

Research Article

Intelligent Diagnosis of Subway Traction Motor Bearing Fault Based on Improved Stacked Denoising Autoencoder

Yanwei Xu ^{1,2}, Chen Li,¹ and Tancheng Xie^{1,2}

¹School of Mechatronics Engineering, Henan University of Science and Technology, Luoyang 471003, China

²Intelligent Numerical Control Equipment Engineering Laboratory of Henan Province, Luoyang 471003, China

Correspondence should be addressed to Yanwei Xu; xuyanweiluoyang@163.com

Received 22 October 2020; Revised 30 November 2020; Accepted 30 December 2020; Published 12 January 2021

Academic Editor: Changqing Shen

Copyright © 2021 Yanwei Xu 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.

Aiming at the problem that the complex working conditions affect the effect of manual feature extraction in bearing fault diagnosis of metro traction motor, a fault diagnosis method of metro traction motor bearing based on improved stacked denoising autoencoder (SDAE) is proposed. This method extracts fault features directly from the original vibration signal through deep learning, reduces the dependence on signal processing technology and diagnosis experience, and solves the problem of unsatisfactory effect of extracting feature values under complex working conditions. The effect of the improved SDAE network structure on the accuracy of bearing fault diagnosis is studied through experiments, and the best network parameters are selected. The test results show that the proposed method can well extract the deep features of the fault under the condition of variable speed and variable load; when using data sets with complex working conditions, the classification accuracy of the proposed method is better than that of many traditional fault diagnosis methods.

1. Introduction

The frequent start and stop of the subway make the bearing of the traction motor continuously bear the changes of speed and load, and the inner and outer rings and rolling elements of the bearing are prone to cracks and pitting corrosion, which cause certain safety risks to the smooth operation of the subway [1, 2]. Therefore, intelligent diagnosis of subway traction motor bearings, timely identification of early failures of bearings [3], and early maintenance and analysis of the causes of failures are of great significance to the normal operation of the subway [4].

The fault diagnosis of bearings has always relied on the signal processing technology of diagnostic experts, and the signal processing method also affects the quality of feature value extraction [5–7]. For example, Chen et al. [8] use empirical wavelet transform (EWT) to decompose the signal into single components under an orthogonal basis, extract the inherent modulation information of the signal, and diagnose bearing faults. Li et al. [9] used a bearing vibration feature extraction method based on multiscale displacement entropy (MPE) and an improved support vector machine

binary tree (ISVM-BT) to diagnose bearing faults. Zvokelj and Zupan [10] conducted bearing fault diagnosis based on EEMD multiscale independent component analysis. The above personnel manually extract features based on relevant signal processing technology and a large amount of prior experience [11], but these manually extracted features are extracted based on specific diagnostic problems and may not be suitable for other problems, and the effect is general under variable working conditions. In addition, the lack of experience of the diagnosis expert will also omit important information in the original signal in the process of extracting the feature value [12], which will have a certain impact on the result of the fault diagnosis [13].

With the development of computer technology and the advent of the era of electromechanical big data [14], deep learning technology has begun to be applied to the field of fault diagnosis [15], solving many problems in fault diagnosis [16]. Wen et al. [17] proposed a LeNet-5-based CNN for fault diagnosis, which adaptively extracts fault features by converting the signal into a two-dimensional image. Guo et al. [18] proposed a hierarchical learning rate-adaptive deep convolutional neural network based on an improved

algorithm and studied its application in bearing fault diagnosis and severity determination. Shao et al. [19] proposed an adaptive deep confidence network (DBN) with a dual-tree complex wavelet packet (DTCWPT), which is applied to the fault diagnosis of rolling bearings. These methods all use deep learning technology for adaptive extraction of fault features and still have good feature extraction effects under complex working conditions. Deep learning is mainly divided into four types: deep autoencoder network, convolutional neural network, recurrent neural network, and deep confidence network. Among them, the autoencoder network is an unsupervised learning algorithm, which can compress the input information and extract the most representative information in the data. It has an incomparable advantage in feature adaptive extraction [20].

Based on this, an intelligent diagnosis network for subway traction motor faults based on improved SDAE is proposed. The BN layer normalization and L2 regularization are added to the network, and the Leaky ReLU activation function is introduced. The improved SDAE is used for adaptive extraction of fault features, and the data sets of variable speed and variable load of subway traction motor bearings are tested to verify the effectiveness of the proposed method.

