Research on Abnormal Feature Extraction and Early Fault Alarm Method of Rolling Bearings Based on CDAE and KLD

Zheng Qin

School of New Energy, East China University of Petroleum, Qingdao, Shandong, China.

Qin Chang

College of Science, East China University of Petroleum, Qingdao, Shandong, China.

Qiang Li, Yao Wang, Jie Wang and Weiwei Xu

School of New Energy, East China University of Petroleum, Qingdao, Shandong, China.
E-mail: liqiangsydx@163.com

(Received 20 November 2022; accepted 3 April 2023)

A rolling bearing is an important part of rotating machinery, and it is widely used in the petrochemical industry, aerospace industry and other industries. Hence, it is of great significance to carry out condition monitoring and fault alarms for rolling bearings. Aiming at the problem of the rolling bearing fault, a method of an improved deep convolutional denoising auto encoder abnormal feature extraction and the Kullback-Leibler divergence threshold alarm is proposed. The experiment verification is carried out on the rotor bearing experiment platform. The experiment results show that the proposed method has good denoising performance and micro fault feature extraction ability under the condition of no fault data training and no frequency domain transformation. High accuracy, good efficiency and strong robustness of the proposed method for an early fault alarm are demonstrated by the experiment as well.

1. INTRODUCTION

As one of the most important parts of rotating machinery, a rolling bearing plays an important role in the safe operation of rotating machinery.[1] Once there is a failure in the equipment operation process, it may cause huge economic losses and even catastrophic accidents.[2,3] In order to ensure the safe and stable operation of rolling bearings, the research of abnormal feature extractions and an early fault alarm method is of highest importance. Although the traditional artificial fault alarm method has resulted in many achievements and wide applications, there are still some limitations, such as inaccurate feature selection, strong dependence on the training set, strong noise interference, alarm time lag, poor intelligence, weak generalization ability, etc. Therefore, it is still of great significance to carry out condition monitoring and early intelligent fault alarm research on rolling bearings.[4,5]

Artificial neural network is a mathematical model that imitates biological brain neurons. In 1998, LeCun proposed the LeNet neural network, which is the earliest convolutional neural network model.[6] It is verified by the MNIST data set, and the accuracy rate reaches 98%. In 2012, AlexNet proposed by Krizhevsky on ILSVRC was the first large-scale deep convolutional neural network model in human history, and Krizhevsky won the championship of the challenge by overwhelming advantage.[7] In 2014, Simonyan proposed the method of constructing deep convolutional neural network with a small convolutional kernel, and this method has been widely expanded and applied by future generation scholars.[8] In 2016, He proposed ResNet neural network, which greatly improved the efficiency and the overall stability of the model.[9] In 2017, Xie combined the advantages of GoogleNet and ResNet and proposed the ResNeX model, which solved the problem of excessive super parameters and improved the accuracy of recognition.[10] In 2020, Selvaraju proposed a technique for producing visual explanations for decisions from a large class of Convolutional Neural Network (CNN)-based models, making them more transparent and explainable.[11] In 2021, Gao proposed Res2Net block, which can be plugged into the state-of-the-art backbone CNN models, e.g., ResNet.[12]

As a branch of artificial intelligence, intelligent fault diagnosis has made great progress with the development of computer peripheral functions. Neural network algorithm and feature extraction have been applied in the field of rolling bearing fault diagnosis. Shen et al. proposed many dimensionless features as feature parameters and applied them to fault diagnosis of rolling bearings under various conditions, and the model achieved good recognition results.[13] Jiang and Purushotham used a series of signal processing methods, such as EMD, EEMD and wavelet transform.[14,15] Wu et al. proposed an enveloped analysis method based on wavelet scale and applied it to the fault diagnosis of rolling bearings.[16] Zhang et al. proposed a deep convolutional neural network (WDCNN), which was verified by the marked rolling bearing data set provided by Case Western Reserve University (CWRU).[17] Lee et al. established a convolutional neural network model and applied it to the fault classification and detection of a power system semiconductor manufacturing process, and the network achieved good recognition results.[18] Gou et al. used the deep auto encoder to study the fault diagnosis of rolling bearings, taking the

original signal features as the input, and using the deep auto encoder network for pattern recognition.[19] Guo et al. used convolutional neural network for fault diagnosis and multi-linear principal component analysis of multi-channel data.[20]

For a long time, among the methods of fault diagnosis and early warning based on vibration signal processing technology, the selection and extraction of fault features and the determination of fault threshold play a decisive role.[21] After reading and sorting out the literature, the authors found the following problems. First, in the early stage of new fault detection, there are often not enough labeled samples, which leads to the lack of training data and greatly limits the use of the supervised learning neural network. Second, the previous intelligent fault diagnosis methods of rolling bearings mostly rely on frequency domain transformation, such as FFT, WT and EEMD, but lack of mining the depth characteristics of the time-domain signal itself. Third, the application scene of rotating machinery is complex and changeable, which is often accompanied by strong environmental interference and noise, so it is imperative to improve the robustness of the model.[22] Finally, the early fault symptoms are small and difficult to detect, and the traditional threshold alarm method is not sensitive to weak symptoms.

In order to solve the above problems, this paper proposes a new intelligent fault alarm method combining deep Convolutional Denoising Auto Encoder(CDAE), PCA, T-SNE and Kullback-Leibler Divergence (KLD). The unsupervised feature of CDAE is used to effectively solve the problem of sample missing in the early stage of fault occurrence. The feature is extracted directly from the time domain signal of rolling bearings without frequency domain transformation. The convolutional kernel has excellent denoising performance and it can be used to reduce the influence of the environment and noise on the feature extraction results. PCA algorithm and T-SNE algorithm are used to reduce and visualize the abnormal features. Due to KLD's wonderful data set distance measurement ability and as it is used for the micro feature recognition, the automatic threshold alarm of an early weak fault is realized.

2. DESIGN OF FEATURE EXTRACTION AND EARLY FAULT WARNING MODEL

The overall design of the abnormal feature extraction and early fault alarm method of rolling bearings based on CDAE and KLD is shown in Fig. 1. The method mainly includes six processes: pretreatment process, training process, feature extraction process, calculating alarm value process, test process, and the alarm process.

Firstly, the rolling bearing historical health data set X is obtained by the experiment platform. In order to verify the superiority of the subsequent method more strictly, the original signal is not transformed in the frequency domain such as FFT. The original time domain signal is directly normalized, $x = \{x_i | 1 \leq i \leq n\}$, where x_i is the i-th group of the sample data, n is the total number of samples, and n is 1024 in this paper. Then, the normalized sample set x is processed by adding 0.3 times of Gaussian white noise to get the fault sample set x'. This is the pretreatment process.

The CDAE network is highly effective at automatically extracting features and is an excellent feature extraction tool. Each convolutional kernel can be regarded as a filter. After filtering layer by layer, the input data are finally restored to the original healthy signal. Because of its unsupervised training characteristics, CDAE can be built and trained without the existing fault labels and the types of fault samples in the training set. The sample set x' from the pretreatment process is used as the input of the neural network, and the normalized healthy sample set x is used as the output of the neural network. After the training of the CDAE, the iterative model is obtained, and its coding part is saved. Principal Component Analysis (PCA) and T-distributed Stochastic Neighbor Embedding (T-SNE) are used for feature reduction and extraction to get the feature set P. After calculating the expectation p_0 of P and the variance σ_0 of P, the alarm threshold can be calculated by $K = p_0 + 3\sigma_0$.[23,24]

The data of nine kinds of working conditions and one kind of healthy working condition after implanting the fault are normalized. These data are imported into the trained model as the test set. Additionally, through feature extraction, feature dimensionality reduction and KLD calculation, finally, the fault sample alarm is realized by comparing it with the alarm threshold.

3. RESEARCH ON ABNORMAL FEATURE EXTRACTION OF DENOISE BASED ON CDAE

3.1. Basic Principles of CDAE

CDAE is made up of two convolutional neural networks. It has the advantage of weight sharing of convolutional neural networks and can map low-dimensional shallow features to high-dimensional high-level features. So, all the data satisfying local correlation can be processed by the convolutional auto encoder in theory.[25] The model structure is shown in Fig. 2.

The convolutional operation can share parameters, and the encoder can extract features after several convolutional operations and pooling operations.[26] The output of convolutional operation is shown in Eq. (1):

$$h^{l(i,j)} = \sum_{j=0}^{W} K_i^l(j') x^{l(j+j')}; \tag{1}$$

where K is the convolutional kernel, l is the l-th layer and i is the i-th. j is the j-th local region and j' is the j'-th weight in the convolutional kernel. In general, a rectified linear unit (ReLU) is selected as the activation function, and its output value is shown in Eq. (2):

$$a^{l(i,j)} = f_{ReLU} h^{l(i,j)} = \max\{0, h^{l(i,j)}\}. \tag{2}$$

Pooling operation can reduce data dimensions to simplify training. Usually, maximum pooling operation is used to reduce the total amount of data while retaining features.[27] Pooling operation is shown in Eq. (3):

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq s \leq jW} \{a^{l(i,s)}\}. \tag{3}$$

The decoder performs several deconvolutions and depooling operations to reconstruct the input signal and restore the original data. The process is shown in Eq. (4):

$$y_t = f^{-1}(P_t); \tag{4}$$

where f^{-1} is the decoder operation, that is the reconstruction process.

