

Multi-source fault identification based on combined deep learning

Dongqiu Xing^{1,}, Rui Chen², Lihua Qi¹, Jing Zhao¹ and Yi Wang³*

¹Information Communications Academy, National University of Defense Technology, No.5 Guangming Road, Wangqu Town, Chang'an District, Xi'an City, 710106, China

²Faculty of School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, 710121, China

³Xi'an Institute of Applied Optics, No.9 Dianzisan Road, Dianzicheng Town, Yanta District, Xi'an City, 710102, China

Keywords: Multi-source fault, Stacked denoising autoencoder, Multiple classifier combination, Deep learning.

Abstract. This study establishes a multi-source fault identification method based on a combined deep learning strategy to identify a multi-source fault effectively in the fault diagnosis of complex industrial systems. This framework is composed of feature extraction and classifier design. In the first state, the signal is transformed to the time-frequency domain and the time-frequency feature is learned using stacked denoising autoencoders. A learning method that consists of unsupervised pre-learning and supervised fine-tuning is used to train this deep model. In the second state, a model for an ensemble multiple support vector machine classifier is created to recognize fault information. Ten types of rolling bearing signals were adopted in a simulation experiment to validate the effectiveness of the proposed framework. The results demonstrate that the joint model helps to obtain higher recognition accuracy.

1 Introduction

The components of modern complex industrial systems and their interiors generally have many complicated and strongly correlated coupling relationships, many uncertainty factors, and a large amount of uncertainty information, which frequently causes failures because of such characteristics as randomness, subsequence, concurrency, and transmissibility[1]. It is difficult to determine the correlated coupling relationships between constituent units using the traditional fault diagnosis method used for a single piece of equipment, subsystem, or sub-unit, and the selection of an effective feature combination typically depends on the experience of experts. The identification accuracy is also difficult to guarantee. Therefore, it is of great significance to determine a data-based autonomous feature learning method to improve the performance of fault identification. With the rapid development of artificial intelligence technology, deep learning has become a topic of common concern in the field of signal processing.

* Corresponding author: angelaxing@126.com

Deep learning implements automatic feature extraction and complex mapping relationship fitting by constructing a deep model and simulating the brain learning process, and finally depicts rich internal information of data and improves the accuracy of fault identification [2]. Many researchers have attempted to solve problems in the field of fault diagnosis using the theory of deep learning. Van[3] studied vibration signals with transient impact and complex background noise, and used a deep belief nets (DBN) diagnosis model composed of a Bernoulli hidden layer and Gaussian visual unit to realize the ability to process actual operation data and promote its development in the field of fault diagnosis. Shao [4] and Wang [5] adopted DBN to realize the fault diagnosis of rolling bearings and gearboxes, and demonstrated better robustness and accuracy than some mainstream fault diagnosis algorithms. Wei Dong [6] proposed, for the first time, a convolutional neural network (CNN) structure using two Soft-max classifiers. It realized the classification of two kinds of non-independent classification problems and the sharing of weights to a greater extent. Sun [7] and others used a sparse self-coding model with unsupervised feature extraction advantages to learn fault features, and effectively proposed interference terms for feature extraction under the effect of noise reduction coding and improved the robustness of feature representation. Each of these methods has advantages, but the fusion synergy between different models is not strong and the composition of classifier is single. The single classifier and the weak fusion synergy between different models have led to the weak generalization ability of fault diagnosis when system failures or complex failures occur. Based on this, in this paper, a fault identification method is proposed that combines deep learning theory with a multi-classifier. Because of characteristics of the research object that can carry a large number of time-frequency features in the time-frequency domain, the method first transforms the signal using the method of time-frequency analysis in the preprocessing stage. A deep neural network model with multiple hidden layers is created for the detected object to learn its inherent common characteristics so that the system fault state can be identified. Second, at the classifier design stage, a multi-support vector machine (SVM) combination classifier is constructed for classification and recognition training to realize fault location in the system fault state and provide accurate decision information for fault diagnosis decision-making.

