

Research Article

A Stacked Autoencoder-Based Deep Neural Network for Achieving Gearbox Fault Diagnosis

Guifang Liu,¹ Huaiqian Bao ,^{1,2} and Baokun Han¹

¹College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China

²AUCMA Company Limited, Qingdao 266510, China

Correspondence should be addressed to Huaiqian Bao; bhqian@163.com

Received 19 March 2018; Accepted 12 June 2018; Published 25 July 2018

Academic Editor: Muddasser Alam

Copyright © 2018 Guifang Liu 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.

Machinery fault diagnosis is pretty vital in modern manufacturing industry since an early detection can avoid some dangerous situations. Among various diagnosis methods, data-driven approaches are gaining popularity with the widespread development of data analysis techniques. In this research, an effective deep learning method known as stacked autoencoders (SAEs) is proposed to solve gearbox fault diagnosis. The proposed method can directly extract salient features from frequency-domain signals and eliminate the exhausted use of handcrafted features. Furthermore, to reduce the overfitting problem in training process and improve the performance for small training set, dropout technique and ReLU activation function are introduced into SAEs. Two gearbox datasets are employed to conform the effectiveness of the proposed method; the result indicates that the proposed method can not only achieve significant improvement but also is superior to the raw SAEs and some other traditional methods.

1. Introduction

As one of the most vital components of rotating machinery, gearboxes are widely used in various industrial fields, such as vehicles and machine tools [1]. Due to the complexity structure of gearbox and various working conditions interference, gears always cause fault quite easily. If the fault is not detected in time, it can result in the crash of entire system and serious loss of property. So it is challenging to conduct effective fault diagnosis system. In modern industries, data-driven systems have revolutionized manufacturing through enabling computers to collect a massive amount of data from monitored machines [2]. At the same time, machines have also been more precise than ever before, and machinery fault diagnosis has sufficiently embraced multifault diagnosis revolution in condition monitoring system. Contrasted with top-down modeling proposed by the physics-based fault diagnosis systems, data-driven systems provide a bottom-up model to detect the occurrence of machinery faults [3]. As is well-known, the physics-based methods are unable to be updated online with measured data and also cannot deal well with large-scale data. On the other hand, with the

fast-developing computer systems and sensors, data-driven based fault diagnosis systems have drawn increasing public attentions.

The basic framework of data-driven system usually consists of four consecutive stages: data acquisition, feature extraction, model training, and model testing [4]. Conventional data-driven methods are usually trying to design a right set of features and then put them into some shallow machine learning models such as Naive Bayes (NB) [5], Support Vector Machines (SVM) [6], and logistic regression [7]. But these works usually focus on manual feature extraction (statistical features, frequency, and time-frequency features) that always need plenty of human labor and cannot update online [8]. Meanwhile, the selection of these features could not leave prior knowledge and feature engineering. So it is really a tough problem for these methods to extract intrinsic features behind the raw time-series data. As the hottest subfield of machine learning, deep learning has been regarded as a powerful solution for the intelligent fault diagnosis system to extract salient features through multilayer architecture, such as artificial neural networks (ANN) [9, 10], autoencoders [11, 12], restricted Boltzmann

machine (RBM) [13, 14], and convolutional neural networks (CNN) [15, 16]. Compared with traditional methods, deep learning methods do not need human labor and expert knowledge for feature extraction. All the hyperparameters in model training and pattern classification modules are able to be trained jointly. Therefore, deep learning can be employed to address machinery fault diagnosis in a very general way.

As one of the widely used deep learning techniques, stacked autoencoders (SAEs) have attracted considerable attention in fault diagnosis. It has been investigated as a common component of DNN by Bengio et al. [17]. Jia et al. [18] proposed a SAEs based DNNs for roller bearing and planetary gearbox fault diagnosis with input as frequency spectra after Fourier transform. Guo et al. [19] employed multidomain statistical features of the raw vibration signals as the input of SAEs, which can be viewed as a kind of feature fusion. Liu et al. [20] fed the normalized spectrograms created by STFT into SAEs for rolling bearing fault diagnosis. In the work presented in [21], the nonlinear soft threshold approach and digital wavelet frame were used to process the measured signal and then fed into SAEs for rotating machinery diagnosis. Jia et al. [22] constructed a local connection network based on normalized sparse autoencoder, and L1 norm was employed to find sparse features.

Inspired by the prior researches, a new framework based on SAEs is proposed to resolve the gearbox fault diagnosis. Furthermore, to overcome the deficiency of overfitting problem in the training process and improve the performance for small training set, dropout technique [23] and ReLU activation function are introduced into SAEs. Rest of the paper is organized as follows. Section 2 briefly introduces the algorithms of SAEs, dropout, and ReLU activation function. Section 3 is dedicated to detailing the content of the proposed method. In Section 4, the multifault gearbox dataset is adopted to validate the effectiveness of the proposed method. Furthermore, the superiority of the proposed method is exhibited by comparing with the other traditional methods. Finally, some conclusions are drawn in Section 5.

2. Theoretical Background

2.1. Stacked Autoencoders. Autoencoder is a kind of unsupervised learning structure that owns three layers: input layer, hidden layer, and output layer as shown in Figure 1. The process of an autoencoder training consists of two parts: encoder and decoder. Encoder is used for mapping the input data into hidden representation, and decoder is referred to reconstructing input data from the hidden representation. Given the unlabeled input dataset $\{x_n\}_{n=1}^N$, where $x_n \in \mathbb{R}^{m \times 1}$, h_n represents the hidden encoder vector calculated from x_n , and \hat{x}_n is the decoder vector of the output layer. Hence the encoding process is as follows:

$$h_n = f(W_1 x_n + b_1) \quad (1)$$

where f is the encoding function, W_1 is the weight matrix of the encoder, and b_1 is the bias vector.