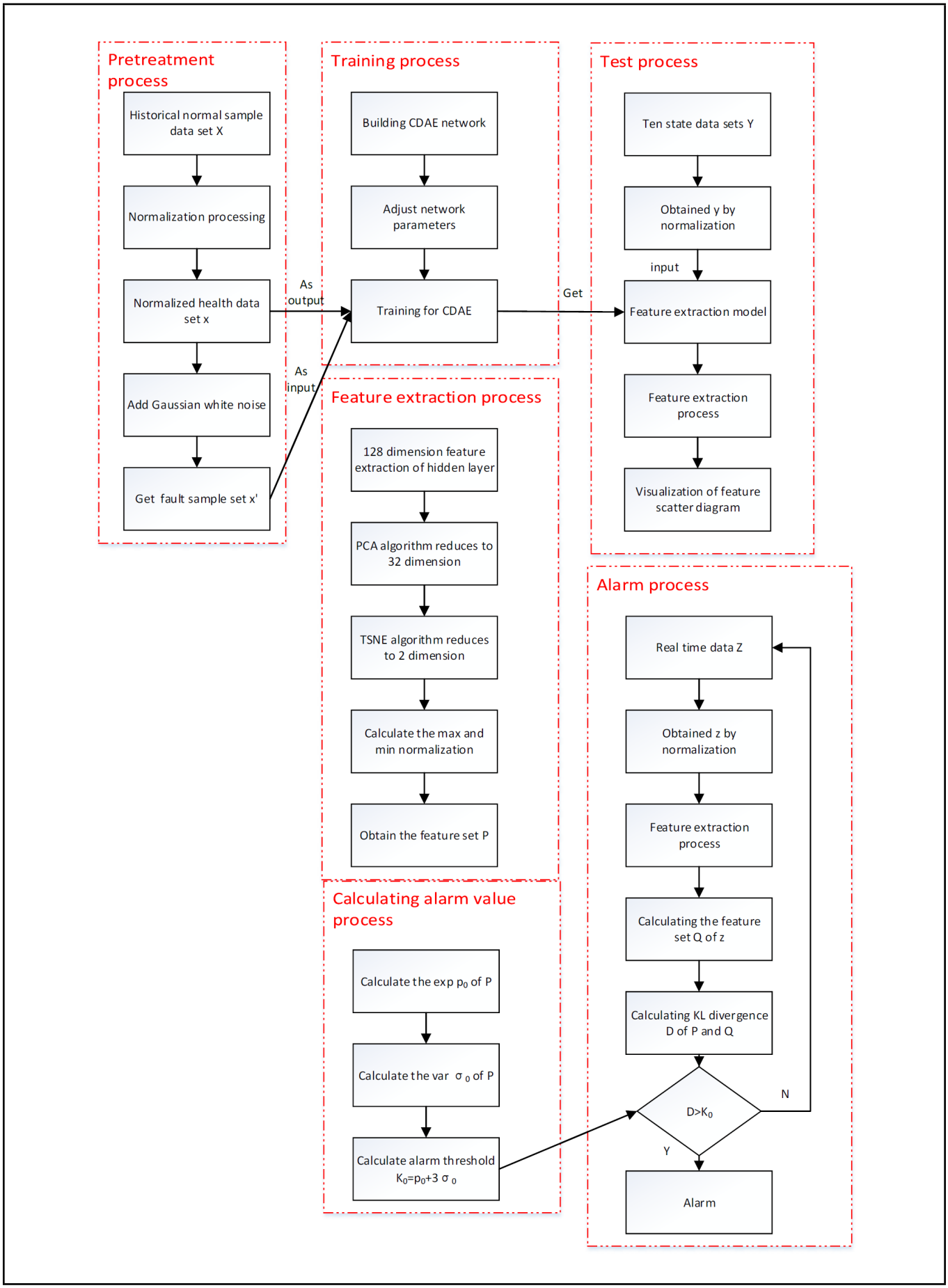

Figure 1. Algorithm flow chart.

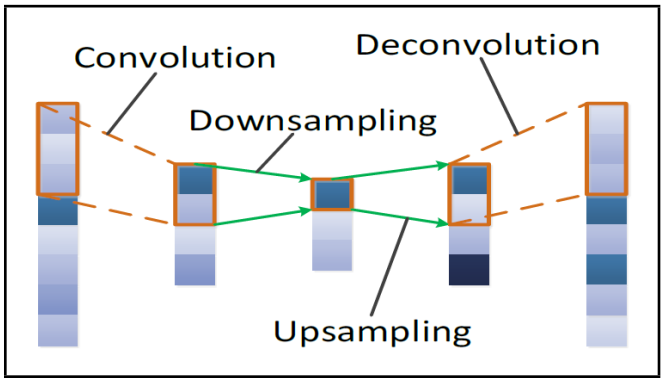

Figure 2. Structure diagram of convolutional auto encoder.

3.2. Construction of Anomaly Feature Extraction Model

In this paper, each convolutional kernel of CDAE is equivalent to a filter. The method of denoising is that in the process of training, artificial noise interference is added to the health data to mimic the damaged data set, and then the CDAE takes the data containing noise as the input and the health data as the output. And the loss function value is minimized through iterations, which makes the features extracted from the model more robust. The denoising process of CDAE is shown in Fig. 3.

The Adam algorithm can intuitively explain the super parameters and only needs few parameters, so the random gradient descent method can be replaced by Adam algorithm in the deep CDAE model. Adam algorithm has faster convergence speed and more effective learning effect compared with the other adaptive learning rate algorithms. And it can avoid the problems such as the disappearance of the learning rate, slow convergence, and the large fluctuation of loss function caused by the parameter update of the high square error.

The mean square logarithmic error (MSLE) is used as the loss function here. Compared to the MSE loss function, which would be guided by some large values and lead to deviation of the final result even if the small value is accurate, MSLE can effectively alleviate the deviation of the final result by taking logarithm of all the values in advance.

Each group of input data is 1024 sampling points. The input layer is added with 0.3 times Gaussian white noise. The sample batch size is 128. The number of iterations is 30. The model parameter settings are shown in Table 1.

T-SNE is a machine learning algorithm for data dimensionality reduction. In the data correlation analysis and visualization processing, T-SNE can fully retain the effective information in the process of data dimensionality reduction, making the extraction results more representative and fully reflecting the differences between samples.

The distribution function of the high dimensional space in a low dimensional space is shown in Eq. (5):

$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k \neq l}(1 + ||y_k - y_l||^2)^{-1}}. \tag{5}$$

The iterative gradient is shown in Eq. (6):

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ji} - q_{ji})(y_i - y_j)(1 + ||y_i - y_j||^2)^{-1} \tag{6}$$

However, it runs very slowly, so it is not suitable for large-scale calculation of big data. The PCA dimensionality reduction method is used to find the covariance matrix C, its eigenvalues and eigenvectors of the data set. It arranges the eigenvectors from top to bottom to form the eigenmatrix, and it takes the first k rows to form the matrix P. The calculation of C matrix and P matrix is shown in Eq. (7):

$$\begin{aligned} E &= (e_1, e_2, \ldots, e_n); \\ E^T C E &= \Lambda; \\ P &= E^T; \end{aligned} \tag{7}$$

where $e_1, e_2, \ldots, e_n$ are eigenvectors, Λ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues of each eigenvector.

In this paper, the application of PCA reduced the data to 16 dimensions, and then the using of T-SNE reduced it to 2 dimensions. The abscissa of the characteristic scatter diagram is retained to participate in the calculation of the subsequent alarm method research.

Table 1. Model parameter settings.

