INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2022.16.54 Volume 16, 2022

A CNN-ABiGRU method for Gearbox Fault Diagnosis

Xiaoyang Zheng[1], Zeyu Ye[2], Jinliang Wu[3],
School of Artificial Intelligence, Chongqing University of Technology,
Chongqing, 400000
China

Received: June 25, 2021. Revised: December 8, 2021. Accepted: January 11, 2022. Published: January 13, 2022.

Abstract—As a key part of modern industrial machinery, there has been a lot of fault diagnosis methods for gearbox. However, traditional fault diagnosis methods suffer from dependence on prior knowledge. This paper proposed an end-to-end method based on convolutional neural network (CNN), Bidirectional gated recurrent unit (BiGRU), and Attention Mechanism. Among them, the application of BiGRU not only made perfect use of the time sequence of signal, but also saved computing resources more than the same type of networks because of the low amount of calculation. In order to verify the effectiveness and generalization performance of the proposed method, experiments are carried out on two datasets, and the accuracy is calculated by the ten-fold crossvalidation. Compared with the existing fault diagnosis methods, the experimental results show that the proposed model has higher accuracy.

Keywords—Gearbox, Fault Diagnosis, CNN, BiGRU, Attention Mechanism.

I. Introduction

GEABOX is an important part of modern mechanical equipment and which has been widely used in various industrial fields. Because of the harsh working conditions, it often breaks down at work. When it happens, there always will be inestimable consequences. Considering the safety and efficiency of industrial operation, it is quite urgent to develope the fault diagnosis method of gearbox.

With the development of modern industry to large-scale and complex, the traditional fault diagnosis method is no longer applicable. The signal processing based method, such as Short Time Fourier Analysis [1] and Empirical Mode Decomposition [2], has been difficult to adapt to the complex and large-scale industrial data [3]-[4]. The machine learning based method, like support vector machine (SVM) with 8 wavelet packet energy features in [5] and Trace ratio-linear discriminant analysis (TR-LDA) proposed by Jin in [6], still can not reach the ideal accuracy. More importantly, both of them rely heavily on prior knowledge to realize feature extraction which seriously reduces the efficiency of fault diagnosis.

Therefore, the method based deep learning has become the focus of scholars' research [7]–[8]. Shao et al. [9] proposed a Pre-trained model to realize highly-accurate machine fault diagnosis through transfer learning. Wang et al. [10] used synchronous extraction transform (SET) to transform the vibration signal into time-frequency representation (TFR) to enhance the robustness of its feature representation, and then trained the data in the deep reinforcement learning (DRL) framework. CNN, as a classic deep learning method, has been widely used in the field of fault diagnosis. Jing et al. [11] applied CNN to the fault diagnosis of gearbox. Then, it has also been improved by many scholars. Azamfar et al. [12] stacked one-dimensional signal data row by row to generate two-dimensional matrix as the input of two-dimensional CNN. Wang et al. [13] proposed a model combining particle swarm optimization and CNN to confirm CNN parameters accurately and automatically.

Although CNN has excellent local feature extraction ability, it can not take advantage of the time sequence of signal data. Consequently, Long Short-Term Memory (LSTM) has been attempted by a lot of researchers in the industrial area [14]-[21]. LSTM is an improved algorithm based on Recurrent Neural Network, which can take the output of the previous time as the input of this time. The improvement of LSTM is to add three gates, input gate, forget gate, and output gate to control the influence of previous input. But, it's not enough to be able to contact the previous data. In [22], a Bidirectional Long Short-Term Memory (BiLSTM) method was used for fault diagnosis of aircraft actuators. Its neurons can receive not only the data of the last moment, but also the data of the next moment. In [23], Cho coupled the input gate and forget gate of LSTM to update gate, and replaced the output gate with reset gate to obtain a Gated recurrent unit (GRU) model. It has also been used in the field of fault diagnosis. Zhao et al. [24] combined handcrafted feature design and automatic feature learning to create a local feature-based gated recurrent unit network (LFGRU) for machine health monitoring. Compared with LSTM, GRU can save more computing resources and achieve the same effect.