The decoder process is defined as follows:

$$\hat{x}_n = g(W_2 h_n + b_2) \quad (2)$$

where g is the decoding function, W_2 is the weight matrix of the decoder, and b_2 is the bias vector.

The parameter sets of the autoencoder are optimized to minimize the reconstruction error:

$$\phi(\Theta) = \arg \min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(x^i, \hat{x}^i) \quad (3)$$

where L represents a loss function $L(x, \hat{x}) = \|x - \hat{x}\|^2$.

As shown in Figure 2, the structure of SAEs is stacking n autoencoders into n hidden layers by an unsupervised layer-wise learning algorithm and then fine-tuned by a supervised method. So the SAEs based method can be divided into three steps:

(1) Train the first autoencoder by input data and obtain the learned feature vector;

(2) The feature vector of the former layer is used as the input for the next layer, and this procedure is repeated until the training completes.

(3) After all the hidden layers are trained, backpropagation algorithm (BP) is used to minimize the cost function and update the weights with labeled training set to achieve fine-tuning.

2.2. Dropout. Dropout is an effective strategy that has been proved to reduce overfitting in the training process of neural networks. The overfitting problem always happens when the training set is small, which would result in a low accuracy on the test set. Dropout can randomly affect the neurons of the hidden layer to lose power in the training process as shown in Figure 3, but the weights of those neurons are preserved. Furthermore, the neurons can recover to work when the next sample is input. Technically, dropout is able to be achieved by setting the output date of some hidden neurons to 0 and then these neurons cannot be related to the forward propagation process. Many researches have tested the effect of dropout on reducing the overfitting problem for the small training set [28], and this paper will also employ it to enhance the feature extraction ability and classification accuracy of SAEs for multifault gearbox fault diagnosis.

2.3. ReLU. For traditional activation functions (sigmoid and hyperbolic tangent functions), the gradients decrease quickly with training error propagating to forward layers. The rectified linear units (ReLU) activation function has received extensive attention in recent years, since its gradient will not decrease with the independent variables increasing. So the network with ReLU does not suffer from gradient diffusion or vanishing. The ReLU function is shown in (4) and the structure is displayed in Figure 4.

$$f_r(x) = \max(0, x) \quad (4)$$

3. Proposed Framework

This section details the proposed intelligent fault diagnosis method. In the method, SAEs are combined with dropout to achieve multifault gearbox fault diagnosis. The framework and illustration of the proposed method are displayed in Figure 5. SAEs combined with dropout model are applied to train

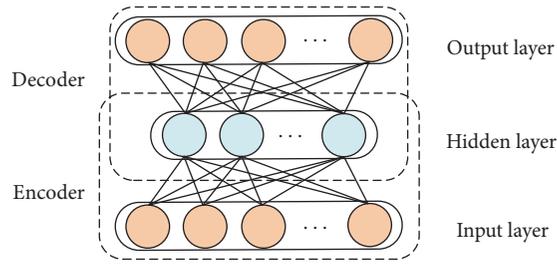


FIGURE 1: Structure of autoencoder.

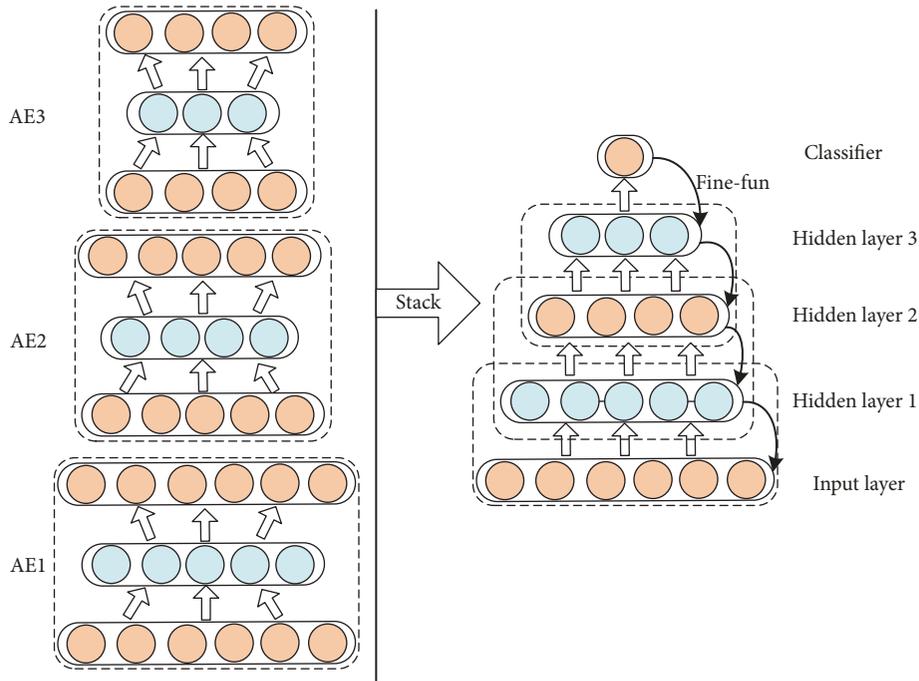


FIGURE 2: Structure of stacked autoencoders.

