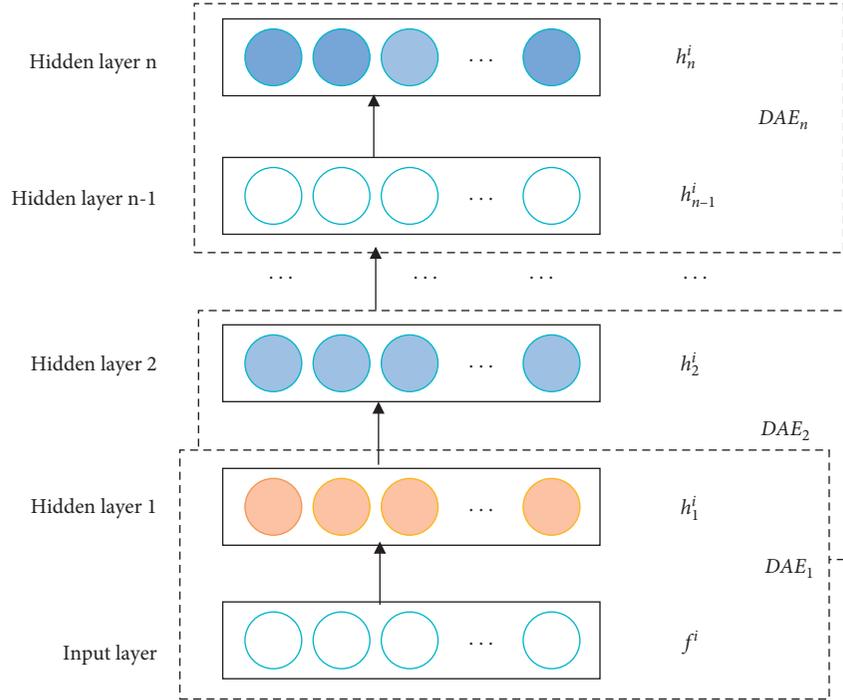


FIGURE 1: DAE model structure.

system. The lack of tag training data is a common problem. It is difficult to reconcile the contradiction between small fault samples and large sample training for deep learning. How to solve the contradiction between small sample fault state data in actual fault diagnosis and prediction and large training samples required by the deep network model is a difficult problem. In order to realize the rapid detection and early warning of faults, this study constructs a fault detection model based on the improved deep self-coding network based on the data in the normal operation state, which provides a solution for the lack of complete samples in the actual diagnosis.

## 2. Fault Early Warning Model Based on Improved DAE

**2.1. DAE Model.** The DAE network is formed by stacking multiple automatic encoders (AE) [38, 39] (shown in Figure 1). The input data in high-dimensional space are converted into the coding vector in low-dimensional space through the coding network, and the coding vector in low-dimensional space is reconstructed back to the original input data through the decoding network.

**2.2. Fault Early Warning Model.** When the equipment is in a normal state, the input and output of DAE fit very well, the original structure and stable relationship between variables have been maintained, and the reconstruction error of corresponding DAE is in dynamic and stable change. When the equipment is abnormal, it will cause the change of relevant variables, and the internal correlation of feature space will be destroyed which will be reflected in the error

trend of variables. The reconstruction error of the DAE model will deviate from the original stable state and increase accordingly. It means that the fault occurs. Therefore, the abnormal change of reconstruction error on monitoring parameters is used as the index to detect the fault in this study.

The reconstruction error of the variable is obtained from the difference between the reconstructed value and the actual value, as shown in the following formula:

$$R_e = (x' - x)^2, \quad (1)$$

where  $R_e$  is the reconstruction error of sample,  $x$  is the actual value, and  $x'$  is the reconstruction value. The threshold is set according to the distribution of reconstruction error of training samples, and the fault detection criteria are shown in the following formula:

$$\begin{cases} R_e > R_{th}, & \text{fault warning.} \\ R_e \leq R_{th}, & \text{normal state} \end{cases} \quad (2)$$

**2.3. The Improved Model.** CNN was originally used to realize image classification and played an important role in the field of image classification. The vibration signal is one-dimensional data. Some scholars choose to process a one-dimensional vibration signal into a time-frequency map and input it into CNN as a two-dimensional image to realize fault diagnosis, which has achieved certain success. However, the above operations make the data processing inconvenient. Therefore, the one-dimensional CNN (1d-CNN) [40, 41] is used to directly process the original vibration signal. With the increase in network depth, the training becomes more