2 Feature extraction based on a stacked denoising autoencoder (SDAE)

Under normal conditions, there are few abnormal data in the actual scenario of a complex industrial system, so this status of system should make it difficult to meet the demand for full-mode samples using a data-driven fault diagnosis model. The traditional solution is to accurately depict all normal states using a data model, and a system state will be considered an anomaly if it is different from the normal state provided by the data model. This method requires accurate data modeling. When some prior knowledge is missing or special conditions are insufficient, it often leads to state misjudgment and increases the uncertainty of the diagnosis decision. Therefore, the approach used to build a universal model to adapt to the impact of exogenous variables on the system and determine the inherent characteristics of the system becomes the key to fault anomaly detection.

The basic unit of the stacked autoencoder(SAE) is the autoencoder (AE);the SAE is formed by stacking multiple AEs, cascading several AEs into a multilayer neural network. Its output can be considered as the characteristic representation of input data after multiple dimensionality reduction [8]. Each AE can be regarded as an artificial neural network with a single hidden layer, and output y can reconstruct input x as much as possible by seeking optimal parameter (W, b) . At this time, output of the hidden layer $y_h^{(k,2)}$ can be regarded as the

low-dimensional feature of input x after dimension reduction. Fig. 1 is an architecture diagram of SAE and AE models.

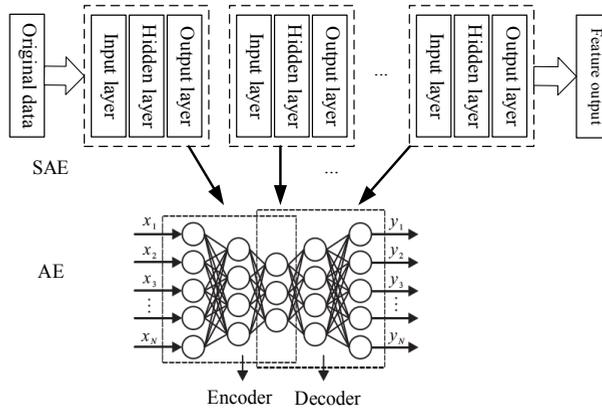


Fig. 1. The architecture of SAE & AE.

Generally, the loss function of AE contains input and output mean square error constraints, weight attenuation constraints, and sparsity constraints, thereby making feature extraction of the hidden layer sparse and robust. Its mathematical description is as follows:

$$\arg \min_{(W, b)} L(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|y^{(i)} - x^{(i)}\|^2 \right] + \frac{\lambda}{2} \sum_{l=1}^{n-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(k,l)})^2 + \beta \sum_{j=1}^{s_2} KL(\rho \| \bar{\rho}_j) \quad (1)$$

where m , n , and S_l represent the number of samples, number of network layers, and number of l -layer units, respectively; $W^{k,l}$ represents the connection weights between the presentation layers; $b_i^{(k,l)}$ represents the unit bias of the l -layer; $\bar{\rho}_i$ represents the average output of the hidden layer units; and

$$\begin{cases} y^{(i)} = f(W^{(k,2)} a^{(k,2)} + b^{(k,2)}) \\ a^{(k,2)} = f(W^{(k,1)} x^{(i)} + b^{(k,1)}) \\ f(z) = 1 / (1 + e^{-z}) \end{cases} \quad (2)$$

AE uses a gradient descent algorithm to train the network parameters to minimize the loss function (1). In the pre-training phase, each AE is trained independently from the lower layer, with the goal of minimizing the error between its output and input. After the lower layer AE training is complete, its hidden layer output is used as the input of the higher layer AE, and the higher layer AE continues to be trained until all AE training is complete. In the global fine-tuning stage, the weights and offsets between the trained AE input layer and the hidden layer are used as the initial weights and offsets, the labels of the data are used as the supervisory signals to calculate the network errors, the Back propagation algorithm is used to calculate the errors of each layer, and the gradient descent method is used to adjust the weights and offsets of each layer. The training process is unsupervised without using input tags. On the basis of the classic AE, some "enhanced" models have been proposed, such as the denoising autoencoder (DAE) [9] and convolution autoencoder (CAE) [10]. Among them, the DAE is a more generalized model because of its ability to learn some "broken" input data; these "broken" data are "cleaned," that is, erased, randomly before input. Therefore, in this

paper, a stacked DAE (SDAE) model that consists of "stacking" multiple DAEs is used for autonomous feature learning of the collected signals.