	Kernel size	Kernel number	Activation function	Nonzero padding
Convolution layer	64	4	ReLU	Same
Convolution layer	16	8	ReLU	Same
Pooling layer	2	–	–	–
Convolution layer	8	16	ReLU	Same
Pooling layer	2	–	–	–
Convolution layer	4	32	ReLU	Same
Convolution layer	1	64	ReLU	Same
Pooling layer	2	–	–	–
Unpooling layer	2	–	–	–
Deconvolution layer	1	64	ReLU	Same
Deconvolution layer	4	32	ReLU	Same
Unpooling layer	2	–	–	–
Deconvolution layer	8	16	ReLU	Same
Unpooling layer	2	–	–	–
Deconvolution layer	16	8	ReLU	Same
Deconvolution layer	64	4	ReLU	Same
Deconvolution layer	1	4	Sigmoid	Same

4. RESEARCH ON FAULT ALARM METHOD BASED ON KLD

4.1. Basic Principles of KLD

In the mathematical statistics and information theory, KLD is also known as relative entropy or information gain. It is a function to measure the information loss of replacing one distribution with another distribution, in other words, it describes the difference between two probability distributions.[28] If KLD is 0, it means that the two vectors are completely coincident. The larger the KLD is, the greater the difference is.[29]

For P and Q, which are two given probability distributions of discrete random variables, the definition of KLD is shown in Eq. (8):[30]

$$D_{KL}(P||Q) = \sum_i P(x_i) \log \left(\frac{P(x_i)}{Q(x_i)} \right). \tag{8}$$

This KLD represents the expectation of the difference between the logarithms of P and Q. For ordinary distribution, the equation derived from the variant of KLD can also be used to calculate the difference between the two distributions. The divergence is shown in Eq. (9):[30]

$$D_{KL}(P||Q) = -E\left(\ln q(x)\right) + E\left(\ln p(x)\right). \tag{9}$$

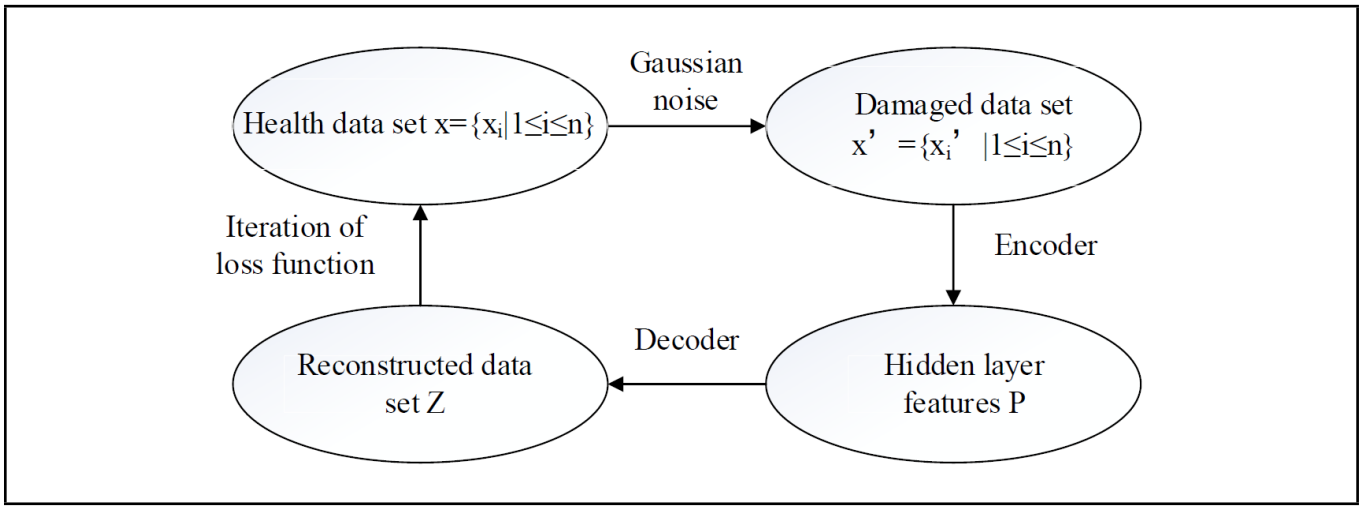

Figure 3. Denoising process of convolutional denoising auto encoder.

The definition and value range of KLD are nonnegative, as shown in Eq. (10):[30]

$$\begin{aligned} P &\geq 0; \\ Q &\geq 0; \\ D_{KL}(P||Q) &\geq 0. \end{aligned} \tag{10}$$

4.2. Fault Alarm Method

Because the definition domain of KLD is nonnegative,[31] the feature is normalized to the range of $[0, 1]$. The feature range can be adjusted by scaling without changing the clustering state. It can effectively avoid infinite results from the calculation results and improve the robustness of the model. The maximum and minimum normalization method is shown in Eq. (11):[30]

$$y_i = \frac{x - i - \min\limits_{1\leq j\leq n}\{x_j\}}{\max\limits_{1\leq j\leq n}\{x_j\} - \min\limits_{1\leq j\leq n}\{x_j\}}. \tag{11}$$

Among them, the old sequence is the x-axis feature of characteristic scatter graph after dimension reduction, and the new sequence $y_1, y_2, \ldots, y_n \in [0, 1]$ is dimensionless. General data can be standardized when needed.

After calculating the expectation p_0 and variance σ_0 of the normalized feature set P and using 3σ principle, the alarm threshold K_0 is calculated:

$$K_0 = p_0 + 3\sigma. \tag{12}$$

KLD does not satisfy the symmetry of distance and the relationship of three inequalities, that is, it does not satisfy the basic attributes of distance.[32,33] In this paper, the distance is constructed by KLD, as shown in Eqs. (13) and (14),[30] and the problems of nonnegativity and symmetry are solved at the same time:

$$D_{KL}(P, Q) = \frac{1}{2}\left(D_{KL}(P||Q) + D_{KL}(Q||P)\right); \tag{13}$$

$$D_{KL}(P||Q) = \sum_i P(x_i) \log\left(\frac{P(x_i)}{Q(x_i)}\right). \tag{14}$$

Normalizing the real-time data set Z to get the set z and the feature set Q by the process of feature extraction of neural network. When using Eq. (13) to calculate the KLD of set P and Q, then if it is larger than the alarm threshold, an alarm will be given. If it is less than the alarm threshold, the real-time data will continue to be read into the algorithm. KLD fault alarm method flow chart is shown in Fig. 4.

The real-time data are segmented by the number of sampling points, that is, 1024 data points are input into the program at a time, and KLD is made with the health reference data in the training set, as shown in Fig. 5. The calculated value between the real-time sample and the reference sample is the monitoring value of condition monitoring curve.

5. EXPERIMENT VERIFICATION

5.1. Acquisition of Experiment Data

In order to verify the feasibility and accuracy of the model, a HZXT-008 small rotor test-platform is used, which can flexibly configure mechanical parameters such as vibration, speed, noise, displacement, and sensors for measurement. It can also simulate the transient process of speed up and down of rotating machinery and the vibration state of steady-state operating conditions, as well as a variety of common rotating machinery faults to analyze the fault characteristics. The test platform mainly includes an 800 mm * 150 mm test-platform, a 400 W motor, a motor mounting frame, a 1000 mm straight steel shaft, two deep groove ball rolling bearings, two adjustable eccentric bearing seats and three couplings. The experiment platform is shown in Fig. 6.[34]

The experiment includes ten common states of rolling bearings in operation, which are health state, severe, moderate, and mild state of outer ring fault, inner ring fault and rolling element fault. Four bearings are deep groove ball rolling bearings. The traditional methods of a rolling bearing fault prefabrication are usually to disassemble the fault parts of the equipment that has been in fault, or to aging the bearing. However, the bearing faults obtained by these two methods are unpredictable, the fault degree is not easy to measure, and the fault period cannot be accurately determined. Therefore, this paper uses a method to simulate the pitting fault of a bearing by a laser spot burning which achieves good data effect. A laser was used to cauterize the outer ring raceway, inner ring raceway and rolling element respectively. Among them, there were