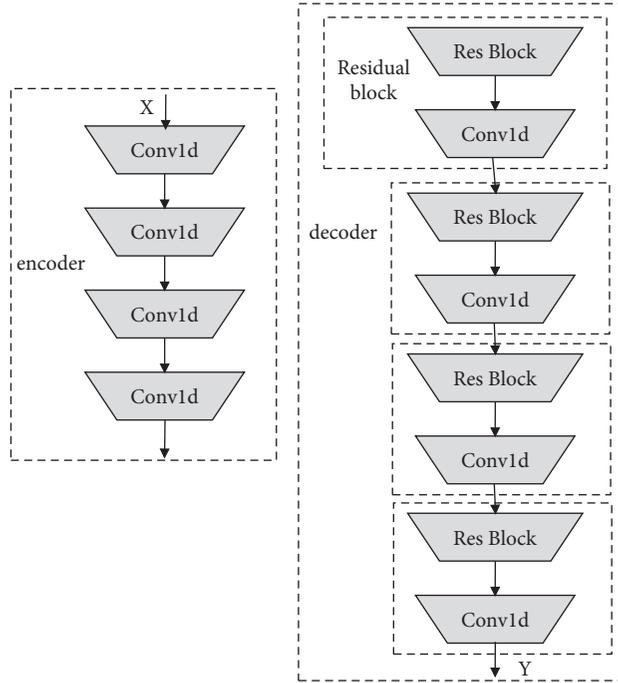


FIGURE 2: Improved AE.

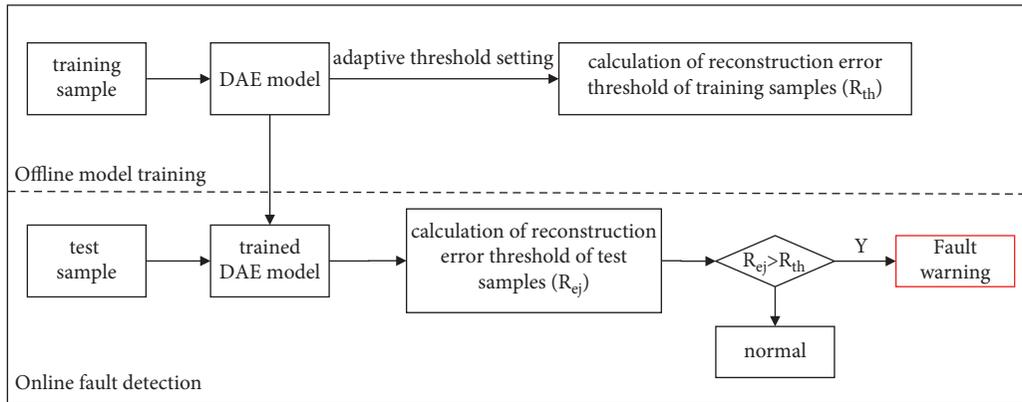


FIGURE 3: Fault early warning model based on DAE.

and more difficult. Some special weight initialization strategies and batch normalization (BN) methods have greatly improved this problem. However, when the model converges, the methods bring another problem: with the increase of network depth, the training error does not decrease, but increases. Residual network (ResNet) solves the problem of training difficulty caused by network depth, and the network performance (accuracy and accuracy) is far better than the traditional network model.

The original vibration signal is used as the input of 1d-CNN, and a self-coding network model is constructed, as shown in Figure 2.

### 3. The Process of Fault Early Warning Method

$R_e$  fluctuates within the threshold range under normal conditions. When  $R_e$  exceeds the threshold and remains above the threshold, it can be determined that the equipment

has failed. The fault prediction method based on DAE mainly includes offline model training, reconstruction error construction, and online fault detection, as shown in Figure 3.

**3.1. Offline Model Training.** The vibration signal in normal conditions collected by the sensor is normalized as samples  $\{x_i\}$  ( $i = 1, 2, \dots, n$ ). The samples are divided into training set and test set. The training of the DAE model is shown in Figure 4.

The calculation between hidden layers is shown in the following formula:

$$x_{i,j}^l = F \left( \sum_{i=1}^L \sum_{j=1}^m (x_{i,j}^{l-1} \times w_{i,j}^{l-1}) + b \right), \quad (3)$$