The main functions of the SDAE model are noise reduction, filtering, and feature extraction. It differs from the fault diagnosis based on DBN and CNN in that both the encoder and decoder in the SDAE model can be used to integrate feature extraction algorithms and classification recognition algorithms [11]. Therefore, the SDAE model adapts to the shortage of fault samples in the actual operation of complex industrial systems, and only needs a small amount of sample data; that is, the SDAE model can achieve higher performance for fault diagnosis coupled with appropriate classification and identification techniques.

3 Fault recognition based on a Multi-SVM classifier

An important link in fault diagnosis is fault location, which can distinguish multi-source faults with a strong correlation through classifiers to realize the diagnosis and identification of fault types and fault locations. The classification performance and identification ability of classifiers directly determine the effect of fault diagnosis. An SVM is widely used in many pattern recognition and data mining tasks. An SVM classifier not only has strong generalization ability but can also map data to high-dimensional space through a kernel function to classify linear inseparable data. Common kernel functions include the linear kernel function, Gaussian kernel function, and polynomial kernel function. Because an SVM classifier based on different kernel functions performs differently on the same training set, the selection of SVM kernel functions often becomes the core component of optimal classification results for a given classification task. However, in the process of fault identification, system faults have the characteristics of randomness, fuzziness, and information uncertainty. The selection of an SVM classifier with a single kernel function clearly affects the accuracy of fault identification for complex industrial systems. To avoid the problem of the kernel function selection of an SVM classifier, an effective method is to combine multiple SVM classifiers. The fault identification framework based on a multi-classifier combination is shown in Fig. 2.

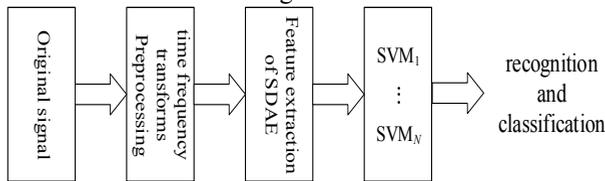


Fig. 2. Framework of multi-classifier combination recognition.

3.1 Multi-classifier combination algorithm

A multi-classifier combination algorithm is a model fusion method. Its basic idea is to assume that different classifiers can provide different side information for the result, and by fusing this information, it can provide a better decision. In [12], the authors proposed a method of the linear integration of deep learning models, which trains a hidden Markov model (HMM) classifier by combining the outputs of different deep neural network models to obtain better results than a single deep neural network model. Inspired by this method, in this paper, a multi-classifier combination model is constructed based on posterior probability, which is described as follows:

Suppose Z represents a classification task in which Z can be assigned to one of C pattern categories. Then the object to be classified has C pattern categories defined by

$$\{\omega_1, \dots, \omega_j, \dots, \omega_c\}$$

Define $P = [p(x_1), \dots, p(x_i), \dots, p(x_N)]^T$ as the true posterior probability of N samples, where $P \in R^{N \times c}$, $p(x_i) \in R^{1 \times c}$. The true posterior probability of the sample x_i is

$$p(x_i) = [p(\omega_1 | x_i), \dots, p(\omega_c | x_i)] \tag{3}$$

Clearly, the posterior probability of sample x_i belonging to a certain category is one; otherwise, it is zero. Assuming M classifiers, the k -th classifier's posterior probability for N samples can be expressed as follows:

$$P_k = [p_k(x_1), \dots, p_k(x_i), \dots, p_k(x_N)]^T \in R^{N \times c} \tag{4}$$

Where $P_k(x_i) \in R^{1 \times c}$. Then the posterior probability of classifier k to the sample x_i is

$$p_k(x_i) = [p_k(\omega_1 | x_i), \dots, p_k(\omega_c | x_i)] \tag{5}$$

More generally, the decision value of a sample belonging to a certain pattern category is obtained by linear combination of the output decision values of multiple classifiers. Depending on this decision value, the combination model can determine what type of category sample is. Therefore, it can be described as:

$$p(\omega_j | x_i) \approx \alpha_1 p_1(\omega_j | x_i) + \dots + \alpha_M p_M(\omega_j | x_i) + b_i \tag{6}$$