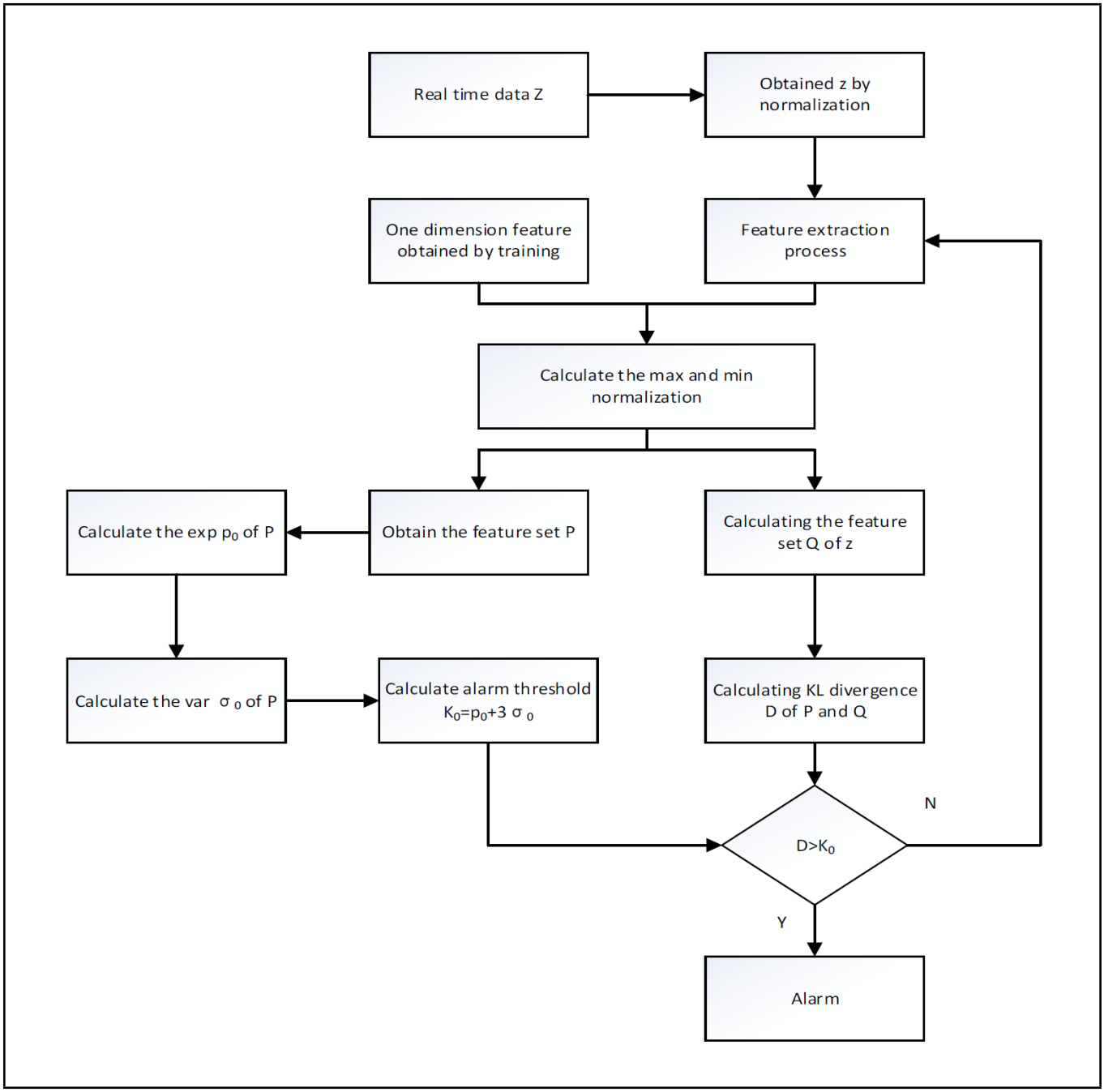

Figure 4. KLD fault alarm method flow chart.

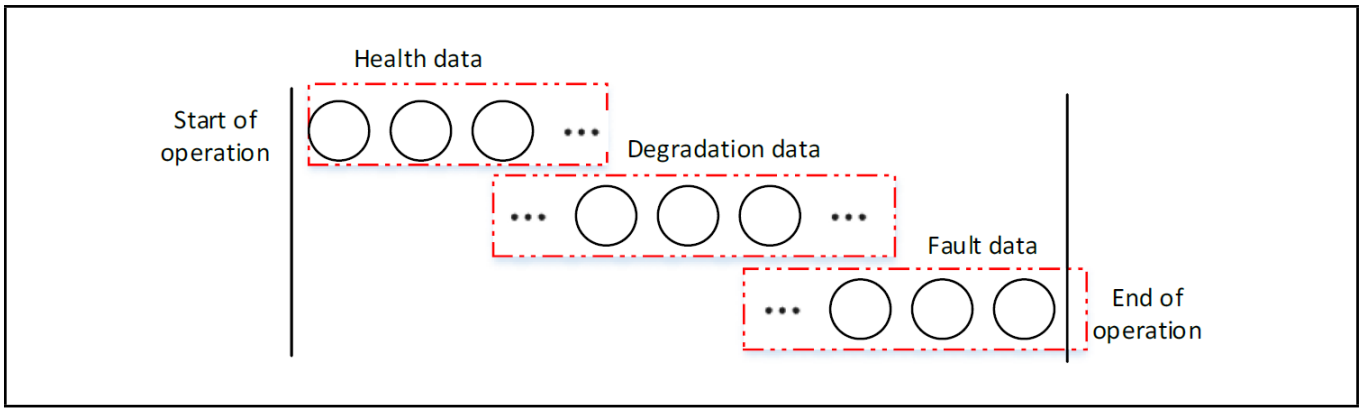

Figure 5. Data change process chart.

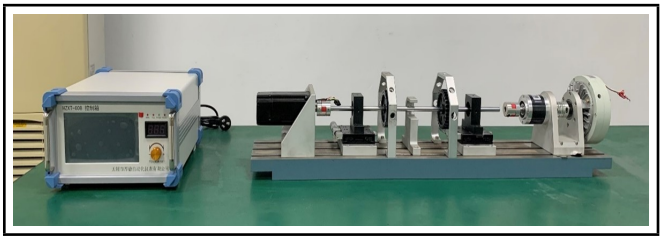

Figure 6. Rolling bearing fault simulation test platform.[34]

Table 2. Sample set partition.

Sample set type	Sample length	Sample number	Training set	Test set	Label
Health	1024	500	400	100	0
9out	1024	100	0	100	1
6out	1024	100	0	100	2
3out	1024	100	0	100	3
9in	1024	100	0	100	4
6in	1024	100	0	100	5
3in	1024	100	0	100	6
9ro	1024	100	0	100	7
6ro	1024	100	0	100	8
3ro	1024	100	0	100	9

nine times of severe failure, six times of moderate failure and three times of mild failure. The pitting size of the simulation results could be detected by a high-power microscope.[35] The faulty parts are shown in Fig. 7.

The sampling frequency of all experiment data is $10k$ Hz ($1k = 1024$), while the rotor speed is 2000 r/min, the number of sampling points is 1024, the motor power is 400 W, and the acceleration sensing sensitivity is 10 mV/(m/s^2). Each fault state uses 100 groups of samples, and the health state uses 500 groups of samples. Randomly selecting 80% of the samples in the health state as the training set, and 20% of the samples as the test set. The specific information is shown in Table 2. The original vibration time domain signals of ten states are shown in Fig. 8 ($n = 4k$).

5.2. Feature Extraction Results and Analysis

After obtaining the data set X of the healthy operation state of the experiment platform with the tag of 0, it is normalized to get the data set x. Gaussian white noise is added to the sample set x to make it become damaged sample set x'. With x' as input and x as output, using training sets to start iteration. If the reconstruction error of the loss function meets the accuracy requirements of the model, the model network is trained and saved, which can be used as a feature auto extraction model. Ten samples of the test set are input into the feature auto extraction model, and the hidden layer feature vector is output. The iterative process curve of model training set and test set is shown in Fig. 9.

The feature scatter diagram of T-SNE after the feature extraction of the test set data and the feature scatter diagram of the comparison method are shown in Fig. 10. It can be seen that the feature of the Fig. 10(b) raw data and Fig. 10(c) CAE method are scattered and mixed. The bearing state cannot be distinguished and there is no effective clustering. The effective alarm standard cannot be formed. In Fig. 10(a), the health features are clustered on the left side by the CDAE method, which is quite different from the fault features. It can provide effective data support for the follow-up research of alarm methods.

Adding 0.3 times Gaussian white noise to the data of ten test sets collected by the experimental platform verified the denoising performance of the algorithm. The feature extraction results are shown in Fig. 10(d) It can be seen from the figure that Fig. 10(d) is greatly different from Figs 10(b) and 10(c). And the Fig. 10(d) is similar to the scatter distribution of Fig. 10(a). The scatter points under healthy working conditions are concentrated on the left side of the figure, which is greatly different from the scatter points of fault data. The subsequent methods can be used for the fault alarm.

5.3. Alarm and Analysis of Rolling Bearing

According to the flow chart in Fig. 4, the maximum and minimum normalization of the x-axis of the feature scatter diagram is performed to obtain the reference standard feature set P. The expectation $p_0 = 0.126$ is calculated and the variance $\sigma_0 = 0.029$ is calculated. The alarm threshold is $K_0 = p_0 + 3\sigma_0 = 0.213$. The degradation process of rolling bearings was simulated by laser cauterization times, and the rolling element fault data, inner ring fault data and outer ring fault data were collected in real time. Each fault type collects health state and three fault depth states. Each state includes 100 groups of experiment data, and each group of experiment data has 1024 sampling points. KLD values of each group of data and reference data were calculated to form condition monitoring curve. The red dotted line represents the bearing alarm threshold.

The experiment results of three fault types are shown in Fig. 11. It can be seen from the figure that for the three different types of faults, this method can find abnormal features in time and make an accurate alarm. The comprehensive fault identification rate can reach more than 95%. The method in this paper can be seen from Fig. 11 that the alarm time is advanced to achieve the early warning effect. It has high sensitivity to early minor faults, and the research on denoising strengthens the robustness and reliability of the system.