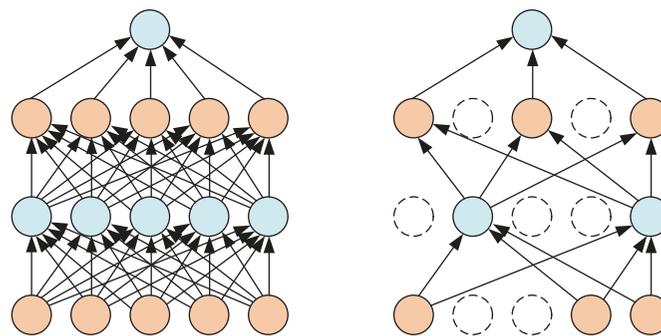


FIGURE 3: Dropout neural net model. Left: a standard neural net. Right: an example of a thinned net produced by applying dropout.

the weight matrix from frequency spectra of vibration signals. Specifically, the procedure can be described as follows:

(1) The spectra of vibration signals are composed the training set $\{X^i, l^i\}_{i=1}^M$, where M is the number of samples, $X^j \in \mathbb{R}^{N \times 1}$ is the i th sample containing N Fourier coefficients, and l^i is the health label of X^i .

(2) Build the DNNs by SAEs, and then employ the unlabeled training set $\{X^i\}_{i=1}^M$ to pretrain the DNNs layer-by-layer.

(3) Utilize BP algorithm to update the weights and fine-tun the parameters of the SAEs with labeled training set $\{X^i, l^i\}_{i=1}^M$.

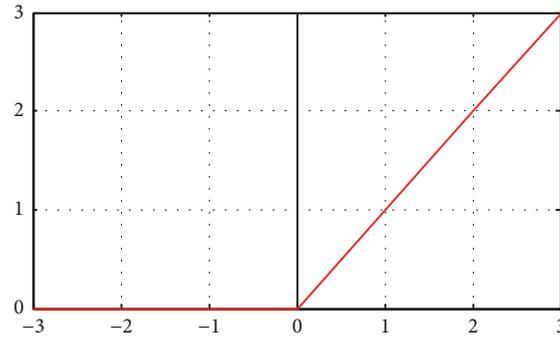


FIGURE 4: Structure of ReLU function.

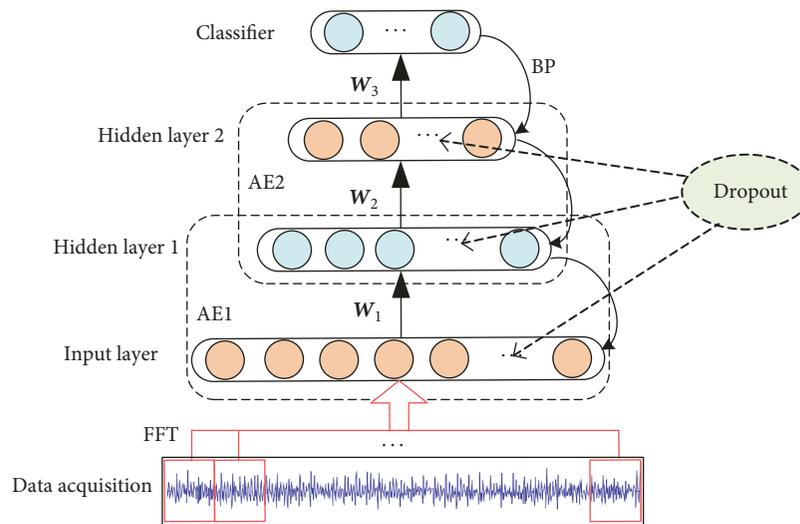


FIGURE 5: Flowchart of the proposed method.

(4) The testing set is adopted to validate the effectiveness of the proposed method.

4. Experiments

4.1. Case 1: Fault Diagnosis of a Multifault Gearbox

4.1.1. Data Description. Gear faults including distributed fault (worn) and localized faults (broken, pit), as well as coupled fault in power train, perhaps cause catastrophic accidents. Therefore, an early recognition of the gear faults is critical for normal operation of a gearbox. Our paper focuses on investigating the multifault gearbox. In this section, a multifault gearbox experimental dataset is employed to validate the effectiveness of the proposed method [29]. The vibration signals were collected on a specially designed bench which consisted of a one phase input and three-phase output motor (the nominal power is 0.75 kW and nominal rotation frequency is 880 rpm), a gearbox, the shaft supporting seats, a flexible coupling, and a magnetic powder brake as show in Figure 6. The sensor is a piezoelectric accelerometer (DH131E) mounted on the flat surface of gearbox and the sampling frequency is 5120 Hz. The gearbox includes two

gears (pinion and wheel gear) and the gear parameters are displayed in Table 1. There are six health conditions under three loads: normal, a single worn pinion, a single pit of wheel, a single broken tooth of wheel, coupled fault of broken wheel and worn pinion, and coupled fault of wheel pit and worn pinion. For brevity, the six fault types of gear are named as Type-1, Type-2, Type-3, Type-4, Type-5, and Type-6, respectively. 100 data samples are collected from each fault type under one load by an overlapped manner, so a total of 1800 samples are obtained from the designed bench and each sample contains 1000 data points. Considering the rotation frequency of shaft is 880 rpm, so each period of rotation contains 350 data points. For avoiding the influence of speed fluctuation, each sample collects almost three periods of rotation data (1000 data points). The frequency spectra are also adopted as input data, and each sample contains 500 Fourier coefficients. The major reason of using frequency spectra is that the frequency spectra can show the distribution of constitutive components with discrete frequencies and more clarity information about the state of rotating machines [18]. Here we randomly select 4 samples from the normal type of gear, and obtain their Fourier coefficients by FFT as shown in Figure 7. It is easy to find that the time-domain features of

TABLE 1: Gear parameters.