In recent years, Attention Mechanism has been successfully used in the field of deep learning. It can get more key and effective features by establishing the dependency between output and input. There are plenty of scholars have combined Attention Mechanism with existing methods such as AdaBoost [25], Residual Network [26], and CNN [27] to improve them.

Inspired by the above methods, a CNN-ABiGRU method combining the advantages of CNN, BiGRU, and Attention Mechanism is proposed in this paper. The main contributions of this paper are summarized below:

1) A novel fault diagnosis method for gearbox combining the advantages of CNN, BiGRU, and Attention Mechanism.
2) The propesed model can directly take the original signal as input without prior knowledge.
3) The use of GRU reduces the consumption of computing resources.

The remainder of this paper is organized as follows. The detailed introduction of the model is in Section 2. The comparative experiments with existing methods are demonstrated in Scetion 3. The final conclusion is in Scetion 4.

II. Methodology

In order to extract more accurate and effective features, we propose a model based on CNN, BiGRU, and Attention Mechanism. As shown in Fig. 2, the proposed model consists of five parts:

1) The input layer: receive the original signal.
2) The Convolution layer: extract local features.
3) The BiGRU layer: extract high-level features.
4) The Attention layer: remove redundant features and retain key information.
5) The ouput layer: realize classification and output results.

among them, the CNN layer, the BiGRU layer, and the Attention layer will be described in detail in this section.

A. The Convolution layer

CNN is a typical feedforward neural network. It has made remarkable achievements in the fields of image and speech because of its strong nonlinear feature extraction ability. Convolution layer is the most important layer in CNN, which can generate feature map by convolution of filter composed of weights [28]. The output feature map c_j of one-dimensional convolution for this model can be calculated as follows:

$$c_j = f\left(\sum_{i=1}^{M} x_i * k_{ij} + b_j\right), j = 1, \cdots, N \tag{1}$$

where f is activation function, x_i is the ith input map, k_{ij} is the convolution kernel between the ith input map and the jth output map, b_j is the bias of the jth output feature map, M is the number of the input maps, N is the number of the output maps. The size of the output feature maps S can be calculated as follows:

$$S = \frac{s_{input} - s_k}{step} + 1 \tag{2}$$

where s_{input} is the size of the input map, s_k is the size of the convolution kernel, $step$ is convolution kernel step size. Finally, there will be N output feature maps of size S, which are extracted from the input maps by the convolution layer.

B. The BiGRU layer

GRU is a simplified model based on BiLSTM. BiLTSM has a strong long dependence capability, which can extract high-level features from time series data. Therefore GRU not only inherits the advantages of BiLSTM, but also has less calculation.

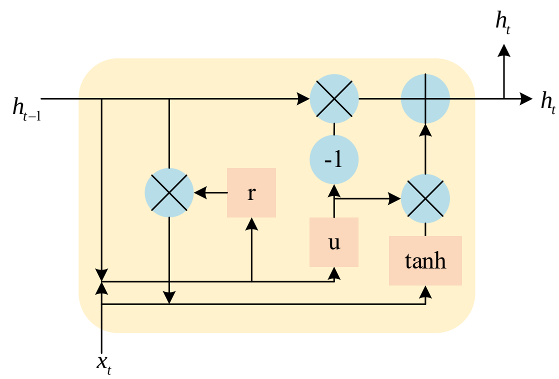

Figure 1. The structure of GRU

As shown in the Fig. 1, GRU contains two types of gates: an update gate and a reset gate. The update gate is used to control the influence of the hidden state at previous time on the hidden state at this time. The calculation of the update gate u_t at time t is as follows:

$$u_t = \sigma\left(W_u \cdot \left[h_{t-1}, x_t\right] + b_u\right) \tag{3}$$

where σ is activation function, W_u is the weight matrix of the update gate, b_u is the bias of the update gate, h_{t-1} is the output of the previous hidden layer. The more information the previous hidden layer retains, the larger the update gate, and the greater the impact on the output of the current hidden layer as in (5). The reset gate is used to control the degree of ignoring the hidden state information at previous time. The calculation of the reset gate r_t at time t is as follows:

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right] + b_r\right) \tag{4}$$

where σ is activation function, W_r is the weight matrix of the reset gate, b_r is the bias of the reset gate, h_{t-1} is the output of the previous hidden layer. The smaller the reset gate, the less information of the previous hidden layer will be retained, and the more will be ignored as in (6). The hidden state h_t and the

DOI: 10.46300/9106.2022.16.54

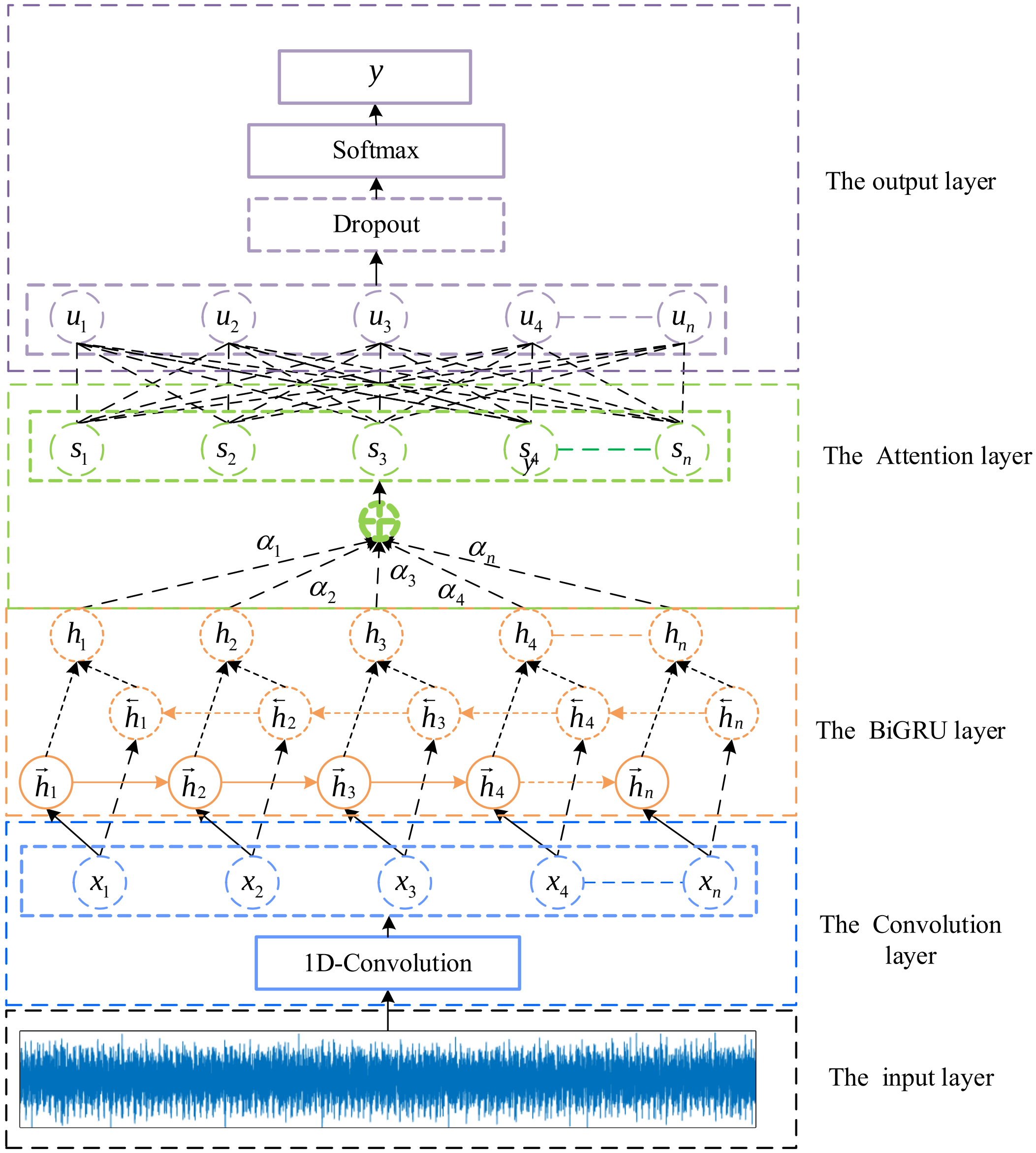

Figure 2. The structure of the proposed model

candidate hidden state h_t at time t can be calculated as follow

$$h_t = \left(1-u_t\right)\cdot h_{t-1} + u_t \cdot h_t \tag{5}$$

$$h_t = \tanh\left(W_h \cdot \left[r_t \cdot h_{t-1}, x_t\right] + b_h\right) \tag{6}$$

where $tanh$ is activation function, W_h is the weight matrix, h_{t-1} is the output of the previous hidden layer, x_t is the input at time t.

GRU can only contact the status information of the previous moment, which limits its ability to extract more accurate high-level features. BiGRU can both obtain the the state information of the previous moment and the next moment. Therefore, BiGRU is used in this paper. BiGRU consists of a forward GRU unit and a backward GRU unit.

As shown in Fig. 3, the hidden state can be divided into forward hidden state $\vec{h}_t$ and backward hidden state $\overleftarrow{h}_t$. They

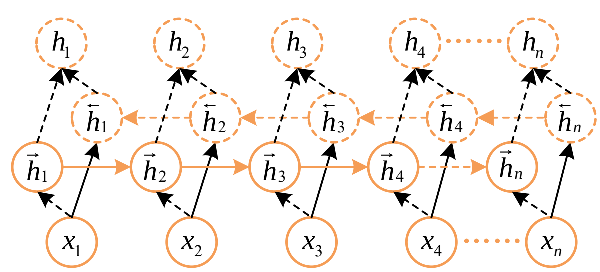
Figure 3. The structure of BiGRU

can be computed as follows:

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right) \tag{7}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t-1}\right) \tag{8}$$

where GRU is GRU computation. The final output of BiGRU is as follows:

$$h_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{9}$$

C. *The Attention layer*

Attention Mechanism is a model inspired by human brain attention and derived from Encoder-Decoder model. Encoder-Decoder is a model for mapping one variable length sequence to another variable length sequence. The starting point of Attention Mechanism is to expect that there is a screening mechanism in the mapping process of two sequences, which can keep more important information and filter out less important information.

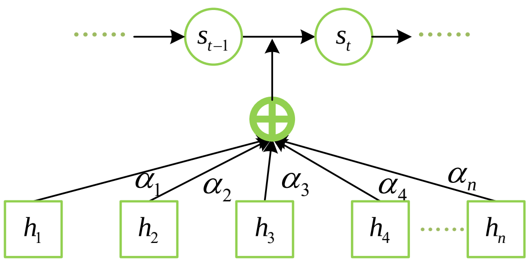

Figure 4. The structure of Attention Mechanism

The structure of Attention Mechanism is presented in Fig. 4. Its calculation is as follows:

$$e_t = u^T \cdot \tanh\left(W \cdot \left[h_t, s_{t-1}\right]\right) \tag{10}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t=1}^{T} \exp(e_t)} \tag{11}$$

$$s_t = \sum_{t=1}^{T} \alpha_t \cdot h_t \tag{12}$$

where h_t is the output of BiGRU, u^T and W are weight matrix, e_t is correlation degree, α_t is attention weight obtained by normalizing the correlation degree. The screening function of Attention Mechanism is realized by weighted summation of attention weight and imput feature. For the output at this time, the more important the input feature is, the greater the attention weight is, and the heavier the proportion in the final output.