2. Algorithm Description

2.1. Autoencoder. Autoencoder (AE) is a typical unsupervised algorithm, which is often used for data compression, feature extraction, and other aspects. It is mainly composed of input layer, hidden layer, and output layer. The input layer and output layer are symmetrical in structure and have the same number of nodes. The encoding process is formed between the input layer and the hidden layer, and the decoding process is formed between the hidden layer and the output layer, as shown in Figure 1.

$x_i (i \in 1, 2, 3, \dots, n)$ is the input data, $h_j (j \in 1, 2, 3, \dots, m)$ is the feature data of the middle hidden layer, and $\tilde{x}_i (i \in 1, 2, 3, \dots, n)$ is the output data. The encoding process is that the input vector x is mapped to the feature vector h of the hidden layer through the activation function f_q . The formula is as follows:

$$H = f_q(W_q X + b_q). \quad (1)$$

The decoding process is that the feature vector h is mapped to the feature vector of the output layer through the activation function f_g . The formula is as follows:

$$X = f_g(W_g H + b_g). \quad (2)$$

Among them,

$$\begin{aligned} X &= x_1, x_2, x_3, \dots, x_n, \\ H &= h_1, h_2, h_3, \dots, h_m, \\ X &= x_1, x_2, x_3, \dots, x_n. \end{aligned} \quad (3)$$

X , H , and X are the data of the input layer, the hidden layer, and the output layer, respectively; W_q and W_g are the weight matrices of the encoding process and the decoding

process, respectively; b_q and b_g are the offset parameters of the encoding process and the decoding process; f_q and f_g are the activation functions of the encoding process and the decoding process. Commonly used activation functions are sigmoid, tanh, ReLU, and so on, but sigmoid and tanh are prone to gradient dispersion during training, and the ReLU activation function is fast in calculation and can better avoid gradient dispersion problems. The function image of ReLU is shown in Figure 2, and the formula is as follows:

$$f(x) = \max(0, x). \quad (4)$$

As can be seen from the figure, when the input value is negative, the output value of the activation function is 0, which results in the failure of updating parameters of the neurons. To avoid this problem, Leaky ReLU activation function was introduced, and its function image is as shown in Figure 3, and the formula is as follows:

$$f(x) = \begin{cases} x, & x > 0, \\ ax, & x \leq 0. \end{cases} \quad (5)$$

2.2. Build Stacked Denoising Autoencoder Network.

Stacked autoencoder (SAE) is a typical unsupervised network model in deep learning. It is composed of multiple autoencoders, and the hidden layer of the former is used as the input layer of the latter. Compared with the structure of a single AE, the stacked network structure has a deeper network structure and a more complex nonlinear mapping relationship, which can extract more abstract expression characteristics from the original data.

Due to the instability of the original signal, in order to improve the robustness of AE, noise is randomly added to the original data X to obtain the contaminated data \tilde{x} , which is used as the input of the autoencoder, and at the same time make the output of the autoencoder as close as possible to the uncontaminated original signal x . Its network structure is shown in Figure 4.

The structure of the improved SDAE is shown in Figure 5. As a basic unit, the denoising AE is placed in the first layer of SDAE. After that, multiple AE are stacked [21], and Softmax classifier is added in the last layer to classify the features extracted in depth. The training flowchart of each AE is shown in Figure 6. Among them, the normalization of the BN layer plays a role in accelerating training and reducing the gradient dispersion and does not change the distribution of the original data. L2 regularization reduces overfitting, and a model with better generalization ability can be obtained, which is suitable for fault diagnosis under variable operating conditions.

3. Construction of Fault Diagnosis Network for Metro Traction Motor Bearing

A three-layer SDAE network structure (3750-128-64-6) is designed for fault diagnosis of subway traction motors, including a double hidden layer structure and a Softmax classifier. Among them, 3750 is the number of input layer nodes, 128 is

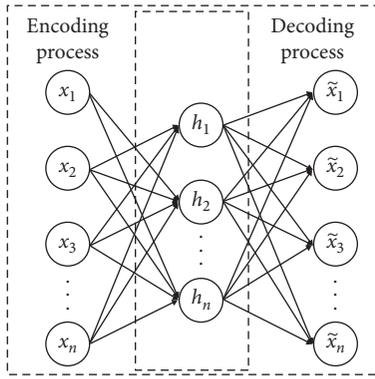


FIGURE 1: Autoencoder.