Gear	Teeth	Module(mm)	Pressure angle (deg.)	Materials
Pinion	55	2	20	S45C
Wheel	75	2	20	S45C

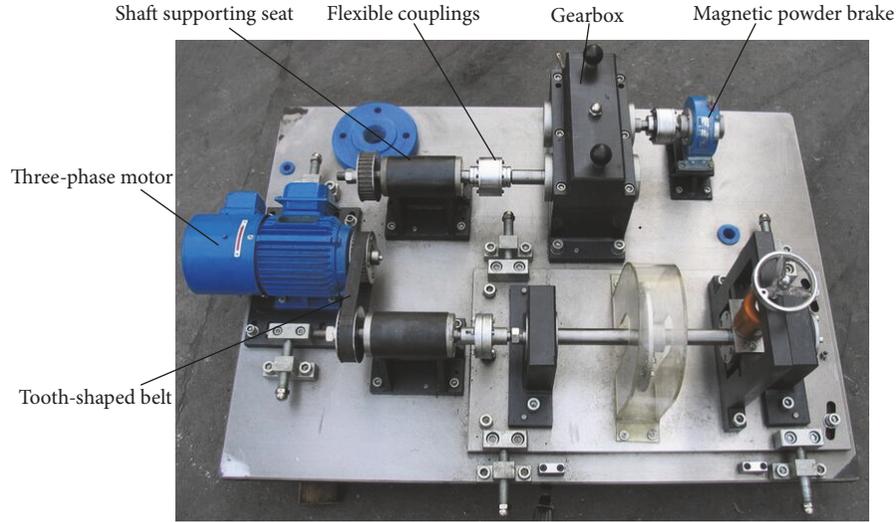


FIGURE 6: Bench of multifault gearbox.

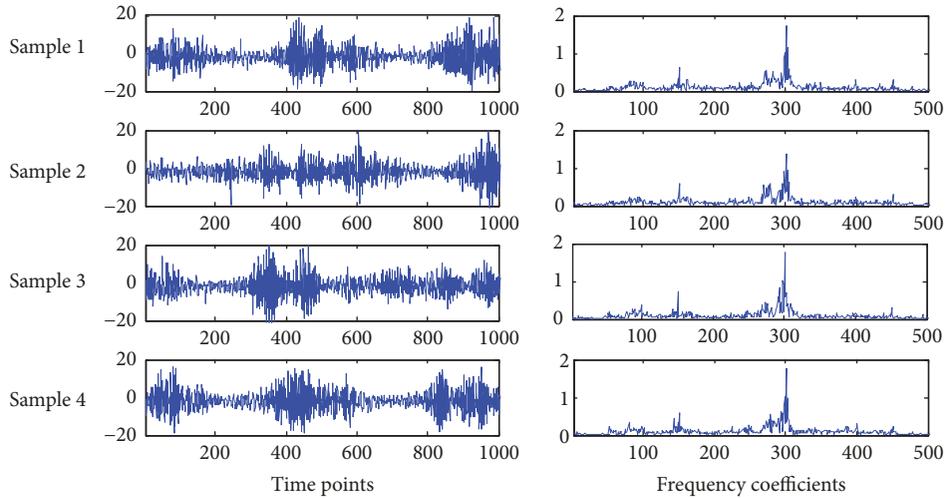


FIGURE 7: Comparison of time-domain and frequency-domain coefficients.

each sample are different, but the frequency spectra features are becoming regularity with each other. The structure of the designed DNNs is 500, 200, 100, and 6, respectively.

4.1.2. Diagnosis Results. The parameter of dropout rate α is changed from 0 to 0.7 with a step size of 0.1, and 15 trials are carried out for the experiment in order to reduce the effective of randomness. 10% of samples are randomly selected to train the model, and the rest are used for testing. The diagnosis accuracies are shown in Figure 8. It is clearly seen that when α is 0.3, the diagnosis accuracy is the highest and the standard

deviation is the lowest, so 0.3 is chosen as the dropout rate in this experiment.

To classify the six health conditions of the gears, 10% samples are employed to train the proposed model and the rest are used for testing. The learning rate is 0.01 and the iteration number is 100. The training and testing accuracies of 15 trials are displayed in Figure 9 and the average training and testing accuracies are 100% and $99.34\% \pm 0.25\%$ respectively, which indicates that the proposed model can also distinguish the six health conditions of gear with a high accuracy. To illustrate the process concretely, the classification results of

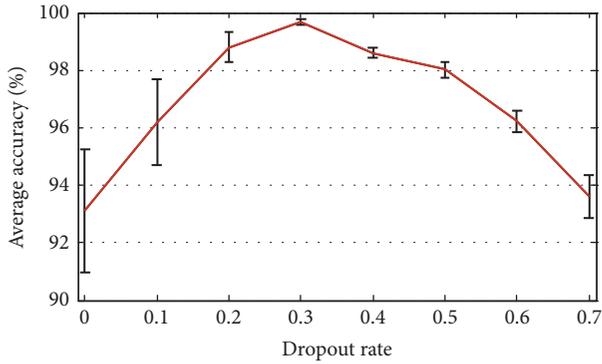


FIGURE 8: Diagnosis results using different learning rates.