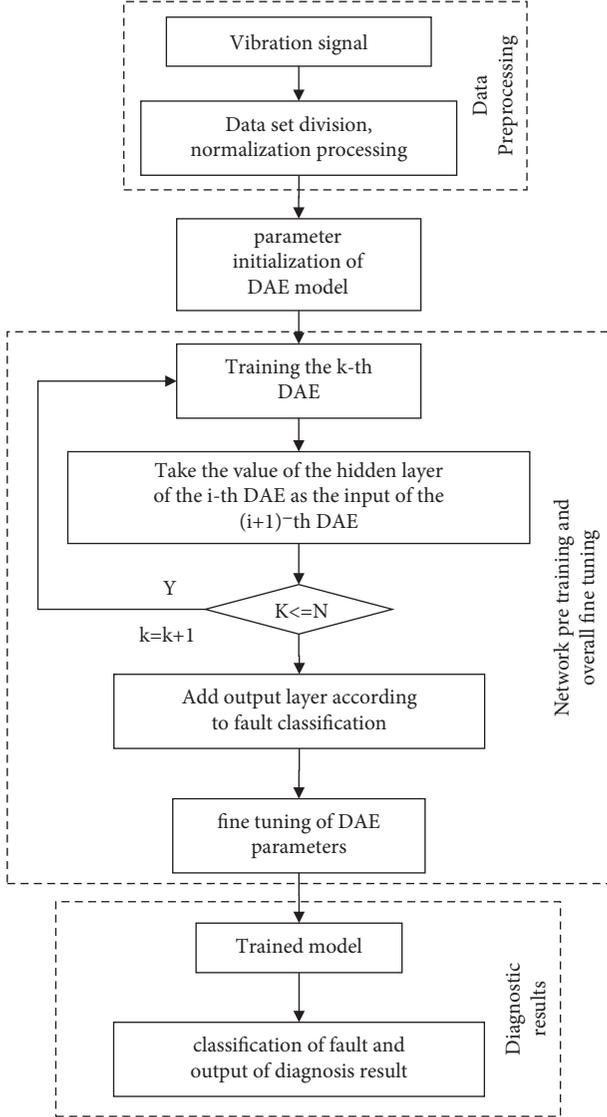


FIGURE 4: Training flow chart of DAE.

where  $x_{i,j}^l$  is the  $i$ -th characteristic value of the  $j$ -th feature vector value in the  $L$ -th layer of DAE,  $w_{i,j}^{l-1}$  is weight coefficient,  $L$  is the convolution kernel size,  $F$  is the activation function, and  $b$  is the deviation value.

Add an output layer after the last hidden layer, the output of DAE is shown in the following formula:

$$x'_i = F \left( \sum_{i=1}^L \sum_{j=1}^m (x_{i,j} \times w_{i,j}) + b \right), \quad (4)$$

where  $x'_i$  is the  $i$ -th output value of DAE,  $x_{i,j}$  is the  $i$ -th characteristic value of the  $j$ -th feature vector in the last hidden layer of DAE,  $w_{i,j}$  is weight coefficient,  $L$  is the convolution kernel size,  $F$  is the activation function, and  $b$  is the deviation value.

The error between output and expected value is calculated by the following formula:

$$R = \frac{1}{m} \sum_{i=1}^n (x_i - x'_i), \quad (5)$$

where  $x_i$  is the  $i$ -th true value of the sample, and  $x'_i$  is  $i$ -th output value of the DAE.

$\theta$  updates with the following formula:

$$\theta = \theta - \alpha \frac{\partial R}{\partial \theta}, \quad (6)$$

where  $\theta = \{W_1, W_2, \dots, W_{n-1}, W, b_1, b_2, \dots, b_{m-1}, b\}$ , and  $\alpha$  is the parameter learning rate.

**3.2. Online Fault Detection.** The vibration signal  $f$  of the equipment is collected in real time by the acquisition device, and the samples are normalized to obtain the data  $\{x_i\}$ .

Input  $\{x_i\}$  into the trained DAE model to obtain the corresponding reconstruction error  $\{R_c^i\}$ , and  $\{R_c^i\}$  is compared with evaluating the equipment status.

#### 4. Determination of Adaptive Threshold

In the fault prediction, it is commonly used to set the fixed threshold of the reconstruction error. However, due to the influence of environment, modeling error, and other factors, the reconstruction error in a normal state may fluctuate. So, if a fixed threshold is set for state evaluation, it is easy to cause a false alarm. Therefore, the adaptive reconstruction error is set for fault detection and early warning in this study.

According to the principle of statistics, the mean and variance of residuals are calculated as follows:

$$\mu(R_j, t_k) = \frac{1}{n} \sum_{i=1}^n r_i(t_k) | R_j, \quad (7)$$

$$\sigma^2(R_j, t_k) = \frac{1}{n-1} \sum_{i=1}^n (r_i(t_k) - \mu(R_j, t_k))^2 | \mu_j,$$

where  $R_j$  is the reconstruction error corresponding to different times.

The confidence interval of the mean value (the confidence is  $1 - \alpha$ ) can be expressed as the following formula:

$$P\{\bar{\mu} - z\alpha < \mu < \bar{\mu} + z\alpha\} = 1 - \alpha, \quad (8)$$

where  $\alpha$  is the confidence level, and  $z$  is the correlation coefficient of the confidence level. In practical application, the confidence is usually set to be 95%–99%.