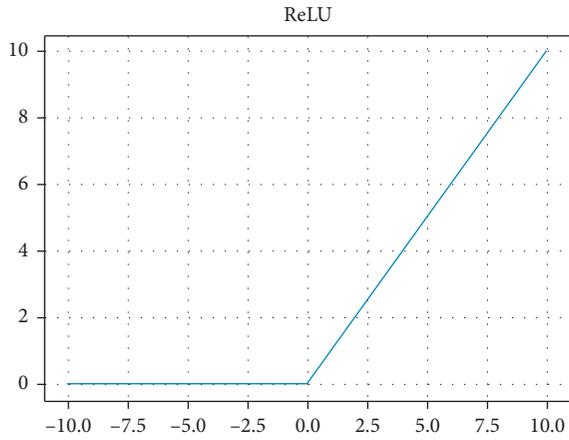


FIGURE 2: ReLU activation function.

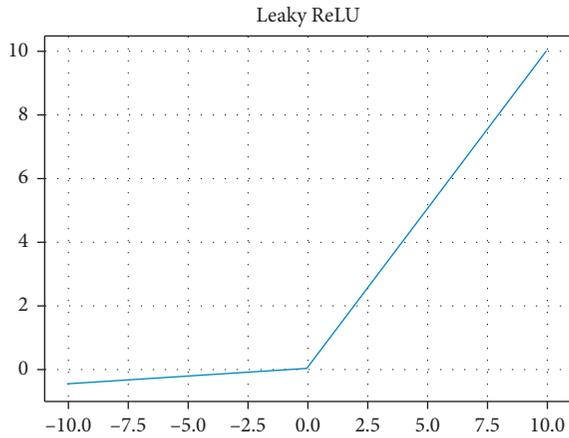


FIGURE 3: Leaky ReLU activation function.

the number of hidden layer nodes in the first layer, 64 is the number of hidden layer nodes in the second layer, and 6 is the number of nodes in the final output layer. The samples input to the network undergo dimensionality reduction, fusion, and nonlinear transformation, compressed and fused by the Softmax classifier and transformed into a 6-dimensional output.

The flowchart of metro traction motor fault diagnosis network based on improved stack autoencoder is shown in Figure 7.

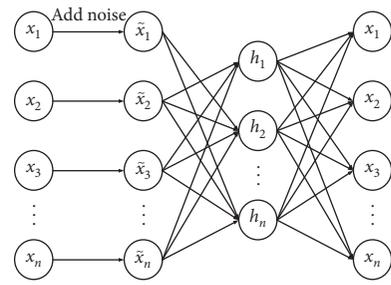


FIGURE 4: Denoising autoencoder.

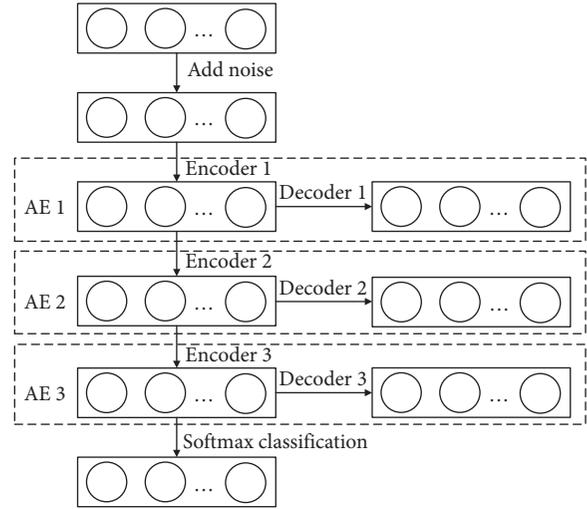


FIGURE 5: Stacked denoising autoencoder network structure.

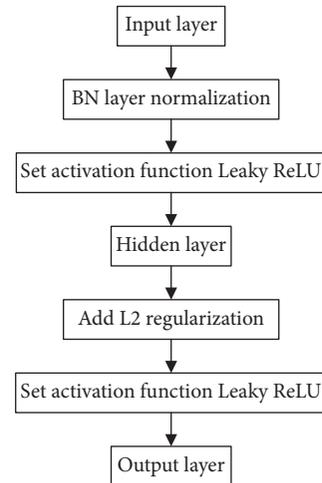


FIGURE 6: AE training flowchart.