At present, the main effective methods of an early weak fault early warning of rolling bearing affected by noise are as follows. The feature extraction method of an early fault of rolling bearing based on: 1) optimal wavelet scale cyclic spectrum;[36] 2) singular point identification and feature extraction of early fault of rolling bearing based on instantaneous envelope scale spectral entropy;[37] 3) the early fault diagnosis method of rolling bearings based on parameter optimization variational modal decomposition method;[38] 4) sparse feature extraction of a bearing's early fault using coherent cumulant piecewise orthogonal matching tracking method;[39] 5) determination of optimal reconstruction scale of continuous wavelet and early fault identification;[40] and, 6) a bearing early fault diagnosis method based on wavelet correlation permutation entropy.[41] These methods can realize the denoising and micro fault early warning of rolling bearings, but on one hand, their condition monitoring depends on the frequency domain transformation or time-frequency analysis. It is rare to explore the alarm technology combining time-domain signal directly with intelligent algorithms. On the other hand, the above methods are to study the mechanism of faulty parts and explore the special characteristics of parts. Relying too much on the expert experience of external mechanical disciplines leads to insufficient intelligence. This highlights the advantages of this method.

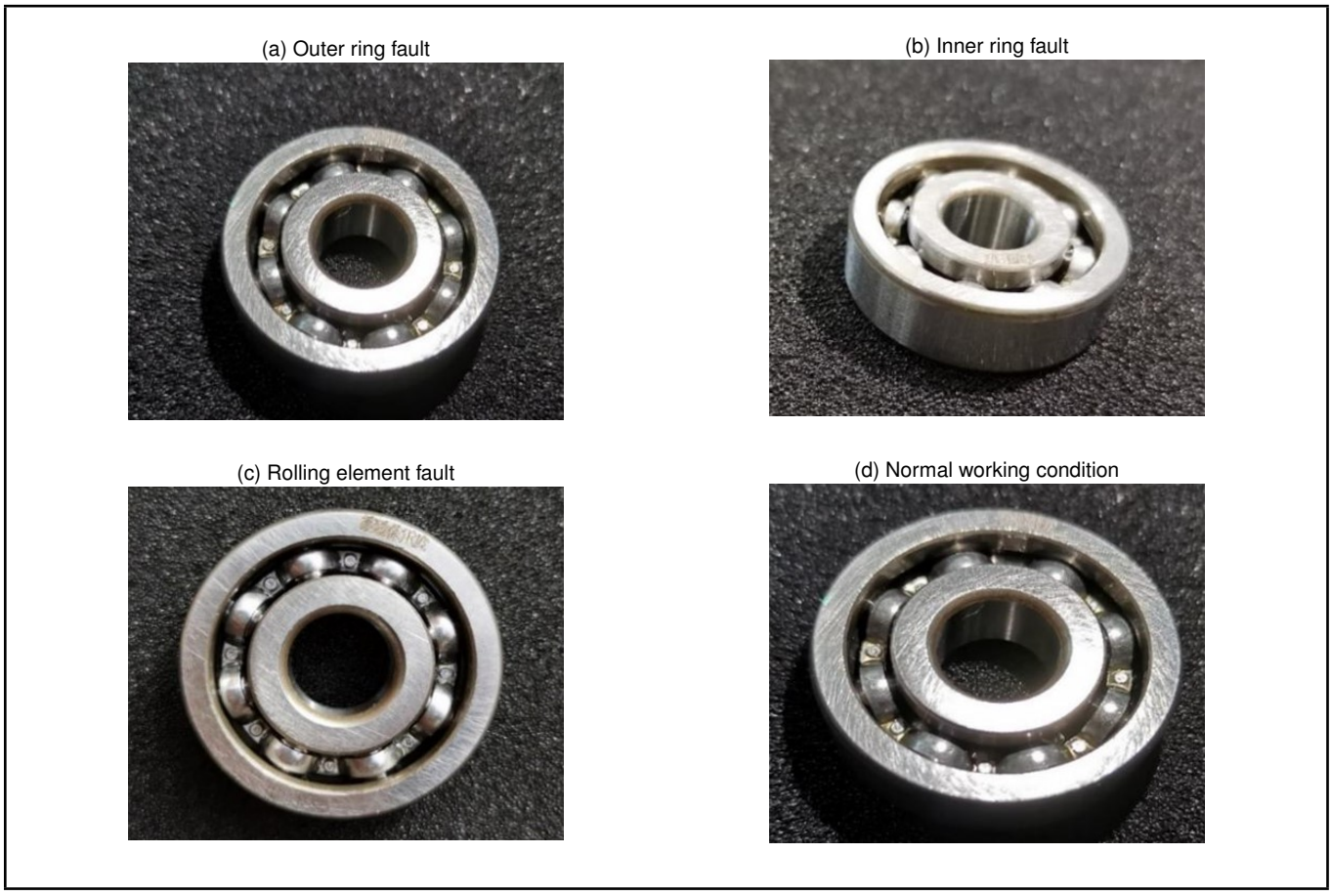

Figure 7. Rolling bearing faulty parts.

5.4. Verification of Rolling Bearing Data of Case Western Reserve University

In order to further verify the effectiveness of the algorithm, the rolling bearing database of CWRU is used for experimental verification. The database is an international public standard in the field of fault diagnosis. Ten common states in the operation of rolling bearings are selected; namely, health state, outer ring fault, inner ring fault and rolling element fault. The sampling frequency is 48K, and the specific working conditions of the bearing are shown in Table 3. In order to maintain the consistency of the experiment, the data set division method is shown in Table 2. The training set and test set are obtained after normalization and noise processing on the basis of the original data.

The feature extraction results are shown in Fig. 12. According to the figure, the healthy scattered points are gathered on the left side of the figure. Similar to the extraction results in the previous section, under the interference of noise, the healthy state and fault state can also be specifically distinguished to form an effective alarm standard.

After the fault characteristics are processed by the alarm method in Section 3.2, the condition monitoring curve is shown in Fig. 13. The expectation $p_0 = 0.052$ is calculated and the variance $\sigma_0 = 0.019$ is calculated. The alarm threshold is $K_0 = p_0 + 3\sigma_0 = 0.109$. Due to the large fault implantation degree of CWRU data set, the fault characteristics of the degradation process are more obvious, the algorithm discrimination is higher, and the effect is better. At the boundary between the health state and the fault state, the monitoring value rises obviously, has high sensitivity to faults, and the alarm effect is better.

6. CONCLUSION AND PROSPECT

In this paper, by using the unsupervised learning characteristics of deep CDAE and healthy samples as a training set, the problem of insufficient training samples in the early stage of equipment commissioning was solved. By using a convolutional kernel which has excellent filtering characteristics to denoise the original signal with Gaussian white noise, the robustness of the model was enhanced. By using CDAE, PCA algorithm and the T-SNE algorithm, feature extraction and feature dimension reduction were carried out. The optimal feature set and feature standard were selected. The KLD measured the distance between two random distributions, and an early micro fault alarm of rolling bearings was carried out by using test platforms and the CWRU database to verify the algorithm. The experimental results show that the method had high accuracy in the feature extraction and alarm. It effectively resisted the interference of environmental noise and detected weak features of early faults in time. Moreover, it can be generalized and applied to feature extraction and early fault alarms of various rotor bearing systems.

Although the research work of this paper made some achievements, due to the limited research conditions and the author's ability, there are still limitations that need to be further studied. On one hand, for the intelligent identification of weak faults, the research for this paper was limited to distinguishing a healthy state and fault state, and was not able to