III. Experiment

In this section, the effectiveness of the proposed model is verified. Since the gearbox is composed of a gear and a bearing, a gearbox dataset [9] and a bearing dataset is used in the comparative experiment. In order to make the experimental results more objective, the ten-fold cross validation evaluation are used in each group of experiments.

A. *Gearbox dataset*

The gearbox dataset is from Southeast University, China. These data are collected from Drivetrain Dynamic Simulator (DDS). DDS consists of brake dvice, motor, two-stage parallel gearbox, two-stage planetary gearbox, motor controller, and brake controller.

There are seven vibrating 608A11 sensors in the surface of DDS test-bed to collect the vibration signal. Their frequency range, measuring range and accuracy are 0.5 Hz, 50 g, and 100 mV/g. Three of these sensors received the vibration signals of planetary gearbox in x, y, z three directions, other three of them received the three direction signals of gearbox, and the last of them was used for measurement of drive motor.

Table 1. The fault type of gearbox dataset

Component	Type	Description
Bearing	Health	Bearing is in health condition
	Ball	Crack occurs in the ball
	Combo	Crack occurs in inner and outer ring
	Inner	Crack occurs in inner ring
	Outer	Crack occurs in outer ring
Gear	Health	Gear is in health condition
	Chipped	Crack occurs in the feet
	Miss	One of feet is missed
	Root	Crack occurs in the root of feet
	surface	Wear occurs in the surface

The gearbox dataset contains vibration signals of gear and bearing. As listed in Table 1, each of them consists of four fault states and one normal state. All of these data are collected under two conditions of speed and load configuration of 20 Hz-0V and 30 Hz-2V.

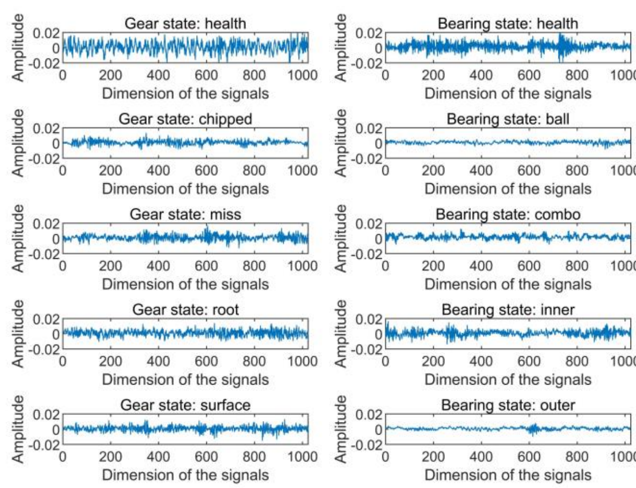

Figure 5. Signal examples of various states of gearbox

In the Fig. 5, 1200 data points are randomly selected from the original vibration signals of gear and bearing in each state as data samples. There are totally 10,000 samples under each conditions for both gear and bearing. Each state consists of 2,000 samples, of which 1,800 are for traning and 200 are for testing.

Table 2. The experimental results of gearbox dataset

INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2022.16.54 Volume 16, 2022

Fault Diagnosis Method	**Bearing**		**Gear**	
	20-0	***30-2***	***20-0***	***30-2***
BiGRU[24]	93.00%	93.60%	93.80%	90.70%
LFGRU[24]	93.20%	94.00%	94.80%	95.80%
Pre-trained model [9]	99.94%	99.42%	99.64%	99.02%
SET+CNN[10]	99.82%	99.80%	99.21%	99.92%
TFR+DRL[10]	100%	100%	99.94%	99.92%
The proposed model	100%	100%	99.97%	99.94%