If the confidence is 95%, then  $z$  is 1.96, and the threshold can be obtained from the following formula:

$$R_{thl} = \mu(R_j, t_k) \pm 1.96\sigma^2(R_j, t_k). \quad (9)$$

Take  $R_{thl}$  as the threshold. When the residual exceeds this threshold and remains above the threshold, it indicates that the system is abnormal and in the early stage of fault.

#### 5. Experiment and Analysis

**5.1. Experimental Equipment and Data.** As shown in Figure 5, the data are collected from the two-stage gearbox.

The gear speed change is controlled by a motor. The torque is provided by the magnetic brake and can be adjusted by changing the input voltage.

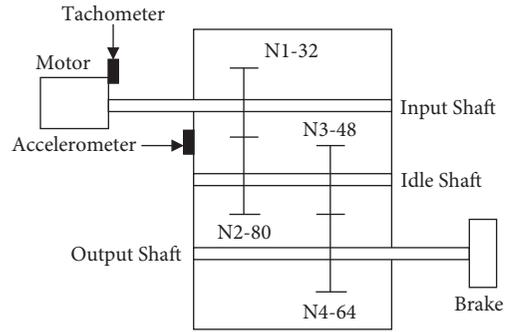


FIGURE 5: Gearbox experimental system.

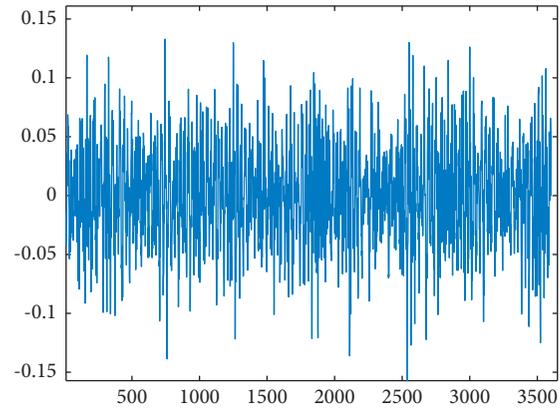
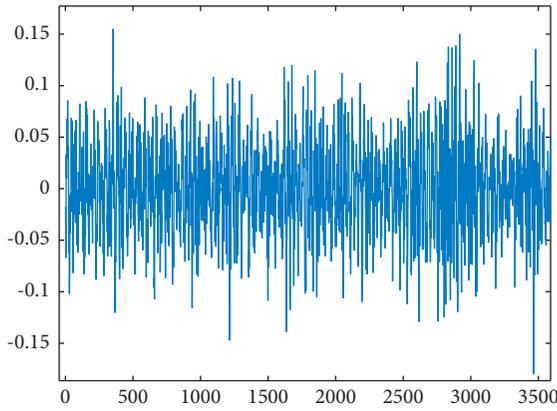


FIGURE 6: Normal sample.

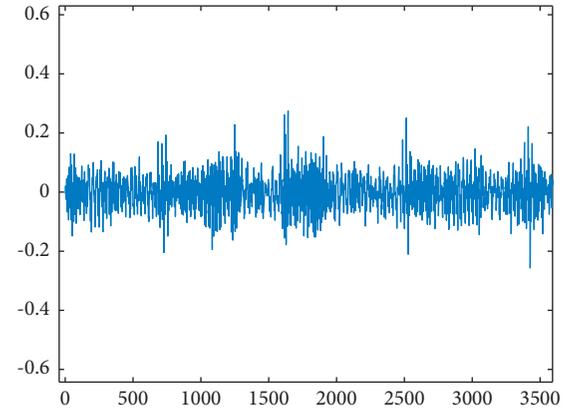
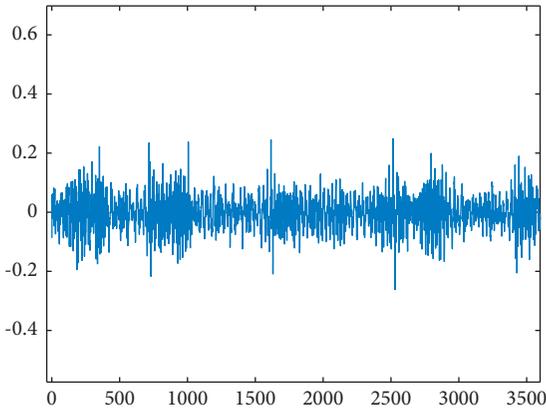


FIGURE 7: Fault sample.

The pinion with 32 teeth and 80 teeth is installed on the first-stage input shaft. The second stage consists of a 48 teeth pinion and a 64 teeth pinion. The input shaft speed is measured by the tachometer, and the gear vibration signal is measured by the accelerometer. The vibration signals of normal and fault samples used in this study are shown in Figure 6 and Figure 7.

5.2. *Model Parameter of Improved DAE.* A seven-layer DAE model is constructed in this study. The network structure parameter is shown in Table 1, and the parameter of the model is shown in Table 2.