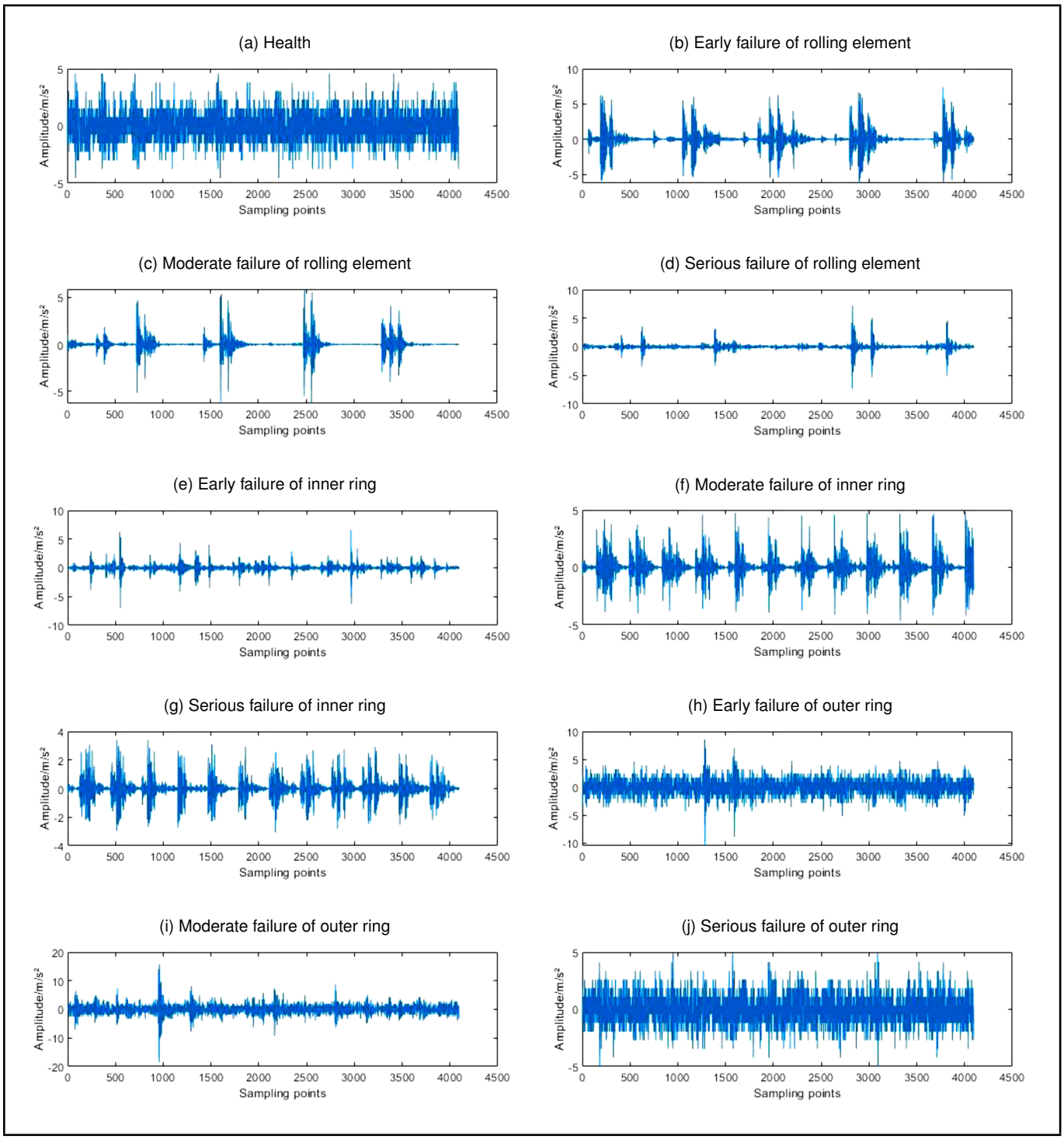

Figure 8. Alarm diagram.

Table 3. CWRU database.

Fault diameter	Motor load	Motor speed	Outer ring fault	Inner ring fault	Rolling fault
0.5334 mm	1 HP	1772 r/min	OR021@6_1_239	IR021_1_214	B021_1_227
0.3556 mm	1 HP	1772 r/min	OR014@6_1_202	IR014_1_175	B014_1_190
0.1778 mm	1 HP	1772 r/min	OR007@6_1_136	IR007_1_110	B007_1_123

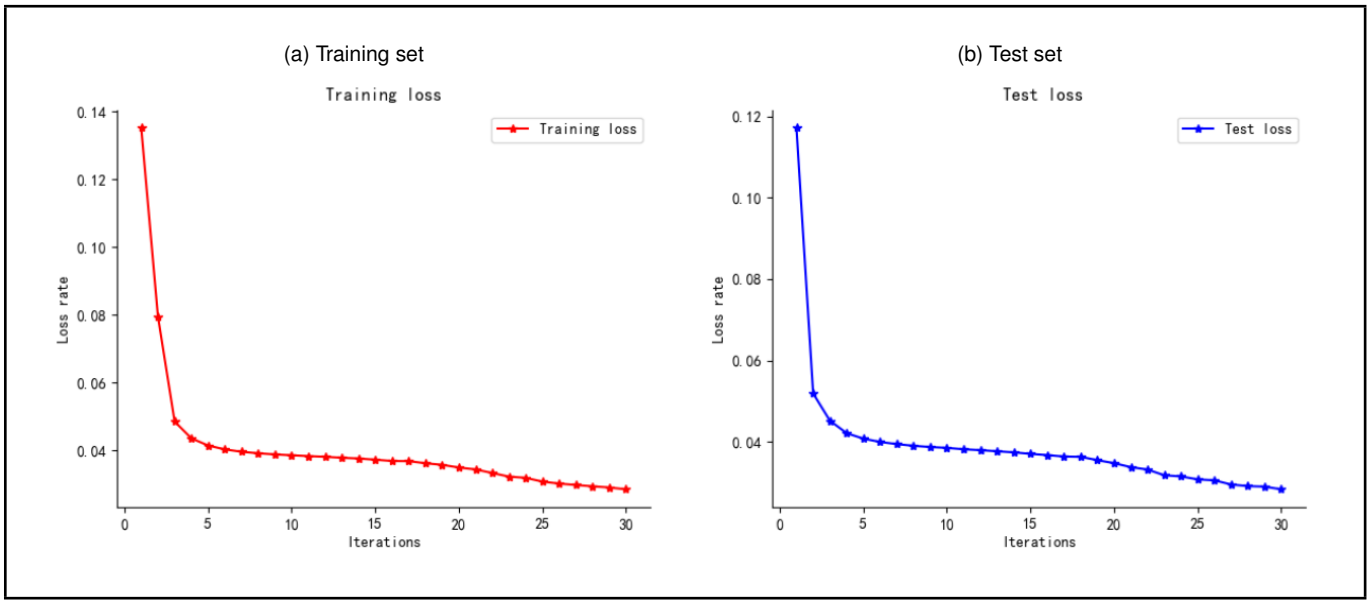

Figure 9. Model iteration process curve.

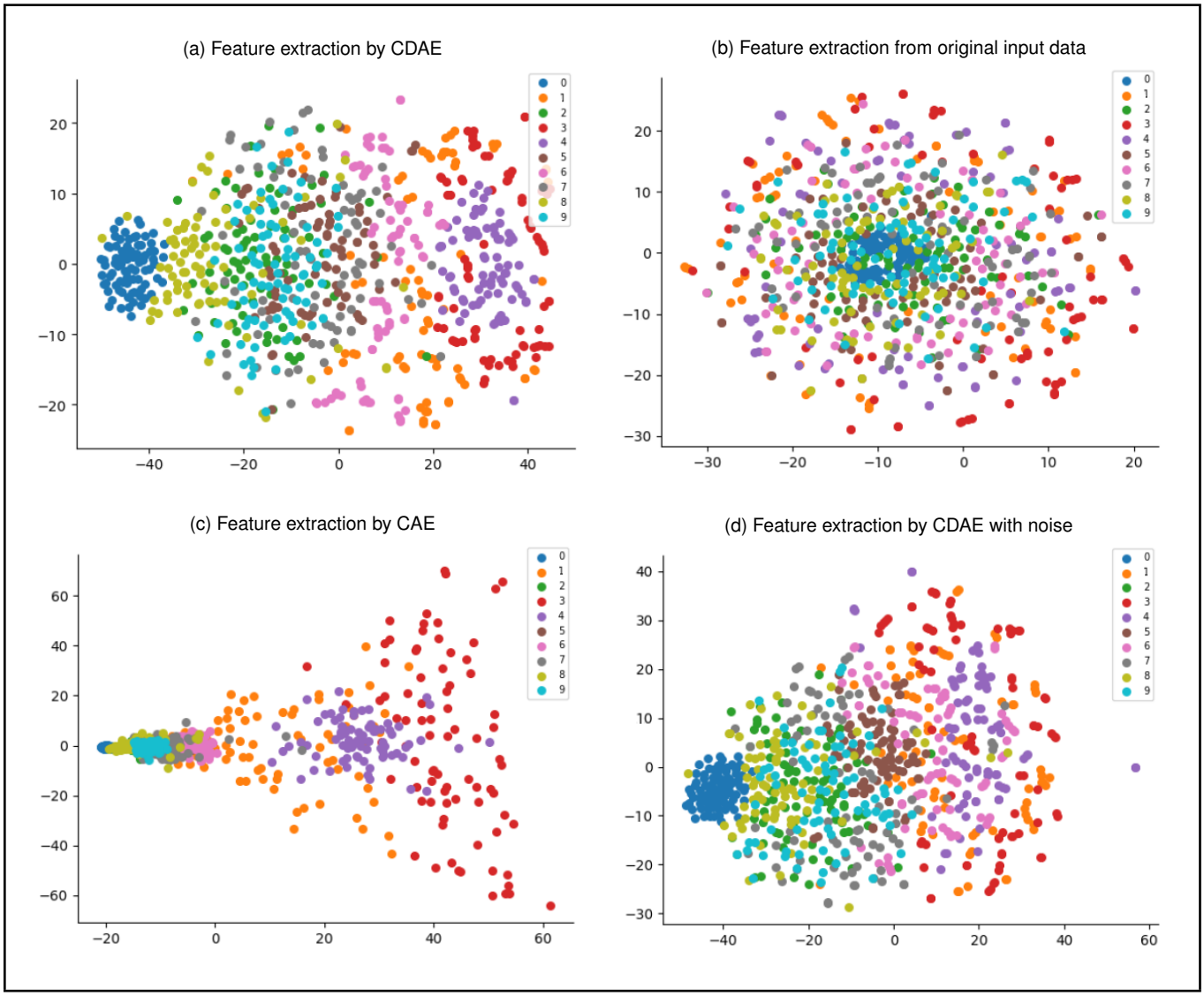

Figure 10. Scatter diagram of feature extraction.