Then, the combination model of multiple classifiers can be described as:

$$p_{com}(x_i) = \sum_{k=1}^M W_k p^T(x_i) + b^T \tag{7}$$

where $W_k \in R^{c \times c}$ represents the weight matrix of the classifier and $b^T \in R^{c \times 1}$ represents the offset vector. The parameter estimation of equation (7) is used to minimize the difference between the output value and the real value of the combined model, which can be measured using the least square error, that is,

$$\min_{(W,b)} C = \frac{1}{2N} \sum_i \|p_{com}(x_i) - p^T(x_i)\|^2 \tag{8}$$

To avoid overfitting of the model, after adding the L2 regularization term to the loss function, equation (8) can be rewritten as follows:

$$\min_{(W,b)} C = \frac{1}{2N} \sum_i \|p_{com}(x_i) - p^T(x_i)\|^2 + \frac{1}{2N} \sum_k \gamma_k \|W_k\|^2 \tag{9}$$

where γ_k is the regularization parameter and can be selected by cross-validation. It can be seen that equation (9) is a multivariate function extreme value problem, which can be solved using the least square method. First, by solving the partial derivative of W, b^T , that is,

$$\frac{\partial C}{\partial W} = 0 \quad \text{and} \quad \frac{\partial C}{\partial b^T} = 0 \tag{10}$$

then,

$$\begin{cases} \sum_i \left(\left(\sum_{k=1}^M W_k p_k^T(x_i) + b^T - p^T(x_i) \right) p_1(x_i) \right) \\ \quad + \gamma_1 W_1 = 0 \\ \quad \vdots \\ \sum_i \left(\left(\sum_{k=1}^M W_k p_k^T(x_i) + b^T - p^T(x_i) \right) p_M(x_i) \right) \\ \quad + \gamma_M W_M = 0 \\ \sum_i \left(\sum_{k=1}^M W_k p_k^T(x_i) + b^T - p^T(x_i) \right) = 0 \end{cases} \tag{11}$$

The solution equations can be solved using $[W \quad b^T] = BA^{-1}$. Where the values of A and B are

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & N \end{bmatrix} \tag{12}$$

$$A_1 = \begin{bmatrix} P_1^T P_1 P + P \gamma_1 & \cdots & P_1^T P_M \\ \vdots & \ddots & \vdots \\ P_M^T P_1 P + P \gamma_1 & \cdots & P_M^T P_M + \gamma_M \end{bmatrix} \tag{13}$$

$$A_2 = [\sum_i p_1^T(x_i), \dots, \sum_i p_M^T(x_i)] \tag{14}$$

$$A_3 = [\sum_i p_1(x_i), \dots, \sum_i p_M(x_i)] \tag{15}$$

$$B = [P^T P_1, \dots, P^T P_M, \sum_i p^T(x_i)] \tag{16}$$

Thus, the adaptive solution of the optimal weight matrix and the optimal offset vector of the combined classifier can be obtained using the above method. However, a standard SVM cannot provide the posterior probability output of samples, so it is necessary to determine an SVM posterior probability output method to achieve the purpose of using a multi-SVM classifier combination.

3.2 Posterior probability output of SVM

In pattern classification decision-making, a standard SVM can obtain the decision value that the sample belongs to a certain category, but cannot provide its corresponding posterior probability. However, there is a certain correlation between the decision value and the posterior probability. The larger the absolute value of the decision value of the sample, the higher the reliability that it belongs to a certain category. Therefore, the posterior probability of SVM can be obtained using the method proposed in the literature [13]. The method fits a sigmoid model through training data, which can map the decision value of the SVM to the posterior probability. The mathematical model can be expressed as:

$$P(y = 1 | f) = \frac{1}{1 + \exp(A * f + B)} \tag{17}$$

In Equation (17), f is the decision value output by the SVM; and parameters A and B can be obtained through training. It should be noted that this method is only applicable to the two-class classification problem. For the multi-class classification problem, it can be divided into a plurality of two-class classification problems, and the corresponding posterior probabilities can be obtained and normalized.