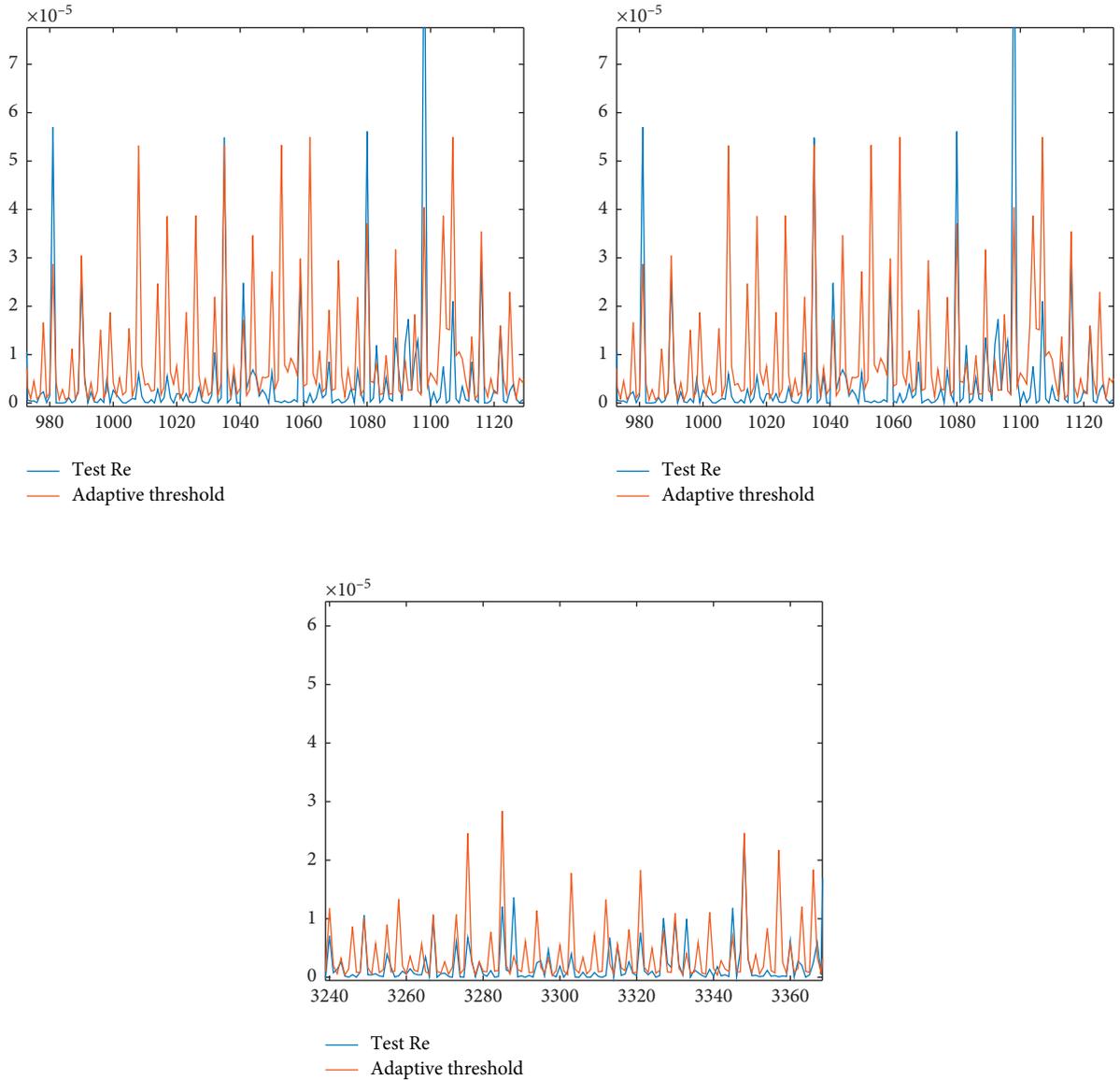
TABLE 1: Network structure parameter.

Network layer	Combination form of node
7	3600-1200-400-134-400-1200-3600

5.3. *DAE Model Test of Gearbox under Normal Condition.* After the model is trained, the data of the gearbox under normal state are tested. The adaptive threshold is calculated by (9). The change trend of  $R_e$  obtained from DAE is shown

TABLE 2: Parameter of the model.