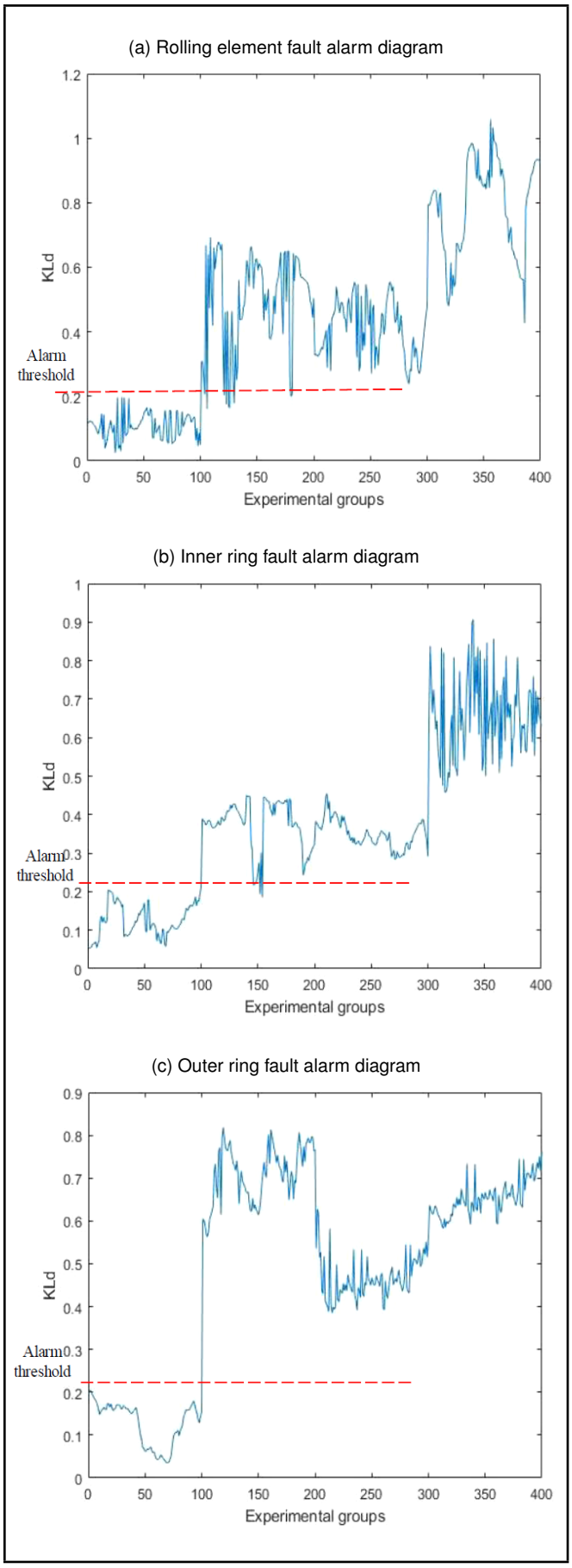

Figure 11. Alarm diagram.

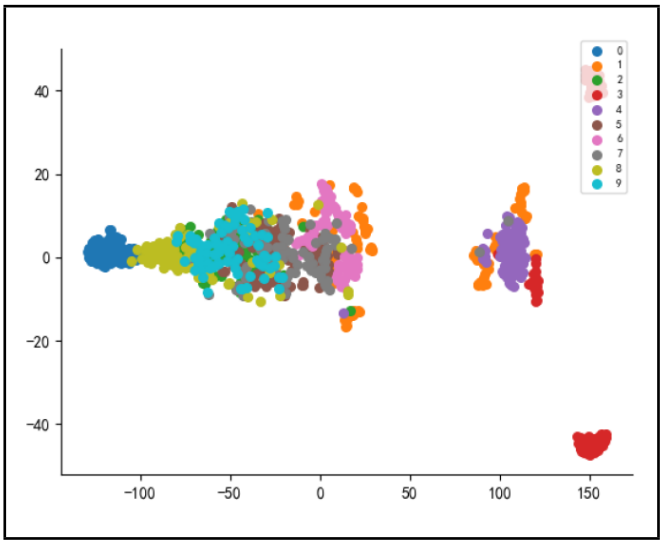

Figure 12. Feature extraction by CDAE.

accurately pattern identify each fault type and fault degree. On the other hand, the data used in this paper were from the laboratory data collection and an international general database download. Although the effectiveness of the method can be confirmed, it needs to be debugged by practical application in the factory.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 52176050), the General Program of Natural Science Foundation of Shandong Province (NO. ZR2020ME174).

REFERENCES

[1] Lei, Y. G. *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, Elsevier Inc, Amsterdam, (2017). https://doi.org/10.1016/C2016-0-00367-4

[2] Zhou, W., Habetler, T. G., and Harley, R. G. Bearing fault detection via stator current noise cancellation and statistical control, *IEEE Transactions on Industrial Electronics*, **55** (12), 4260–4269, (2008). https://doi.org/10.1109/TIE.2008.2005018

[3] Li, C., Zong, X., and Gudake, M. A survey of online fault diagnosis for pv module based on BP network, *Proc. of the International Conference on Smart City and Systems Engineering*, IEEE Computer Society, Dalian, China, 483–486, (2016). https://doi.org/10.1109/TIE.2008.2005018

[4] Muralidharan, V. and Sugumaran, V. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump, *Measurement*, **46** (1), 353–359, (2013). https://doi.org/10.1016/j.measurement.2012.07.007

[5] Ivanova, T. A. *Principles and Operation of Rotating Machinery*, Tritech Digital Media, (2018).

[6] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to documengt recognition, *IEEE*

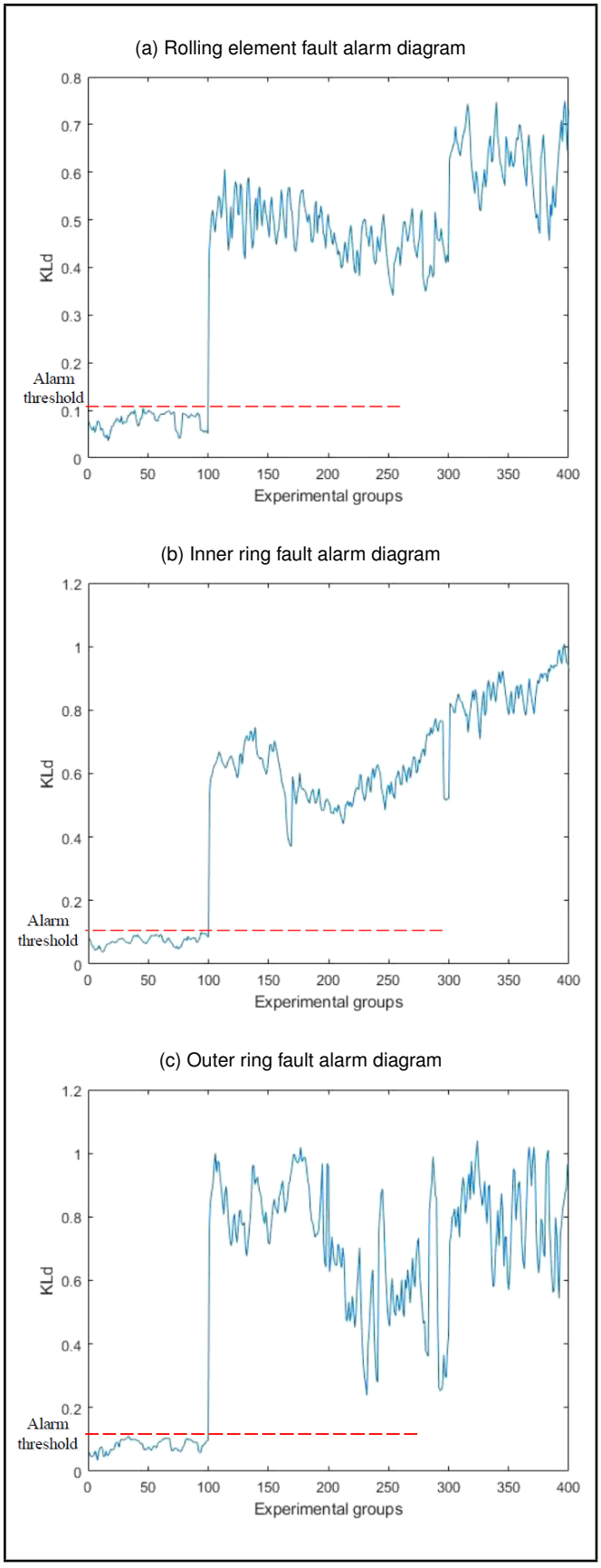

Figure 13. Alarm diagram.