4 Experiments

In this section, an ensemble of multi-SVM classifiers is used for pattern recognition experiments. It consists of a linear combination of the linear kernel-based SVM (labeled SVM_LK), polynomial kernel-based SVM (labeled SVM_PK), and Gaussian kernel-based SVM (labeled SVM_GK). And the simulation table is shown in Fig.3.

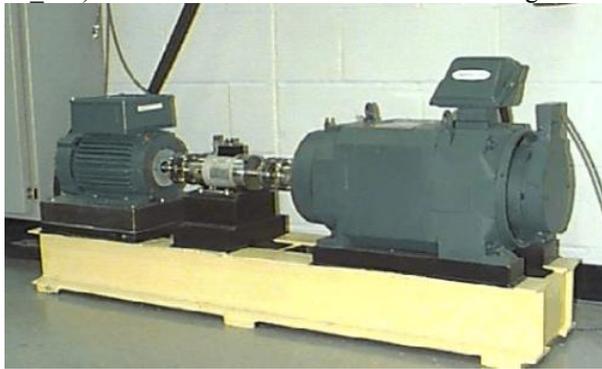


Fig. 3. Rolling bearing simulation table.

The experiments simulated four bearing operation states: normal (NOR), rolling element fault (Ball), inner ring fault (IR), and outer ring fault (OR). Each mild, moderate, and severe fault was simulated for three diameters: 0.018 mm, 0.036 mm, and 0.053 mm. The 10 states of the test platform were classified and identified, and the experimental conditions are described in Table 1.

Table 1. Description of the experimental conditions.

Type	Fault location	Fault size/mm	Bearing condition
1	NOR	0	normal
2	Ball	0.018	mild rolling fault
3	Ball	0.036	Moderate rolling fault
4	Ball	0.053	Severe rolling fault
5	IR	0.018	mild inner ring fault
6	IR	0.036	Moderate inner ring fault
7	IR	0.053	Severe inner ring fault
8	OR	0.018	mildouter ring fault
9	OR	0.036	Moderate outer ring fault
10	OR	0.053	Severe outer ring fault

The platform environment for the experiments was Matlab2017A + LibSVM3.12, and the SDAE model was built using the Deeplearntoolbox toolbox. The settings were an empty load, the rotating shaft speed was 1792r/min, sampling frequency of the acceleration sensor was 48kHz, and sampling time was approximately 5s.

Clearly, the data of bearing rotation one week is about 1600 data points. During the test, eight groups ($1600/200 = 8$) of data was set for different sample lengths, with a step size of 200 data points. The time-domain waveforms of the ten state of bearing was shown in Fig.4.

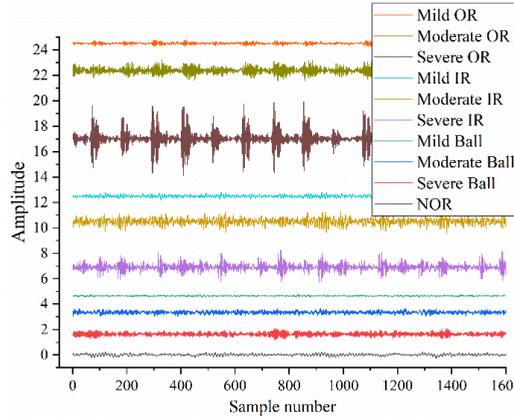


Fig. 4. The time-domain waveforms of different bearing state.

A training set that contained 50 samples and a test set that contained 20 samples were randomly selected from each group during the test. For each group of data, by selecting the training sample set and test sample set from the 10 types of data in Table 1, a training set that contained a total of 500 samples and a test set that contained a total of 200 samples were finally obtained. After the data was simply normalized, it was input into the hidden layer of the fault identification model designed in this study.