Layer	Structure name	Structure parameter	Number of channels	Output size
1	Input	(3600,1)	1	(3600,1)
2	Convolution layer	(1,1)	5	(3600,1)
3	Convolution layer	(3,1)	4	(1200,1)
4	Convolution layer	(3,1)	8	(400,1)
5	Convolution layer	(3,1)	4	(134,1)
6	Convolution layer	(3,1)	8	(400,1)
7	Convolution layer	(3,1)	4	(1200,1)
8	Convolution layer	(1,1)	5	(3600,1)
9	Convolution layer	(3600,1)	1	(3600,1)

FIGURE 8:  $R_e$  within the adaptive threshold.

in Figure 8. It can be seen that  $R_e$  of the gearbox is always within the adaptive threshold range under normal conditions.

**5.4. DAE Model Test of Gearbox under Normal Condition.** The data before and after the fault are used to verify the fault detection effect of the model. When the model is used in the

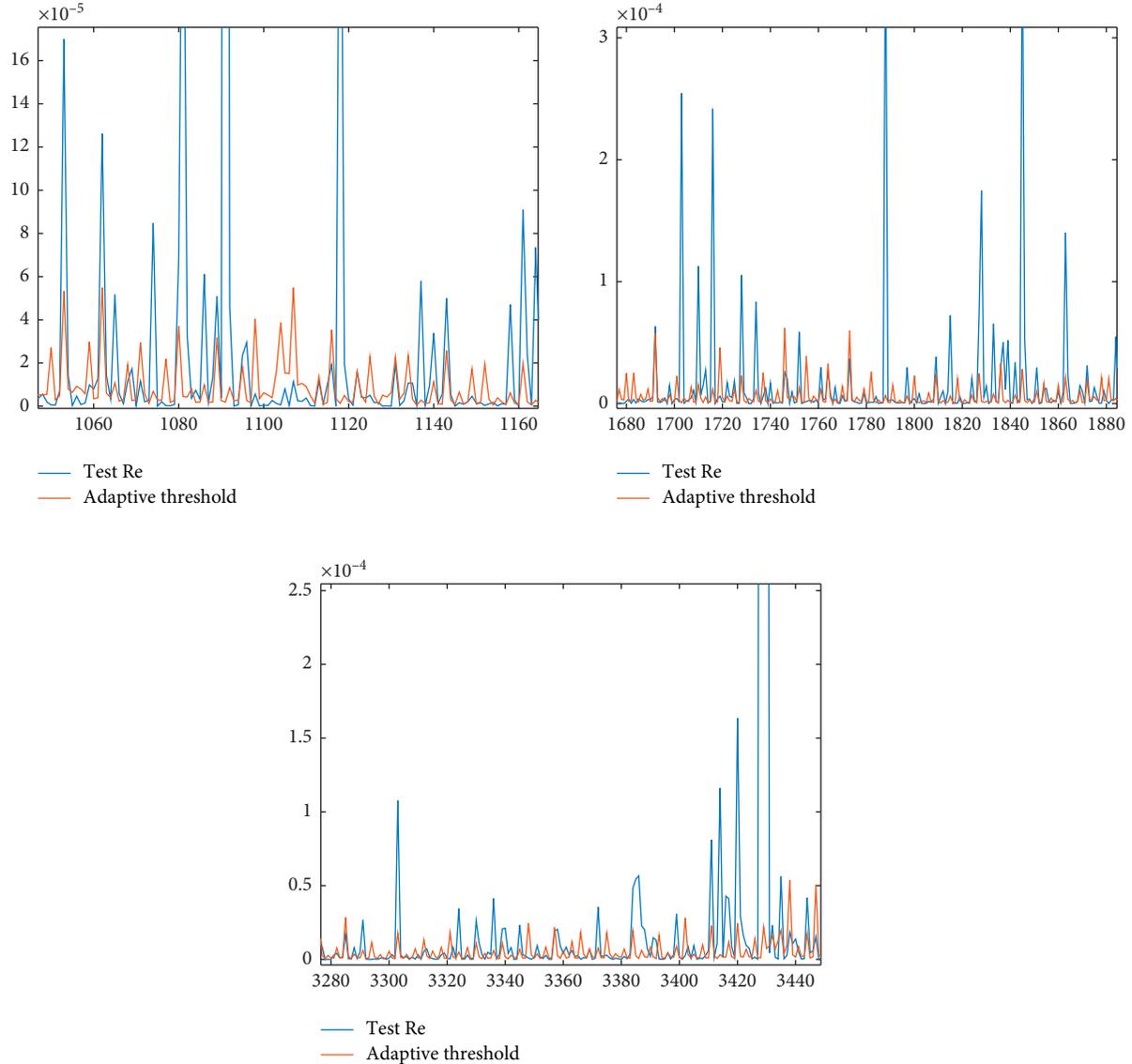


FIGURE 9:  $R_e$  exceeded the adaptive threshold.

gearbox fault state, calculate  $R_e$ . The trend of  $R_e$  is shown in Figure 9.