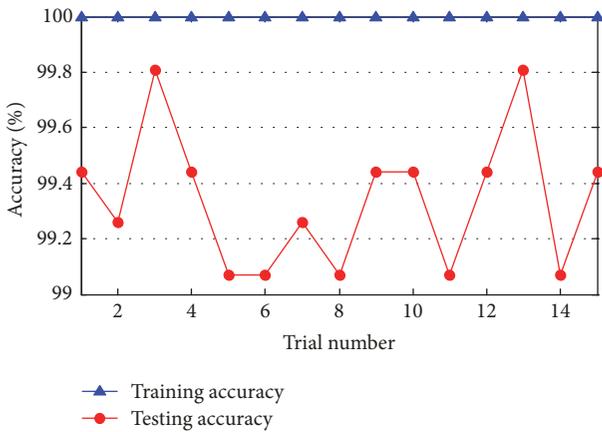


FIGURE 9: Diagnosis result using the proposed method.

the 14th testing trial are drawn in Figure 10. It can be seen that 3 of the testing samples are misclassified, yielding the success rates 99.44%. Among them, 1 sample of type-3 is misclassified as type-4, 2 samples of type-4 are misclassified as type-3, and 1 sample of type-5 is misclassified as type-6, respectively. To further display the ability of the proposed method, t-distributed stochastic neighbor embedding (t-SNE) [30] is employed to visualize the learned features. Therefore, the 100-dimension feature vector is embedded into a 3-dimension feature vector. The classification result is shown in Figure 11. It is easy to find that the same types of samples are gathered together and different types are separated excellently.

For comparison, several diagnosis methods are presented and the diagnosis results are displayed in Table 2. Li et al. [24] proposed a method combining 19 time-domain and frequency-domain features with self-organizing map, when their method was adopted to classify the six types of the gearbox dataset and achieved $92.51\% \pm 4.23\%$ testing accuracy. In [25], wavelet multifractal features and SVM model were used to represent the six gear fault types, and finally obtained $87.65\% \pm 5.37\%$ classification accuracy. Lin et al. [26] proposed a fault diagnosis method using multifractal detrended fluctuation (MFDFA) and here achieved $96.23\% \pm 1.69\%$ accuracy. Lou et al. [27] applied multiple domain features and ensemble fuzzy ARTMAP neural networks to distinguish the health conditions and $98.83\% \pm 0.42\%$

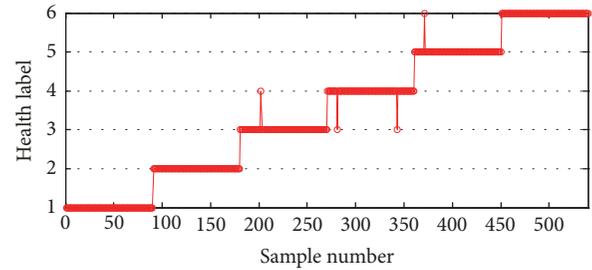


FIGURE 10: Diagnosis result of the 14th trial.

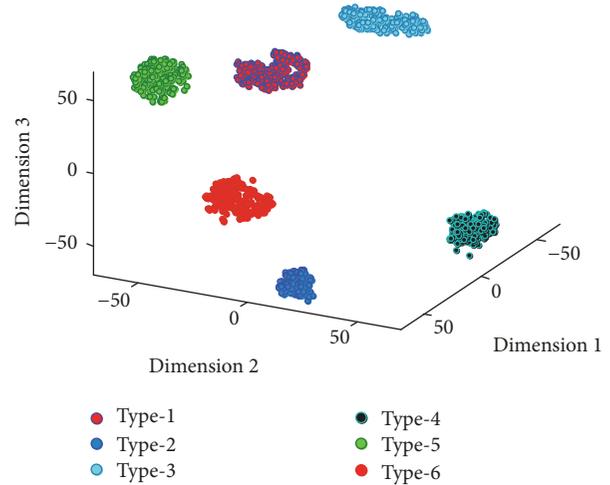


FIGURE 11: Feature visualization map.

accuracy was obtained. Furthermore, the raw stacked autoencoders without dropout (Raw SAEs) are also adopted for comparison, and the testing accuracy is just $93.16\% \pm 3.78\%$ which exhibits the effectiveness of dropout in feature extraction. Compared with the methods above, it shows that the proposed method can not only automatically distinguish the six health conditions of gearbox, but also achieve a higher accuracy with a lower percentage of training samples.

To further investigate the learned features in the proposed model, another experiment is conducted as shown in Figures 12 and 13. As a result, two level features can be obtained from two hidden layers of the DNNs which can be called learned features, so 200- and 100-dimensional learned feature vectors of each sample are obtained, respectively. For achieving a good view on the visualization, all the learned feature vectors of the same type test samples are gathered together [31]. Figure 12 displays the learned 100-dimensional features of gearbox dataset using the raw SAEs without dropout, and Figure 13 shows the learned 100-dimension features using the proposed method. It can be clearly seen that all the learned feature vectors of one health condition by the proposed method are almost the same trend with each other. By contrast, the feature vectors of each health condition by the raw SAEs method are mixed with each other and can not find a unit tendency. And also, the amplitudes of feature vectors by the two methods are different, the proposed method is able to learn more distinguished feature vectors that own

TABLE 2: Comparison of classification accuracy.

Method	Training samples	Testing accuracy
[24]	40%	92.51% \pm 4.23%
[25]	75%	87.65% \pm 5.37%
[26]	N/A	96.23% \pm 1.69%
[27]	50%	98.83% \pm 0.42%
Raw SAEs	10%	93.16% \pm 3.78%
Proposed	10%	99.34% \pm 0.25%

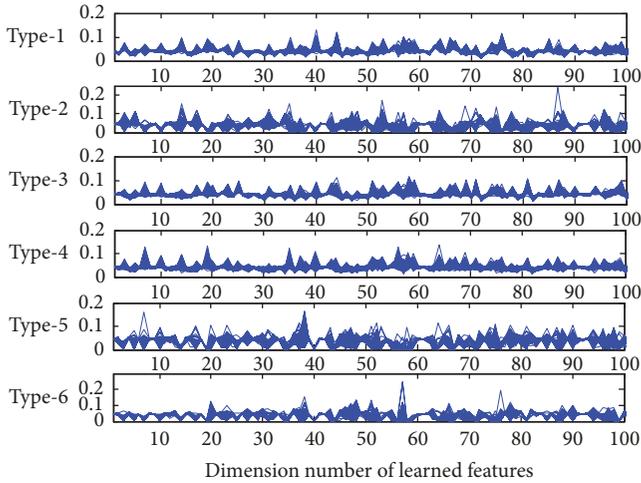


FIGURE 12: Learned features of gearbox dataset using the raw SAEs.