4. Bearing Test Analysis and Information Acquisition System of Metro Traction Motor

4.1. *Information Collection System.* The subway traction motor bearing information collection system is a rolling bearing experimental platform jointly developed by Henan University of Science and Technology, Luoyang Bearing Research Institute, and Henan Engineering Laboratory of

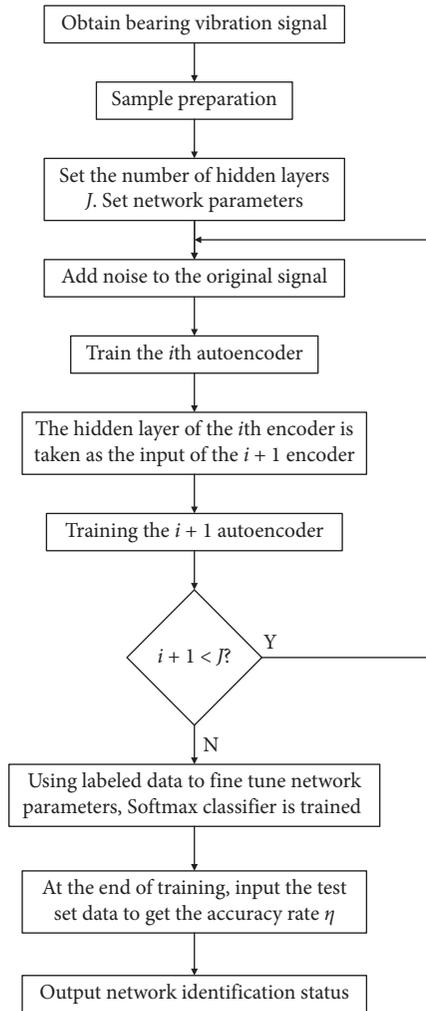


FIGURE 7: Improved SDAE network structure.

Intelligent CNC Equipment, as shown in Figure 8. The information acquisition system consists of a test bench, subway traction motor bearings, hydraulic loading system, sensor information acquisition module, and computer.

Among them, the sensor information acquisition module is mainly composed of a vibration sensor and a data acquisition card. The sensor adopts the LC0151T high-precision vibration acceleration sensor, which has a range of 33 g and a sensitivity of 150 mv/g. It also uses a PCI8510 data acquisition card that can collect 8 channels of data at the same time, which can well meet the needs of experimental data collection. In the process of data collection, the signal collected by the vibration sensor is a current signal; after a signal conditioner, the current signal is converted into a voltage signal and the computer collects and stores the signal through the data acquisition card. The acquisition system is shown in Figure 9.

4.2. Test Parameters and Scheme. Nu216 metro traction motor bearing is used as the test bearing. In order to test and verify, the defects on the inner ring, outer ring, and rolling element of the bearing are prefabricated, and the simulated

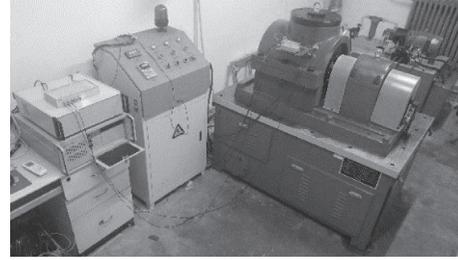


FIGURE 8: Metro traction motor bearing information acquisition system.

faults are processed by laser marking machine, which are inner ring pitting, inner ring cracking, outer ring pitting, outer ring cracking, rolling element pitting, and rolling element cracking. The crack width is 30 μm , the pitting diameter is 40 μm , and the depth is 30% of laser energy.

Due to the frequent start and stop of the subway, the bearing of the traction motor is constantly subjected to changes in speed and load. The variables of the test are selected as speed, load, and bearing failure. The experiment collects data under 54 ($3 \times 3 \times 6$) working conditions and collects 4 signals for each working condition. The test arrangement is shown in Table 1.

4.3. Analysis of Test Data. The sampling frequency of the test is 50 kHz, the speed of bearing is 800 rpm, 1600 rpm, and 2400 rpm, and the load of bearing is 5 kN, 7 kN, and 9 kN. The vibration data of the bearing under the working conditions of 2400 rpm and 7 kN are selected to construct the data set s to preliminarily analyze the deep learning network and carry out the intelligent fault diagnosis of the bearing. The fault details are shown in Table 2.