It can be seen that  $R_e$  exceeds the threshold at 2200, and the value is basically above the threshold. The gearbox fault can be determined, and the early warning of fault can be carried out.

### 6. Conclusion

This study presents a fault detection method based on normal operation data, which solves the contradiction between small sample fault data and large training samples required by the deep network model. The main conclusions are as follows:

- (1) An improved DAE network is designed, which directly uses the original vibration data for fault detection, and the data in the normal state are used for

unsupervised pretraining and supervised fine-tuning of the network. In this method, the reconstruction error is selected as the fault state parameter. In order to reduce the misjudgment caused by environmental factors, the adaptive threshold of reconstruction error is introduced as the decision-making basis of fault early warning.

- (2) The experimental results show that under the premise of small data sets, this model has a very high accuracy and can quickly realize fault detection and early warning.
- (3) This study constructs an improved self-encoder prediction model mainly for small data sets of gearboxes, but in practice, the factors affecting the high-precision prediction of gears are extremely complex. With the continuous development of deep learning algorithm research, combined with gearbox

monitoring data, advanced deep learning algorithm and comprehensive high-precision prediction model of multiple environmental factors outside the gearbox will be the next focus of this study.

## Data Availability

The data set used in this article can be obtained from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

## Acknowledgments

This work was supported by the Innovation ability improvement project of scientific and technological small and medium-sized enterprises in Shandong Province (Grant no. 2021TSGC1089), Major scientific and technological innovation projects in Shandong Province (Grant no. 2019JZZY010117), and 2020 industrial Internet innovation and development project - solution application and promotion public service platform(Grant no. TC200802C).

## References

- [1] X. Chen, *Intelligent Maintenance and Health Management*, China Machine Press, Beijing, China, 2018.
- [2] Y. Lei, F. Jia, and X. Zhou, "A deep learning-based method for machinery health monitoring with big data," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 49–56, 2015.
- [3] Y. Lei, F. Jia, and D. Kong, "Opportunities and challenges of machinery intelligent fault diagnosis in big data era," *Journal of Mechanical Engineering*, pp. 1–12, 2017, <http://kns.cnki.net/kcms/detail/11.2187.TH.20170822.1648.032.html>.
- [4] X. Jin, Y. Wang, and B. Zhang, "Industrial big data-driven fault prognostics and health management," *Computer Integrated Manufacturing Systems*, pp. 1–27, 2021, <http://kns.cnki.net/kcms/detail/11.5946.TP.20200814.1703.006.html>.
- [5] H. Hao, M. Wang, Y. Tang, and Q. Li, "Research on data fusion of multi-sensors based on fuzzy preference relations," *Neural Computing & Applications*, vol. 31, no. S1, pp. 337–346, 2018.
- [6] S. Li, L. Jiang, Q. Zhang, Z. Wang, Z. Tian, and M. Guizani, "A malicious mining code detection method based on multi-features fusion," *IEEE Transactions on Network Science and Engineering*, vol. 2022, Article ID 3155187, 2022.
- [7] P. Yu and J. Cao, "Deep learning approach and its application in fault diagnosis and prognosis," *Computer Engineering and Applications* 2020, vol. 56, no. 3, pp. 1–18, 2021.
- [8] K. Yu, L. Jia, Y. Chen, and W. Xu, "Deep learning: yesterday, today, and tomorrow," *Journal of Computer Research and Development*, vol. 50, no. 9, pp. 1799–1804, 2013.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 60, pp. 1097–1105, 2012.
- [10] S. Hoermann, P. Henzler, M. Bach, and D. Klaus, "Object detection on dynamic occupancy grid maps using deep learning and automatic label generation," 2018.
- [11] H. Yang, L. Luo, L. P. Chueng, L. David, and C. Francis, "Deep learning and its applications to natural language processing," *Deep Learning: Fundamentals, Theory and Applications*, Springer, New York, NY, USA, 2019.
- [12] S. Li, Y. Li, W. Han, X. Du, M. Guizani, and Z. Tian, "Malicious mining code detection based on ensemble learning in cloud computing environment," *Simulation Modelling Practice and Theory*, vol. 113, Article ID 102391, 2021.
- [13] D. T. Hoang and H. J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.
- [14] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing & Applications*, vol. 32, pp. 1–14, 2019.
- [15] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," *Nature Communications*, vol. 5, no. 1, pp. 1–9, 2014.
- [16] H. Pan, X. He, and S. Tang, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *Journal of Mechanical Engineering*, vol. 64, no. 7/8, pp. 443–452, 2018.
- [17] J. Latif, S. Tu, C. Xiao, S. U. Rehman, M. Sadiq, and M. Farhan, "Digital forensics use case for glaucoma detection using transfer learning based on deep convolutional neural networks," *Security and Communication Networks*, vol. 202113 pages, Article ID 4494447, 2021.
- [18] H. Ren, J. Qu, C. Yi, and T. Qiu, "Deep learning for fault diagnosis: the state of the art and challenge," *Control and Decision*, vol. 32, no. 8, pp. 1345–1358, 2017.
- [19] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.
- [20] T. de Bruin, K. Verbert, and R. Babuska, "Railway track circuit fault diagnosis using recurrent neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 523–533, 2017.
- [21] W. Elmasry, A. Akbulut, and A. H. Zaim, "Deep learning approaches for predictive masquerade detection," *Security and Communication Networks*, vol. 2018, Article ID 327215, 24 pages, 2018.
- [22] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [23] M. Alawad and M. Lin, "Stochastic-based deep convolutional networks with reconfigurable logic fabric," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 4, pp. 242–256, 2016.
- [24] H. Ren, J. F. Qu, Y. Chai, and T. Qiu, "Deep learning for fault diagnosis: the state of the art and challenge," *Control and Decision*, vol. 32, no. 8, pp. 1345–1358, 2017.
- [25] W. Jiang, C. Wang, J. Zou, and S. Zhang, "Application of deep learning in fault diagnosis of rotating machinery," *Processes*, vol. 9, no. 6, pp. 919–922, 2021.
- [26] R. M. Souza, E. Nascimento, U. A. Miranda, W. Silva, and A. L. Herman, "Deep learning for diagnosis and classification of faults in industrial rotating machinery," *Computers & Industrial Engineering*, vol. 153, Article ID 107060, 2020.
- [27] Y. Yang, M. Menul Haque, D. Bai, and T. Wei, "fault diagnosis of electric motors using deep learning algorithms and its application: a review," *Energies*, vol. 14, no. 21, pp. 2–26, 2021.
- [28] C. Li, J. Xiong, X. Zhu, Q. Zhang, and S. Wang, "fault diagnosis method based on encoding time series and