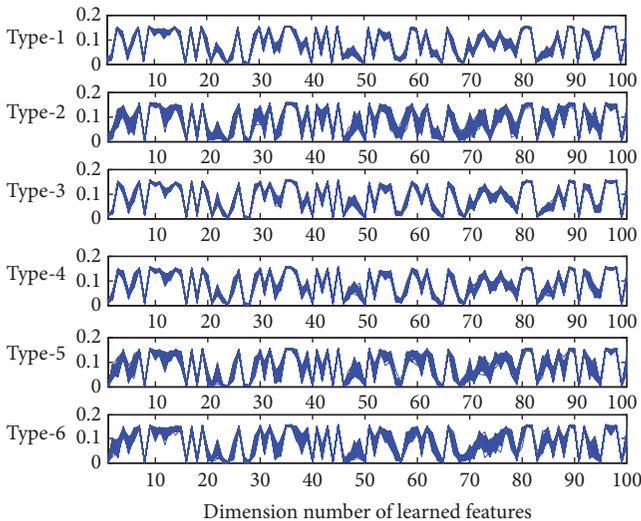


FIGURE 13: Learned features of gearbox dataset using the proposed method.

larger amplitudes than the raw SAEs method. Therefore, the proposed method can effectively mine the main variations in a high order space from different fault signals than the raw SAEs method without dropout.

4.2. Case 2: Fault Diagnosis of a Motorcycle Gearbox. To further validate the proposed method, a motorcycle gearbox dataset [32] is taken as the second case analysis. Figure 14 displays the gearbox being referred to; besides the gearbox, there are an electrical motor with the rotation speed 1420 rpm, a data acquisition system, a tachometer, a triaxial accelerometer, and a load mechanism. The sample frequency was 16384 Hz. There are four health types of gear as shown in Figures 14(b)–14(c): normal condition (NC), slightly worn (SW), medium worn (MW), and broken tooth (BT). In Figure 14(e), the gears, which have 24 teeth and 29 teeth (tested gear), are a pair of driven and driving gears. The vibration signals of four gear health types are depicted in Figure 15. It can be easy to find that the NC and BT are easily distinguished, but the SW and MW are hardly discriminated. 50 samples of NC and 100 samples of SW, MW, and BT are collected, and each sample contains 1000 data points.

Similarly, 10% samples are employed to train the proposed model and the rest are used for testing, and the parameter set is the same as Case 1. The training and testing accuracies of 15 trials are displayed in Figure 16 and the average training and testing accuracies are 100% and 99.26% \pm 0.41% respectively, which also indicates that the proposed model can distinguish the four health types of the motorcycle gearbox with a high accuracy. Then, the classification results of the 3rd testing trial are displayed in Figure 17. Only 2 samples are misclassified; i.e., 1 sample of SW is misclassified as MW, and 1 sample of MW is misclassified as SW, respectively. Meanwhile, the visualization of the 2-dimensional feature vectors mapped by t-SNE is shown in Figure 18. The excellent classification result is also obtained which illustrates the robustness of the proposed method.

5. Conclusions

An intelligent fault diagnosis method based on SAEs is presented for gearbox fault diagnosis. In order to reduce overfitting problem and improve the performance of traditional SAEs for small training set, the dropout technique and ReLU activation function are both adopted. As illustrated in the experimental study, the proposed method can extract useful features from different fault gear types and achieve a high diagnosis accuracy. Comparison studies show that the proposed method outperforms the raw SAEs method and some other traditional methods. On the other hand, the

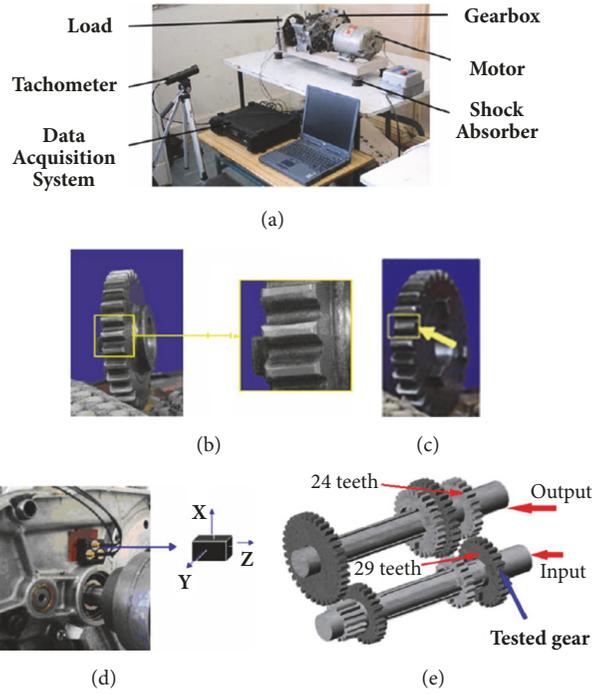


FIGURE 14: (a) Experimental setup; (b) worn teeth; (c) broken teeth; (d) accelerometer location; (e) schematic of the gearbox.