4.4. Sample Preparation and Labeling. In the experiment, the sampling frequency is 50 kHz, and the sampling time of each sample is 10 s. Therefore, each sample has 500000 data points, which needs to be segmented to obtain a large number of training data. According to Shannon's sampling theorem, the sampling period is at least two revolutions of the bearing; considering the influence of rotation speed, every 7500 data points are finally selected as a group of small samples. Based on this, the original data are processed as follows:

- (1) The 500000 data points are divided into two parts: 425000 data points and 75000 data points.
- (2) The first 425000 data points are used as a training set to segment data, and the data are overlapped to achieve the purpose of data amplification. 425000 data points are divided into 100 groups of small sample data, and each group of small sample data contains 7500 data points, as shown in Figure 10.
- (3) The last 75000 data points need to be used as the test set, without overlapping sampling, and directly truncated. Ten sets of test set samples with a data length of 7500 are obtained.

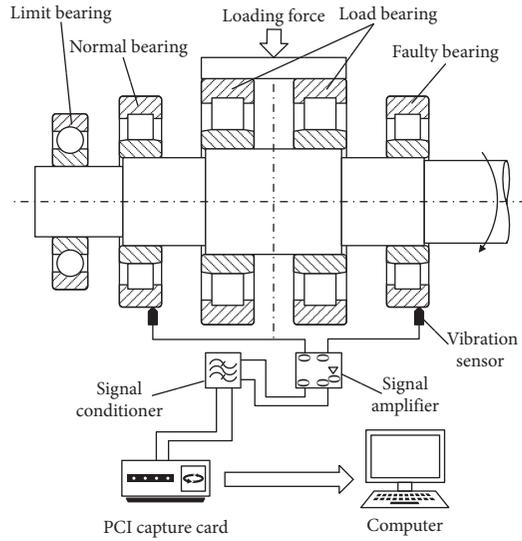


FIGURE 9: Schematic diagram of the acquisition system.

TABLE 1: Test schedule.

Speed (r/min)	Inner ring (kN)		Outer ring (kN)		Rolling element (kN)	
	Pitting	Crackle	Pitting	Crackle	Pitting	Crackle
800	5	5	5	5	5	5
800	7	7	7	7	7	7
800	9	9	9	9	9	9
1600	5	5	5	5	5	5
1600	7	7	7	7	7	7
1600	9	9	9	9	9	9
2400	5	5	5	5	5	5
2400	7	7	7	7	7	7
2400	9	9	9	9	9	9

TABLE 2: Fault details.

No.	Fault type	Speed (rpm)	Load (kN)
0	Rolling element pitting	2400	7
1	Rolling element crack	2400	7
2	Pitting of inner ring	2400	7
3	Inner ring crack	2400	7
4	Pitting of outer ring	2400	7
5	Outer ring crack	2400	7

Taking data set s as an example, the number of training set samples and test set samples can be obtained after the original data is processed in the above sections, as shown in Table 3. Each type of fault is labeled according to the number in Table 3, and it is divided into 6 types of faults.

Before inputting the data to SDAE, it is necessary to process the data to meet the input format of the deep neural network. After the experiment, all the vibration data are finally selected to do FFT, and the frequency domain data of the signal are input into the deep neural network for training and testing. Each type of fault is labeled according to the number in Table 3, which is divided into six fault types.

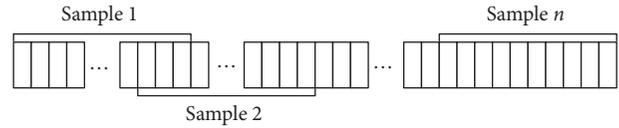


FIGURE 10: Data segmentation.

TABLE 3: Description of test data set.

No.	Fault type	Sample number of training set	Sample number of test set
0	Rolling element pitting	400	40
1	Rolling element crack	400	40
2	Pitting of inner ring	400	40
3	Inner ring crack	400	40
4	Pitting of outer ring	400	40
5	Outer ring crack	400	40
Total		2400	240

4.5. Influence of SDAE Network Structure on Bearing Fault Diagnosis Results. SDAE is stacked by multiple autoencoders. The number of nodes, number of iterations, learning rate, and even batch training number of each autoencoder will have a certain impact on the results of fault diagnosis. Particularly, the number of layers and nodes has a great influence on the accuracy of diagnosis results. In order to find the appropriate network structure, the number of layers and nodes of the network is studied under the condition of other parameters unchanged (learning rate is 0.001, batch training number is 256, and iterations are 20 times). In this paper, the influence of the following network structure on the diagnosis results under the data set s is studied. The average value of 10 diagnostic results is taken as the accuracy rate, as shown in Table 4.