4.1 Experiments on the SDAE model

In this section, to verify the effectiveness of SDAE feature learning, it was compared with bi-dimensional empirical mode decomposition (BEMD) [14] and multi-scale permutation entropy (MSPE) [14] feature extraction methods, and Gaussian noise of -5dB, -4dB, -3dB, and -2dB was added to verify the feature recognition capability of different methods under different noise backgrounds. There are 6 feature recognition frameworks were selected: BEND+SVM_LK, BEMD+SVM_GK, MSPE+SVM_LK, MSPE+SVM_GK, SDAE+SVM_LK and SDAE+SVM_GK. The SDAE model with three hidden layers was constructed. The parameters of the model were set using the default values of the Deeplearn Toolbox toolbox. The average classification accuracy of each method is shown in Table 2.

Table 2. Comparison of different feature extraction methods (%).

Method of Feature extraction	Noise/dB			
	-5	-4	-3	-2
BEMD+SVM_LK	71.35	82.86	90.34	94.45
BEMD+SVM_GK	70.67	81.57	89.12	93.67
MSPE+SVM_LK	70.88	81.78	89.38	93.75
MSPE+SVM_GK	70.65	81.35	88.83	93.18
SDAE+SVM_LK	73.33	83.48	91.56	96.68
SDAE+SVM_GK	72.58	83.20	91.27	96.53

The results in Table 2 show that SDAE model shows a good feature learning ability under strong background noise, and accuracy of the linear kernel-based SVM classifier is better than the Gaussian kernel-based SVM classifier.

4.2 Experiments on the Multi-SVM classifier

To verify the recognition rate effect of the multi-SVM classifier, the output value of the last hidden layer of the SDAE model was used as the input feature of each comparison classifier (SVM_LK, SVM_PK, SVM_GK and Soft-max), where the default values of the LibSVM toolbox were used for parameter setting; and the three SVM linear combinations were compared with each single classifier. The experimental results are shown in Table 3.

Table 3. Comparison of different SVM classification accuracy rate (%).

Classifier model	Noise/dB			
	-5	-4	-3	-2
SDAE+SVM_LK	73.33	83.48	91.56	96.68
SDAE+SVM_PK	72.78	83.22	91.48	96.61
SDAE+SVM_GK	72.58	83.20	91.27	96.53
SDAE+Soft-max	71.66	82.92	91.15	96.32
Methods adopted in this paper	74.23	85.53	93.46	98.58

As can be seen from Table 2 and Table 3, the SVM based on linear kernel function has improved the classification accuracy to a certain extent compared with the traditional Soft-max classifier. At the same time, it is easy to see that SVM_PK, SVM_LK and SVM_GK are very similar in classification performance, among which SVM_LK has the best classification performance. Compared with SDAE + Soft-max model, the combination model in this paper can improve the classification accuracy by about 2 percentage points, and the combination model can also achieve better performance than the method using single SVM classifier.

Finally, to further compare the performance of each classification model. Under the same noise environment, the classifiers based on different kernel functions were tested with different test sample lengths. For convenience, they are labeled LK, PK, and GK, and the method in this article is labeled MSVM. The sample length increased from zero to 1,600 according to the previously set of eight groups of samples. The relationship between the recognition rate and number of samples is shown in Fig. 5.

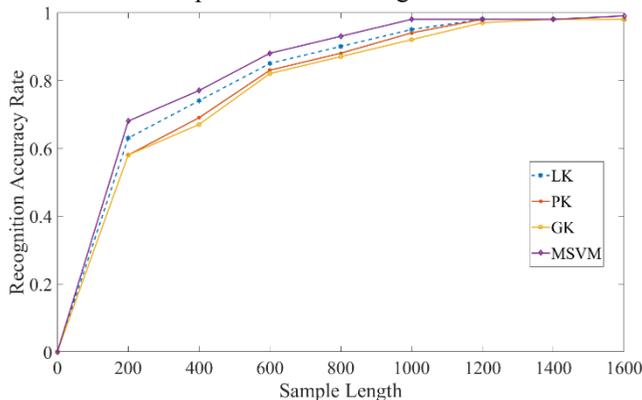


Fig. 5. 5Recognition accuracy of different SVM classifications.