- convolutional neural network,” *IEEE Access*, vol. 8, Article ID 165232, 2020.
- [29] S. Li, Q. Zhang, X. Wu, W. Han, and Z. Tian, “Attribution classification method of APT malware in IoT using machine learning techniques,” *Security and Communication Networks*, vol. 2021, Article ID 9396141, 12 pages, 2021.
- [30] L. V. Utkin, V. S. Zaborovskii, and S. G. Popov, “Detection of anomalous behavior in a robot system based on deep learning elements,” *Automatic Control and Computer Sciences*, vol. 50, no. 8, pp. 726–733, 2016.
- [31] H. Lee, Y. Kim, and C. O. Kim, “A deep learning model for robust wafer fault monitoring with sensor measurement noise,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 1, pp. 23–31, 2017.
- [32] C. Lu, Z. Wang, W.-L. Qin, and J. Ma, “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification,” *Signal Processing*, vol. 130, pp. 377–388, 2017.
- [33] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, “Unsupervised electric motor fault detection by using deep autoencoders,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, pp. 441–451, 2019.
- [34] V. H. Nguyen, J. S. Cheng, Y. Yu, and V. T. Thai, “An architecture of deep learning network based on ensemble empirical mode decomposition in precise identification of bearing vibration signal,” *Journal of Mechanical Science and Technology*, vol. 33, no. 1, pp. 41–50, 2019.
- [35] Y. Zhang, X. Li, L. Gao, and P. Li, “A new subset based deep feature learning method for intelligent fault diagnosis of bearing,” *Expert Systems with Applications*, vol. 110, pp. 125–142, 2018.
- [36] S. Liu, Y. Liu, Y. Gu, and X. Xu, “Method of extracting gear fault feature based on stacked autoencoder,” *The Journal of Engineering*, vol. 2019, no. 23, 8769 pages, 2019.
- [37] F. Yu, J. Liu, D. Liu, and H. Wang, “Upervised convolutional autoencoder-based fault-relevant feature learning for fault diagnosis in industrial processes,” *Journal of the Taiwan Institute of Chemical Engineers*, vol. 132, Article ID 104200, 2022.
- [38] Z. Zhao, L. Li, S. Yang, and Q. Li, “A frequency domain feature extraction auto-encoder and its applications on fault diagnosis,” *CHINA MECHANICAL ENGINEERING*, vol. 32, no. 20, pp. 2468–2474, 2021.
- [39] H. Shao, H. Jiang, H. Zhao, and F. Wang, “A novel deep autoencoder feature learning method for rotating machinery fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 95, pp. 187–204, 2017.
- [40] V. Tra, J. Kim, S. A. Khan, and J. M. Kim, “Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal levenberg-marquardt algorithm,” *Sensors*, vol. 17, no. 12, p. 2834, 2017.
- [41] 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, Article ID 10278, 2019.