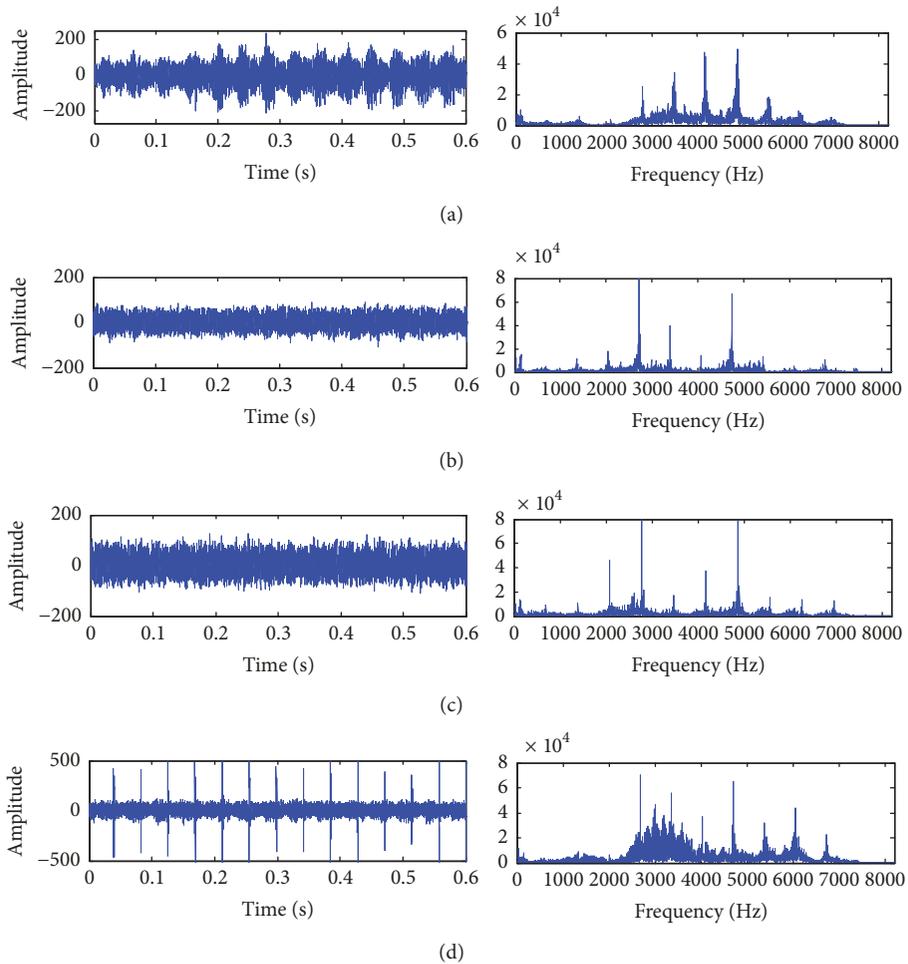


FIGURE 15: Vibration data and corresponding frequency spectra of the four different types of gear conditions: (a) NC; (b) SW; (c) MW; and (d) BT.

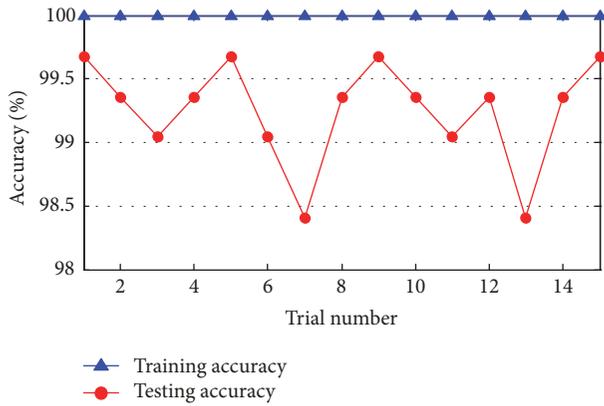


FIGURE 16: Diagnosis result using the proposed method.

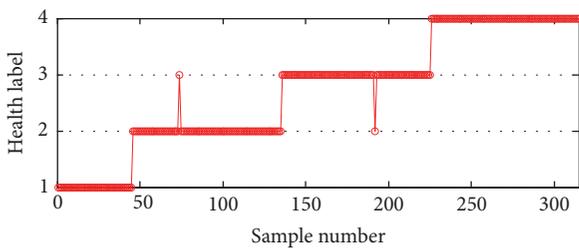


FIGURE 17: Diagnosis result of the 4th trial.

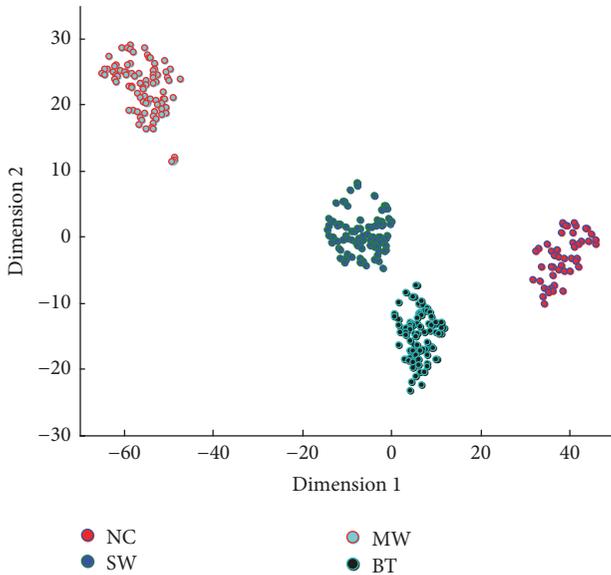


FIGURE 18: Feature visualization map.

exhibition of the learned features illustrates that with the help of dropout technique and ReLU activation function, the proposed method can capture salient features and obtain a higher diagnosis result than the raw SAEs method. Meanwhile it can clearly describe the process on how DNNs deal with mechanical signals, which is worth further study in fault diagnosis. In future work, a wide range of experiments will be investigated to evaluate the robustness of the proposed method.