As can be seen from Fig. 2, when the number of samples was small, the classifier recognition rate was low, which could be caused by the fact that the original signal contained all the associated signals, whereas the effective fault feature signal occupied only a small portion, and valuable fault information was seriously disturbed or masked by other information after the original signal was time-frequency converted, thereby making it difficult to classify correctly. However, with the increase of the number of test samples, when the sample length increased to a certain extent, the sample length had little influence on the correct identification.

5 Conclusion

In this paper, a multi-source fault identification method based on a hybrid deep learning model was proposed, which was applied to the classification and identification of bearing faults and verified by 10 common bearing state signals. The results demonstrate that, first, compared with the traditional vibration signal feature extraction method, the SDAE model-based feature extraction method obtains higher accuracy. Second, the multi-SVM combined classifier model can further improve signal recognition performance.

Additionally, deep learning can directly mine valuable information from existing data, whereas it is difficult to explain the physical meaning of feature parameters extracted based on the deep neural network model and training of the model requires a large amount of labeled data, so complete data collection is required to support the model. For the multi-sensor monitoring data of complex industrial systems, the approach used to deeply fuse the statistical model with the fault mechanism model is a research direction to be further studied.

References

1. LEI Yaguo, et al. "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol.63, Issue: 5, pp.3137-3147, May 2016.
2. Feng Jia, et al. "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical System and Signal Processing*, Vol. 72–73, pp. 303-315, May 2016.
3. Van Tung Tran, Faisal Al Thobiani, Andrew Ball. "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, Vol. 41, Issue 9, pp. 4113-4122, July 2014.
4. Haidong Shao, Hongkai Jiang, Xun Zhang, et al. "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, Vol. 26, Issue 11, article id. 115002, November 2015.
5. Wang Xingqing, Li Yanfeng, Ting Rui, et al. "Bearing fault diagnosis method based on Hilbert envelope spectrum and deep belief network," *J of Vibroengineering*, vol.17, Issue 3, pp. 1295-1308, May 2015.
6. Wei D, Gong Q W, Lai W Q, et al. "Research on internal and external fault diagnosis and fault-selection of transmission line based on convolutional neural network," *Proc of the Chinese Society for Electrical Engineering*, Issue 36, pp. 36: 21-28, Sep.2016.
7. Wenjun Sun, Siyu Shao, Rui Zhao, et al. "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, Vol.89, pp. 171-178, July 2016.

8. Martin Langkvist, Lars Karlsson, AmyLoutfi. "A review of unsupervised feature learning and deep learning for time-series modeling," Pattern Recognition Letters, Vol. 42, Issue 1, pp.11-24, June 2014.
9. VINCENT P, LAROCHELLE H, LAJOIE I, et al. "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," Journal of Machine Learning Research, Vol.11, Issue 12, pp.3371-3408, Jan. 2010.
10. MASCI J, MEIER U, CIREŞAN D, et al. "Stacked convolutional auto-encoders for hierarchical feature extraction," Proc. Of the 21st International Conference on Artificial Neural Networks, Espoo, Finland, Vol. Part I, pp. 52-59, June 2011.
11. Hoyeop Lee, Youngju Kim, Chang Quk Kim. "A deep learning model for robust wafer fault monitoring with sensor measurement noise," IEEE Trans on Semiconductor Manufacturing, Vol.30, Issue1, pp.23-31, Feb.2017.
12. DENG L, PLATT J C. "Ensemble deep learning for speech recognition," Proc. of the Annual Conference of the International Speech Communication Association, Singapore, pp.1915-1919, Sep. 2014.
13. PLATT J C. "Probabilistic Outputs for support vector machines and comparisons to regularized likelihood methods," Advances in Large Margin Classifiers, Vol.10, Issue 4, pp. 61-74, Mar. 2000.
14. Titijaronroj, etal. "Iteration-Free Bi-Dimensional Empirical Mode Decomposition and Its Application," IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS, Vol.E100D, Issue 9, pp.2183-2196, Sep. 2017.