It can be concluded from Table 4 that the number of network layers and nodes per layer has a certain impact on the diagnosis results, but it is not that the more the network layers and nodes, the higher the accuracy rate. The accuracy rate of the test set of three kinds of network structure is 100%, which are 3750-128-64-6, 3750-256-128-6, and 3750-256-128-64-6. Because of the more layers and nodes of the network and because the amount of calculation is greater and the training time is longer, the network structure of 3750-128-64-6 is selected.

4.6. Visualization of Fault Characteristics. In order to reflect the advantages of SDAE in feature extraction, this paper chose the t-SNE algorithm to visualize the data and perform cluster analysis and visualization on the original data and the extracted features layer by layer, as shown in Figure 11. The same fault types are marked according to the number in Table 3, and the numbers 0-5 and different colors are used to distinguish. It can be clearly seen that the original data is irregular and confused with each other. After the first autoencoder extracts features, the data of the

TABLE 4: SDAE accuracy under different network structures.

Network layers	Network structure	Test set accuracy rate (%)
3	3750-64-6	99.78
3	3750-128-6	99.91
4	3750-128-64-6	100
4	3750-256-128-6	100
4	3750-256-64-6	99.96
5	3750-512-256-128-6	98.89
5	3750-256-128-64-6	100
5	3750-512-128-64-6	99.24

same category are close to each other, and the class spacing of the data of the same category will be further reduced when features are extracted by an autoencoder, which fully shows that the proposed SDAE has a good effect on feature extraction.

5. Analysis and Comparison of Test Results

The above tests all adjust the SDAE network on the data set collected under the stable working conditions of the subway traction motor bearing. However, in the actual operation of the subway, frequent start and stop and changes in passenger capacity will also cause changes in the load on the bearings. Fault diagnosis under a single working condition is not enough to explain the stability and effectiveness of the proposed method. It is still necessary to verify the effectiveness of the proposed method under variable load and variable working conditions.

5.1. Intelligent Diagnosis of Bearing Failure in Single Working Condition Change. The change of signal during bearing operation is mainly affected by the speed and load. Here, the speed and load are, respectively, controlled to be constant, and the proposed SDAE is used to diagnose the bearing fault. The collected bearing vibration signal data is processed according to the method proposed in Section 4.4. According to the number, it is divided into 6 types of faults, the label is consistent with the fault number, and the correctness of the fault diagnosis of the SDAE network is tested through experiments. The results are shown in Tables 5 and 6.

From Tables 5 and 6, it can be concluded that the accuracy of fault diagnosis under the same speed and variable load can basically reach 100%, while under the same load and variable speed, the accuracy of fault diagnosis decreases slightly, especially when the training set speed is higher than the test set speed. The reason is that the rotation speed of the training set is high, and obvious fault characteristics are extracted, while the rotation speed of the test set is low, and the fault characteristics are lower than the failure amplitude of the training set, so the diagnostic accuracy rate drops slightly.

5.2. Intelligent Diagnosis and Comparative Verification of Bearing Faults with Changing Operating Conditions. In order to verify the effectiveness of SDAE again, a more complex

data set was selected as the training set and test set (as shown in Table 7) and, compared with other methods on this data set, reflect the advantages of SDAE in the case of variable speed and variable load. The following common fault diagnosis methods are selected for comparison and verification. The diagnostic accuracy of different methods is shown in Table 8:

- (1) Method 1: the vibration signal is decomposed by 8-layer wavelet packet, and the root mean square value, kurtosis value, and margin value are extracted, respectively, in 256 frequency bands after decomposition and then input into convolution neural network for fault diagnosis. According to the number of input layer elements 768, the convolution layer is selected as 2 layers, which are (28×28) and (12×12) , the pooling layer is 2 layers (14×14) and (6×6) , the dense layer is 1 layer, and the output layer uses "Softmax" classifier. The convolution kernel is set to 3×3 , the sliding step size is set to 1, the learning rate is 0.001, the activation function is "ReLU" activation function, and the training times are 25.
- (2) Method 2: the vibration signal is decomposed by 8-layer wavelet packet, and the root mean square value, kurtosis value, and margin value are extracted, respectively in 256 frequency bands after decomposition, which are input into SDAE for fault diagnosis. The network structure of SDAE is 768-128-64-6, the number of batch training is 25, and the number of iterations is 20.
- (3) Method 3: the vibration signal is decomposed by EMD, and 12 characteristic indexes, such as mean, standard deviation, variance, skewness index, kurtosis value, peak-peak value, peak mean square amplitude, average amplitude, root mean square value, waveform index, and peak value index, are extracted from the first five IMF components and input into support vector machine (SVM) for fault diagnosis.
- (4) Method 4: the frequency domain signal of vibration signal is directly input into convolution neural network for fault diagnosis. Two convolution layers (60×60) and (28×28) , two pooling layers (30×30) and (14×14) , one dense layer, and "Softmax" classifier are selected for output layer. The convolution kernel is set to 3×3 , the sliding step is set to 1, the learning rate is 0.001, the activation function is the "ReLU" activation function, and the number of training times is 25.
- (5) Method 5: the frequency domain signal of vibration signal is directly input into the traditional SDAE network. The traditional SDAE network does not add BN layer for normalization, does not use L2 regularization, and uses ReLU activation function, the network structure is 3750-128-64-6, the learning rate is 0.001, the number of batch training is 256, and the number of iterations is 20.

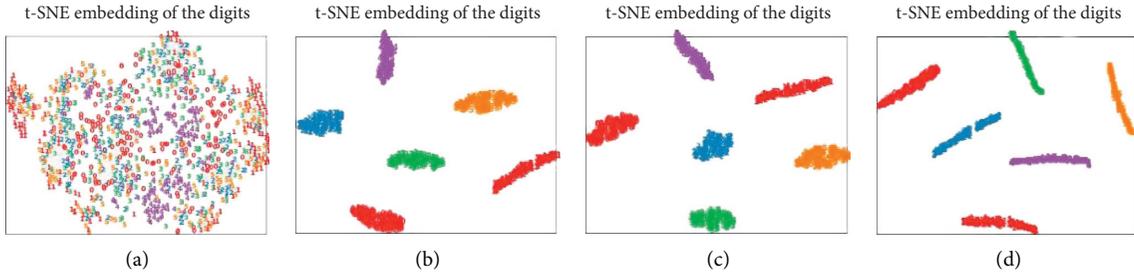


FIGURE 11: Feature layer-by-layer visualization. (a) Raw data. (b) First-layer features. (c) Second-layer features. (d) Classification results.

TABLE 5: Fault diagnosis accuracy of variable load at the same speed.

No.	Training set speed (rpm)	Training set load (kN)	Test set speed (rpm)	Test set load (kN)	Accuracy (%)
1	800	5, 7	800	9	100
2	800	5, 9	800	7	100
3	800	7, 9	800	5	99.15
4	800	5, 7, 9	800	5, 7, 9	100
5	1600	5, 7	1600	9	100
6	1600	5, 9	1600	7	100
7	1600	7, 9	1600	5	100
8	1600	5, 7, 9	1600	5, 7, 9	100
9	2400	5, 7	2400	9	100
10	2400	5, 9	2400	7	100
11	2400	7, 9	2400	5	100
12	2400	5, 7, 9	2400	5, 7, 9	100

TABLE 6: Fault diagnosis accuracy of variable speed under the same load.

No.	Training set speed (rpm)	Training set load (kN)	Test set speed (rpm)	Test set load (kN)	Accuracy (%)
1	800, 1600	5	2400	5	100
2	800, 2400	5	1600	5	99.61
3	1600, 2400	5	800	5	91.88
4	800, 1600, 2400	5	800, 1600, 2400	5	100
5	800, 1600	7	2400	7	100
6	800, 2400	7	1600	7	99.38
7	1600, 2400	7	800	7	96.37
8	800, 1600, 2400	7	800, 1600, 2400	7	100
9	800, 1600	9	2400	9	100
10	800, 2400	9	1600	9	100
11	1600, 2400	9	800	9	98.76
12	800, 1600, 2400	9	800, 1600, 2400	9	100

TABLE 7: Data set of multiple working conditions.