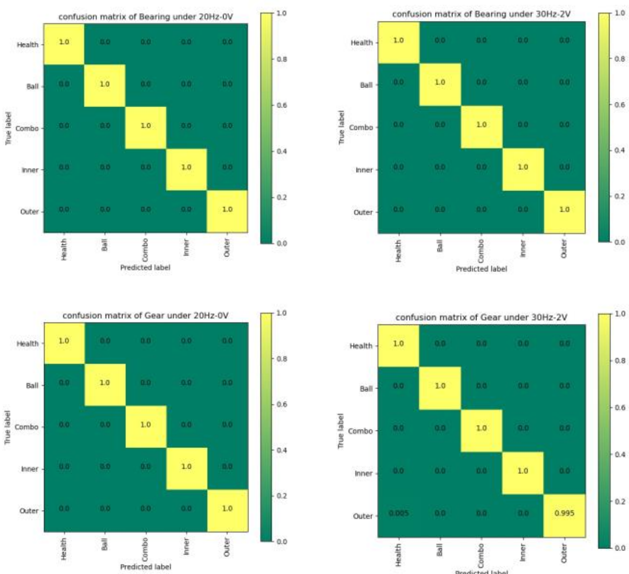

Figure 6. Confusion matrix of experimental results

To prove the superiority of the proposed model, the experimental results are compared with BIGRU, LFGRU, Pre-trained model, the model based on SET and CNN, and the model based on TFR and DRL. As listed in Table 2 comparison results show that the proposed model is more accurate than other fault diagnosis methods based on deep learning. Even though the model based on TFR and DRL is as good as the proposed model in bearing data set, it still lags 0.03\% and 0.02\% in gear data set. As shown in Fig. 6, there are more details of the experimental results. The confusion matrix of experimental results further verified that the proposed model can achieve high-precision fault diagnosis.

A. Bearing dataset

In order to further verify the effectiveness of the proposed model and prove that this model has generalization performance, anothoer experiment will be carried out on a public data set from Case Western Reserve University (CWRU).

Figure 7. The bearing test rig of CWRU

As shown in Fig. 7, the test rig consists of an induction motor, a troque transducer/encoder, a dynamometer and control electronics. There are four different motor loads (0, 1, 2, and 3 HP) while collecting the bearing dataset. Each one of them contains three fault types, outer-race fault (OF), inner-race fault (IF), and ball fault (BF). Each fault type is created by the Electrical Discharge Machining and set to three severity levels 7, 14, and 21 mils, respectively. Therefore 9 fault types and 1 normal condition (NC) are setting in every motor load. As shown in Fig. 8, 2048 continuous data points are selected from each bearing state as data samples.

There are 10,000 data samples for each state of bearing, of which 9,000 are used for trainning and 1,000 are used for testing.

The proposed model is compared with SVM, TR-LDA, BiGRU, BiLSTM, and CNN. As listed in Table 3, CNN, BiGRU, and BiLSTM all have working conditions with excellent accuracy. Even so, the accuracy of the proposed model is still 11.85\%, 2.8\%, 1.55\%, and 1.4\% higher than that of them.