Transaction on Industrial Informatics, **86** (11), 2278–2324, (1998). https://doi.org/10.1016/j.measurement.2012.07.007

[7] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolution neural network, *Advances in Neural Information Processing Systems*, **25**, 1097–1105, (2012). https://doi.org/10.1145/3065386

[8] Simonyan, K. Very deep convolutional networks for large-scale image recognition, *Proc. of the International Conference on Learning Representations*, 1409–1556, (2014). https://doi.org/10.48550/arXiv.1409.1556

[9] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learing for image recognition, *IEEE Computer Vision and Pattern Recognition*, 770–778, (2016). https://doi.org/10.48550/arXiv.1512.03385

[10] Xie, S., Girshick, R., Dollar, P., Tu, Z., and He, K. Aggregated residual transformations for deep neural networks, *IEEE Computer Vision and Pattern Recognition*, 5987–5995, (2017). https://doi.org/10.1109/cvpr.2017.634

[11] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization, *International Journal of Computer Vision*, **128** (2), 618–626, (2017). https://doi.org/10.1109/iccv.2017.74

[12] Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. Res2Net: A new multi-scale backbone architecture, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43** (2), (2019). https://doi.org/10.1109/TPAMI.2019.2938758

[13] Shen, Z., He, Z., Chen, X., Sun, C., and Liu, Z. A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time, *Sensors*, **12**, 10109–10135, (2012). https://doi.org/10.3390/s120810109

[14] Jiang, Y., Zhu, H., and Li, Z. A new compound faults detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator, *Chaos, Solitons & Fractals*, **89**, 8–19, (2016). https://doi.org/10.1016/j.chaos.2015.09.007

[15] Purushotham, V., Narayanan, S., and Prasad, S. A. N. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, *NDT & E International*, **38** (8), 654–664, (2005). https://doi.org/10.1016/j.ndteint.2005.04.003

[16] Wu, Q., He, Q. B., Kong, F. R., Liu, Y. B., and Li, P. Bearing fault diagnosis based on wavelet transform and ICA, *Applied Mechanics and Materials*, **148–149**, 672–675, (2011). https://doi.org/10.4028/www.scientific.net/AMM.148-149.672

[17] Zhang, W., Peng, G., Li, C., Chen, Y., and Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors*, **17**, 425, (2017). https://doi.org/10.3390/s17020425

[18] Lee, K. B., Cheon, S., and Kim, C. O. A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes, *IEEE Transactions on Semiconductor Manufacturing*, **30**, 135–142, (2017). https://doi.org/10.1109/TSM.2017.2676245

[19] Gou, L., Gao, H., Huang, H., He, X., and Li, S.-C. Multifeatures fusion and non-linear dimension reduction for intelligent bearing condition monitoring, *Shock and Vibration*, **16** (6), 895–914, (2016). https://doi.org/10.1155/2016/4632562

[20] Guo, Y., Zhou, Y., and Zhang, Z. Fault diagnosis of multi-channel data by the CNN with the multilinear principal component analysis, *Measurement*, **171**, 108–513, (2021). https://doi.org/10.1016/j.measurement.2020.108513

[21] Shi, M., Cao, Z., Liu, Y., Liu, F., Lu, S., and Li, G. Feature extraction method of rolling bearing based on adaptive divergence matrix linear discriminant analysis, *Measurement Science and Technology*, **32** (7), 075003, (2021). https://doi.org/10.1088/1361-6501/abde72

[22] Liu, S., Ji, Z., Wang, Y., Zhang, Z., Xu, Z., Kan, C., and Jin, K. Multi-feature fusion for fault diagnosis of rotating machinery based on convolutional neural network, *Computer Communications*, **173**, 160–169, (2021). https://doi.org/10.1016/j.comcom.2021.04.016

[23] Wang, Z. *Research on fault diagnosis and condition monitoring method of rolling bearing based on K-L divergence*, Beijing Jiaotong University, (2019).

[24] Li, Y., Bai, L., Jiang, Z., Gao, H., Huang, X., Liu, X., and Huang, M. Identification of caving coal and rock based on EEMD-KPCA and KL divergence, *Journal of China Coal Society*, **45** (02), 827–835, (2020). https://doi.org/10.13225/j.cnki.jccs.2019.0198

[25] Huo, H., Zhang, W., Liu, L., and Li, Y. Collaborative filtering recommendation model based on convolutional denoising auto encoder, *Proc. of the 12th Chinese Conference on Computer Supported Cooperative Work and Social Computing*, (2017). https://doi.org/10.1145/3127404.3127420

[26] Wang, X., Mao, D., and Li, X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network, *Measurement*, **173**, (2021). https://doi.org/10.1016/j.measurement.2020.108518

[27] Guo, Y., Zhou, Y., and Zhang, Z. Fault diagnosis of multi-channel data by the CNN with the multilinear principal component analysis, *Measurement*, **171**, (2021). https://doi.org/10.1016/j.measurement.2020.108513

[28] Bakdi, A., Bounoua, W., Guichi, A., and Mekhilef, S. Real-time fault detection in PV systems under MPPT using PMU and high-frequency multi-sensor data through online PCA-KDE-based multivariate KLD, *International Journal of Electrical Power & Energy Systems*, **125**, (2021). https://doi.org/10.1016/j.ijepes.2020.106457

[29] Zhang, Y., Wang, K., Qian, X., and Gendeel, M. Robust fault-detection based on residual K-L divergence for wind turbines, *IET Renewable Power Generation*, **13** (13), (2019). https://doi.org/10.1049/iet-rpg.2018.6190

[30] Wang, Y., Deng, J.-Z. User similarity collaborative filtering algorithm based on KL divergence, *Journal of Beijing University of Posts and Telecom*, **40** (2), 110–114, (2017). https://doi.org/10.13190/j.jbupt.2017.02.019

[31] Mao, H., Tang, W., Zhu, W., Yang, G., Li, X., Huang, Z., Mao, H., and Si, B. Feasibility study on wheelset fatigue damage with NOFRFs-KL divergence detection method in SIMO, *Journal of Sound and Vibration*, **483**, (2020). https://doi.org/10.1016/j.jsv.2020.115447

[32] Harmouche, J., Delpha,C., and Diallo, D. Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part I, *Signal Processing*, **94**, 278–287, (2014). https://doi.org/10.1016/j.sigpro.2013.05.018

[33] Harmouche, J., Delpha,C., and Diallo, D. Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II, *Signal Processing*, **109**, 334–344, (2015). https://doi.org/10.1016/j.sigpro.2014.06.023

[34] Wang, X. *Fault diagnosis of bearing and planetary gearbox based on data and multi-source information fusion*, University of Electronic Science and Technology, (2020).

[35] Lei, P., Yang, Y., Shi, Y., and Li, W. A method and system for bearing fault prefabrication, Patent Shanxi: CN108593297A, (2018).

[36] Yang, R., Li, H., He, C., and Wang, F. Rolling element bearing incipient fault feature extraction based on optimal wavelet scales cyclic spectrum, *Journal of Mechanical Engineering*, **54** (17), 208–217, (2018). https://doi.org/10.3901/JME.2018.17.208

[37] Sun, X., Liu, H., Zhao, X., and Zhou, B. Singular point recognition and feature extraction for incipient bearing fault based on instantaneous envelope scalogram entropy, *Journal of Mechanical Engineering*, **53** (3), 73–80, (2017). https://doi.org/10.3901/JME.2017.03.073

[38] Tang, G. and Wang, X. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing, *Journal of Xi'an Jiaotong University*, **49** (5), 73–81, (2015).

[39] Yan, B. and Zhou, F. Initial fault identification of bearing based on coherent cumulant stagewise orthogonal matching pursuit, *Journal of Mechanical Engineering*, **34** (9), 149–153, (2015).

[40] Li, H., Liu, H., Xu, F., Zhang, X., and Zhang, X. Method for the optimal continuous wavelet reconstruction scale determination and early fault classification, *Journal of Mechanical Engineering*, **50** (17), 69–76, (2014). https://doi.org/10.3901/JME.2014.17.069

[41] Feng, F., Si, A., Rao, G., and Jiang, P. Early fault diagnosis technology for bearing based on wavelet correlation permutation entropy, *Journal of Mechanical Engineering*, **48** (13), 73–79, (2012). https://doi.org/10.3901/JME.2012.13.073