Data set	Training set speed (rpm)	Training set load (kN)	Test set speed (rpm)	Test set load (kN)
A	800,1600	5, 7, 9	2400	5, 7, 9
B	800,2400	5, 7, 9	1600	5, 7, 9
C	1600,2400	5, 7, 9	800	5, 7, 9

TABLE 8: Diagnostic accuracy of different methods.

Method	Accuracy (%)		
	A	B	C
The method of this paper	100	99.79	96.18
Method 1	77.78	84.72	37.50
Method 2	90.76	95.38	60.85
Method 3	70.23	76.12	42.65
Method 4	98.65	97.28	91.59
Method 5	91.82	96.69	79.54

6. Conclusion

In the bearing fault diagnosis of metro traction motor, the extraction of fault characteristics of nonstationary signal under complex working conditions has been one of the research hotspots. The commonly used fault feature extraction methods are manual feature extraction, but in complex working conditions, the effect of feature extraction is not ideal. The improved SDAE network proposed in this paper can adaptively extract fault information under complex working conditions, and the fault diagnosis accuracy rate is more than 96.18%. The following conclusions can be drawn:

- (1) Through the comparison between the method and method 4, BN layer is added to the traditional SDAE network for normalization optimization, L2 regularization is used to prevent overfitting, and Leaky ReLU activation function is introduced for improvement, which effectively solves the problems of gradient dispersion and overfitting. The extracted features can well express the original signal and improve the accuracy of fault diagnosis.
- (2) Under the condition of constant speed and constant load, the diagnostic accuracy of fault feature extracted by this method is equivalent to that of manual extraction method, which can reach 100% accuracy.
- (3) Under the condition of variable speed and variable load, the accuracy rate of fault feature extracted by this method is far higher than that of traditional manual feature extraction methods in method 1, method 2, and method 3. It shows that the improved SDAE still has strong feature extraction ability under variable working conditions, which can meet the engineering requirements and can be used in the feature extraction stage of mechanical fault diagnosis.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] L. Jianqiang, C. Aifeng, and Z. Nan, "An intelligent fault diagnosis method for bogie bearings of metro vehicles based on weighted improved D-S evidence theory," *Energies*, vol. 11, no. 1, p. 232, 2018.
- [2] S. S. Udmale, S. K. Singh, and S. G. Bhirud, "A bearing data analysis based on kurtogram and deep learning sequence models," *Measurement*, vol. 145, pp. 665–677, 2019.
- [3] R. Chen, S. Chen, M. He, D. He, and B. Tang, "Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion," in *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 231, no. 6, pp. 666–679, 2017.
- [4] Y. Tian and X. Liu, "A deep adaptive learning method for rolling bearing fault diagnosis using immunity," *Tsinghua Science and Technology*, vol. 24, no. 6, pp. 750–762, 2019.
- [5] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 73, pp. 303–315, 2016.
- [6] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, pp. 289–306, 2016.
- [7] A. Glowacz, W. Glowacz, and Z. Glowacz, "Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals," *Measurement*, vol. 113, pp. 1–9, 2017.
- [8] J. Chen, J. Pan, Z. Li, Y. Zi, and X. Chen, "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals," *Renewable Energy*, vol. 89, pp. 80–92, 2016.
- [9] Y. Li, M. Xu, Y. Wei, and W. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80–94, 2016.
- [10] M. Zvokelj and S. Zupan, "EEMD-based multiscale ICA method for slewing bearing fault detection and diagnosis," *Journal of Sound & Vibration*, vol. 370, pp. 394–423, 2016.
- [11] F. Jia, Y. G. Lei, and L. Guo, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2017.
- [12] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and R. Wang, "Development and application of a deep learning-based sparse autoencoder framework for structural damage identification," *Structural Health Monitoring*, vol. 18, no. 1, pp. 103–122, 2018.
- [13] H. Shao, H. Jiang, and H. Zhang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 2727–2736, 2017.
- [14] S. Muhammad, K. Cheol-Hong, and K. Jong-Myon, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, p. 2876, 2017.
- [15] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 101–110, 2018.
- [16] P. Vincent, H. Laroche, and I. Lajoie, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 12, pp. 3371–3408, 2010.
- [17] L. Wen, X. Li, and L. Gao, "A new convolutional neural network based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2017.
- [18] X. Guo, C. L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [19] H. Shao, F. Jiang, 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.

- [20] H. Shao, H. Jiang, and H. Zhao, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 95, pp. 187–204, 2017.
- [21] J. Wang, 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, 2017.