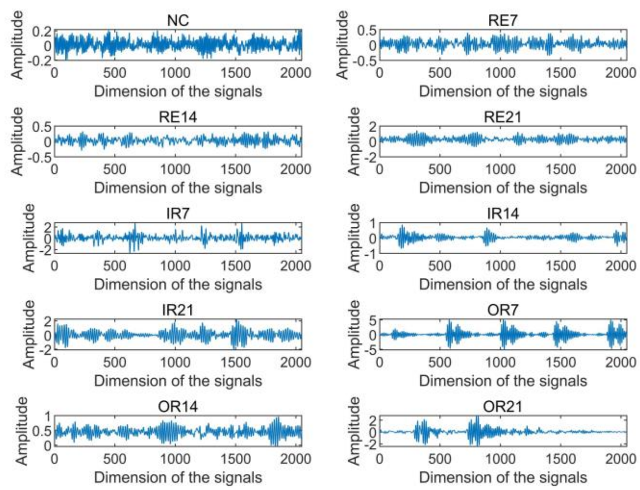

Figure 8. Signal examples of various states of bearing

Table 3. The experimental results of bearing dataset

Fault Diagnosis Method	**Bearing Dataset**			
	0HP	***1HP***	***2HP***	***3HP***
SVM[5]	88.90%	-	-	-
TR-LDA[6]				92.50%
CNN	88.10%	90.40%	87.10%	83.50%
BiGRU	82.95%	96.25%	95.05%	95.15%

Fault Diagnosis Method	Bearing Dataset			
	0HP	*1HP*	*2HP*	*3HP*
BiLSTM	86.50%	94.40%	97.70%	97.65%
The proposed model	99.95%	99.05%	99.25%	99.05%

IV. Conclusion

In order to improve the reliability and safety of modern industrial machinery, a fault diagnosis method of gearbox Based on CNN, BiGRU, and Attention Mechanism is proposed in this paper. To prove the effectiveness and generalization performance of the proposed model, experiments will be conducted on two different datasets. Experimental results show that the proposed model can take the original signal as input and extract more accurate features than the existing fault diagnosis methods. In the real working environment, the signal is often accompanied by noise. In the future work, developing an efficient fault diagnosis method that can resist noise will be the key research direction.

ACKNOWLEDGMENT

This work is funded by Fundamental and Advanced Research Project of Chongqing CSTC of China and the project No. is cstc2019jcyj-msxmX0386.

References

[1] H. Z. Gao, L. Liang, X. G. Chen, G. H. Xu, "Feature extraction and recognition for rolling element bearing fault utilizing short-time Fourier transform and non-negative matrix factorization," Chinese Journal of Mechanical Engineering, vol. 28, pp. 96-105, 2015.

[2] Z. J. Shen, X. F. Chen, X. L. Zhang, Z. J. He, "A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM," Measurement, vol. 45, pp. 30-40, 2012.

[3] D. Q. Zhu, S. L. Yu, “Survey of knowledge based fault diagnosis methods,” J. of Anhui University of Technology, vol. 19, pp. 197-205, 2002.

[4] H. Li, D. Y. Xiao, “Survey of fault diagnosis methods based on data driven,” Control and Decision, vol. 26, pp. 1-9, 2011.

[5] W. L. Du, J. F. Tao, Y. M. Li, C. L. Liu, “Wavelet leaders multifractal features based fault diagnosis of rotating mechanism,” Mechanical Systems and Signal Processing, vol. 43, pp. 57-75, 2014.

[6] X. H. Jin, M. B. Zhao, T. W. S. Chow, M. Pecht, “Motor Bearing Fault Diagnosis Using Trace Ratio Linear Discriminant Analysis,” IEEE Transactions on Industrial Electronics, vol. 61, pp. 2441-2451, 2014.

[7] X. Y. Zheng, J. L. Wu, Z. Y. Ye, “An End-To-End CNN-BiLSTM Attention Model for Gearbox Fault Diagnosis,” IEEE International Conference on Progress in Informatics and Computing, pp. 386-390, 2020.

[8] X. L. Pei, X. Y. Zheng, J. L. Wu, “Intelligent bearing fault diagnosis based on Teager energy operator demodulation and multiscale compressed sensing deep autoencoder,” Measurement, vol. 179, pp. 109452, 2021.

[9] S. Y. Shao, S. McAleer, R. Q. Yan, P. Baldi, “Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning,” IEEE Transactions on Industrial Informatics, vol. 15, pp. 2446-2455, 2019.