Data Availability

The data used to support the findings of this study can be found in the online versions at <https://doi.org/10.1006/mssp.2000.1338> and <http://dx.doi.org/10.1016/j.ymsp.2016.06.012>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant no. 51675315) and Shandong Province key research and development projects (2017GGX40120).

References

- [1] X. Jiang, S. Li, and Q. Wang, “A study on defect identification of planetary gearbox under large speed oscillation,” *Mathematical Problems in Engineering*, vol. 2016, Article ID 5289698, 18 pages, 2016.
- [2] S. Yin, X. Li, H. Gao, and O. Kaynak, “Data-based techniques focused on modern industry: an overview,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 657–667, 2015.
- [3] J. Gao, L. Wu, H. Wang, and Y. Guan, “Development of a method for selection of effective singular values in bearing fault signal de-noising,” *Applied Sciences*, vol. 6, no. 5, p. 154, 2016.
- [4] R. Zhao, R. Yan, J. Wang, and K. Mao, “Learning to monitor machine health with convolutional Bi-directional LSTM networks,” *Sensors*, vol. 17, no. 2, article no. 273, 2017.
- [5] V. Muralidharan and V. Sugumar, “A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis,” *Applied Soft Computing*, vol. 12, no. 8, pp. 2023–2029, 2012.
- [6] A. Widodo and B. Yang, “Support vector machine in machine condition monitoring and fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.
- [7] J. Yan and J. Lee, “Degradation assessment and fault modes classification using logistic regression,” *Journal of Manufacturing Science and Engineering*, vol. 127, no. 4, pp. 912–914, 2005.
- [8] Z. Feng, M. Liang, and F. Chu, “Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples,” *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 165–205, 2013.
- [9] B. A. Paya and I. I. Esat, “Artificial neural networks based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor,” *Mechanical Systems and Signal Processing*, vol. 11, no. 5, pp. 751–765, 1997.
- [10] K. Patan and J. Korbicz, *Artificial Neural Networks in Fault Diagnosis*, pp. 333-379, Springer, Berlin, Germany, 2004.
- [11] C. Lu, Z. Y. Wang, W. L. Qin, and J. Ma, “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification,” *Signal Processing*, vol. 130, pp. 377–388, 2017.
- [12] N. K. Verma, V. K. Gupta, M. Sharma, and R. K. Sevakula, “Intelligent condition based monitoring of rotating machines using sparse auto-encoders,” in *Proceedings of the 2013 IEEE International Conference on Prognostics and Health Management, (PHM '13)*, pp. 1–7, IEEE, Gaithersburg, MD, USA, June 2013.

- [13] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering & System Safety*, vol. 115, no. 7, pp. 124–135, 2013.
- [14] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 72, no. 2, pp. 92–104, 2016.
- [15] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.
- [16] X. Guo, 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.
- [17] Y. Bengio and Y. Lecun, "Scaling learning algorithms towards AI," *Large-Scale Kernel Machines*, pp. 321–359, 2007.
- [18] 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. 72–73, pp. 303–315, 2016.
- [19] L. Guo, H. Gao, H. Huang, X. He, and S. Li, "Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring," *Shock and Vibration*, vol. 2016, Article ID 4632562, 10 pages, 2016.
- [20] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals," *Shock and Vibration*, vol. 2016, Article ID 6127479, 12 pages, 2016.
- [21] J. Tan, W. Lu, J. An, and X. Wan, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *Proceedings of the 27th Chinese Control and Decision Conference, CCDC 2015*, pp. 4608–4613, IEEE, May 2015.
- [22] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2018.
- [23] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, <https://arxiv.org/abs/1207.0580>.
- [24] W. Li, S. Zhang, and G. He, "Semisupervised distance-preserving self-organizing map for machine-defect detection and classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 869–879, 2013.
- [25] W. Du, J. Tao, Y. Li, and C. Liu, "Wavelet leaders multifractal features based fault diagnosis of rotating mechanism," *Mechanical Systems and Signal Processing*, vol. 43, no. 1–2, pp. 57–75, 2014.
- [26] J. S. Lin and Q. Chen, "Fault diagnosis of rolling bearings based on multifractal detrended fluctuation analysis and Mahalanobis distance criterion," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 515–533, 2013.
- [27] X. Lou and K. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1077–1095, 2004.
- [28] S. Wang and C. Manning, "Fast dropout training," in *Proceedings of the International Conference on Machine Learning, JMLR.org*, p. 118, 2013.
- [29] X. Jiang, S. Li, and Y. Wang, "A novel method for self-adaptive feature extraction using scaling crossover characteristics of signals and combining with LS-SVM for multi-fault diagnosis of gearbox," *Journal of Vibroengineering*, vol. 17, no. 4, pp. 1861–1878, 2015.
- [30] J. Wang, S. Li, X. Jiang, and C. Cheng, "An automatic feature extraction method and its application in fault diagnosis," *Journal of Vibroengineering*, vol. 19, no. 4, pp. 2521–2533, 2017.
- [31] J. Wang, X. Jiang, S. Li, and Y. Xin, "A novel feature representation method based on deep neural networks for gear fault diagnosis," in *Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin)*, pp. 1–6, IEEE, Harbin, China, July 2017.
- [32] X. Jiang, S. Li, and Y. Wang, "Study on nature of crossover phenomena with application to gearbox fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 83, pp. 272–295, 2017.