[10] H. Wang, J. W. Xu, C. Sun, R. Q. Yan, X. F. Chen, “Intelligent Fault Diagnosis for Planetary Gearbox Using Time-Frequency Representation and Deep Reinforcement Learning,” IEEE/ASME Transactions on Mechatronics, pp. 1-1, 2021.

[11] L. Y. Jing, M. Zhao, P. Lin, X. Q. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” Measurement, vol. 111, pp. 1-10, 2017.

[12] M. Azamfar, J. Singh, I. B. Imaz, J. Lee, “Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis,” Mechanical Systems and Signal Processing, vol. 144, pp. 106861, 2020.

[13] C. G. Wang, H. K. Li, K. L. Zhang, S. L. Hu, B. Sun, “Intelligent fault diagnosis of planetary gearbox based on adaptive normalized CNN under complex variable working conditions and data imbalance,” Measurement, vol. 180, pp. 109565, 2021.

[14] J. J. Zhang, P. Wang, R. Q. Yan, R. X. Gao, “Long short-term memory for machine remaining life prediction,” Journal of Manufacturing Systems, vol. 48, pp. 78-86, 2018.

[15] H. R. Yan, Y. Qin, S. Xiang, Y. Wang, H. Z. Chen, “Long-term gear life prediction based on ordered neurons LSTM neural networks,” Measurement, vol. 165, pp. 108205, 2020.

[16] D. Cabrera, et al., “Bayesian approach and time series dimensionality reduction to LSTM-based model-building for fault diagnosis of a reciprocating compressor,” Neurocomputing, vol. 380, pp. 51-66, 2020.

[17] R. B. Sun, et al., “Planetary gearbox spectral modeling based on the hybrid method of dynamics and LSTM,” Mechanical Systems and Signal Processing, vol. 138, pp. 106611, 2020.

[18] S. Xiang, Y. Qi, C. C. Zhu, Y. Y. Wang, H. Z. Chen, "LSTM networks based on attention ordered neurons for gear remaining life prediction," ISA Transactions, vol. 106, pp. 343-354, 2020.

[19] H. Y. Wu, A. H. Huang, J. W. Sutherland, "Avoiding Environmental Consequences of Equipment Failure via an LSTM-Based Model for Predictive Maintenance," Procedia Manufacturing, vol. 43, pp. 666-673, 2020.

[20] S. Xiang, Y. Qin, C. C. Zhu, Y. Y. Wang, H. Z. Chen, "Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction," Engineering Applications of Artificial Intelligence, vol. 91, pp. 103587, 2020.

[21] J. Q. Liu, C. L. Pan, F. Lei, D. B. Hu, H. F. Zuo, "Fault prediction of bearings based on LSTM and statistical process analysis," Reliability Engineering \& Systems Safety, vol. 214, pp. 107646, 2021.

[22] J. Yang, Y. Q. Guo, W. L. Zhao, "Aircraft Actuator Fault Diagnosis Using Deep Learning Based Sparse Representation and TSM," IEEE Aerospace Conference, pp. 1-9, 2019.

[23] K. Cho, et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," EMNLP, pp. 1724-1734, 2014.

[24] R. Zhao, et al., "Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks," IEEE Transactions on Industrial Electronics, vol. 65, pp. 1539-1548, 2018.

[25] Z. Long, et al., "Motor fault diagnosis using attention mechanism and improved adaboost driven by multi-sensor information," Measurement, vol. 170, pp. 108718, 2021.

[26] K. Zhang, B. P. Tang, L. Deng, X. L. Liu, "A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox," Measurement, vol. 179, pp. 109491, 2021.

[27] H. Wang, J. W. Xu, R. Q. Yan, C. Sun, X. F. Chen, "Intelligent Bearing Fault Diagnosis Using Multi-Head Attention-Based CNN," Procedia Manufacturing, vol. 49, pp. 112-118, 2020.

[28] G. Q. Qiu, Y. K. Gu, Q. Cai, "A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions," Measurement, vol. 145, pp. 94-107, 2019.