An Improved CNN-LSTM Network Based On Hierarchical Attention Mechanism For Motor Bearing Fault Diagnosis

Yisha Jiao
China Agricultural University
Yaoguang Wei (wyg@cau.edu.cn)
China Agricultural University
Dong An
China Agricultural University
Wenshu Li
China Agricultural University
Qiong Wei
China Agricultural University

Research Article

Keywords: Motor bearing fault diagnosis, Hierarchical attention mechanism, Convolutional neural network, Long short-term memory network

Posted Date: April 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-201800/v1

An Improved CNN-LSTM Network based on Hierarchical Attention Mechanism for Motor Bearing Fault Diagnosis

Yisha Jiao [a,b,c,d], Yaoguang Wei [a,b,c,d,*], Dong An [a,b,c,d], Wenshu Li [a,b,c,d] and Qiong Wei [a,b,c,d]

[a] National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China

[b] Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China

[c] Precision Agricultural Technology Integration Research Base (Fishery), Ministry of Agriculture and Rural Affairs , Beijing 100083, China

[d] College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, China

* Correspondence: wyg@cau.edu.cn; Tel.: +86 10 62737696

Corresponding author

Corresponding author: Prof. Dr. Yaoguang Wei

P.O.Box 63.

China Agricultural University.

Tsinghua East Road 17#, Haidian District, Beijing,100083, China

Tel:+86 10 62737696

Fax:+86 10 62737741

Mob:+86 13521591976

E-mail: wyg@cau.edu.cn

College of Information and Electrical Engineering

China Agricultural University

Beijing, 100083. P. R. China

Abstract

Motor is widely used in industrial production, but the frequent motor bearing fault brings great safety hazard to the production. Traditional fault diagnosis methods often require prior signal processing knowledge and are inefficient. In order to solve this problem, the artificial intelligence fault diagnosis method has been applied in motor bearing fault diagnosis. With the help of the original motor running state signal collected by the sensors, non-invasive real-time detection of motor bearing fault can be realized. This paper presents an improved CNN-LSTM network based on hierarchical attention mechanism(CALSTM) for motor bearing fault diagnosis. In this artificial intelligence method, the fault characteristics of the original data can be learned by convolutional neural network, and then the importance of the features can be obtained by using hierarchical attention mechanism. Finally, the weighted results are sent to the LSTM network for time dimension selection. This method does not need signal processing and adaptively weights the features of each sample learned by the neural network, which enhances the explanatory ability of the learning process of the neural network. When carry out experiments on CWRU data set, and the experimental results indicate that, compared with several common models, CALSTM method has a better diagnosis effect, and the overall accuracy of the model reached 99.22%.

Keywords Motor bearing fault diagnosis · Hierarchical attention mechanism · Convolutional neural network · Long short-term memory network

1、Introduction

As a common industrial equipment, motor is widely used in modern factories for automation and large-scale production. Motor bearing is a common part of motor equipment, and its health has a great impact on the performance, stability and service life of the whole equipment. Due to the complex operating environment and high load operating conditions, the rolling bearing is prone to damage, and its fault probability accounts for up to 40% in all motor faults (Lau et al.2010). In the actual production, the fault of a single motor may cause the stagnation of the whole production line and cause huge economic losses. Therefore, finding an accurate and effective fault diagnosis method for motor bearings has become an urgent need in industrial production.

With the development of modern electronic technology, sensor technology and detection technology, motor fault diagnosis has been further developed. In 1965, Cooley published the Theory of fast Fourier Transform, and spectrum analysis became a research hotspot(Cooley et al.1965). Various kinds of spectrum analyzers have also sprung up and been applied in motor bearing fault diagnosis. By comparing the characteristic frequency of vibration signal of damaged motor bearing and the analysis result of spectrum analyzer, it can be judged whether there is a fault. Since the 1980s, due to the rapid development of computer technology, computers have powerful information storage and processing capacity, which facilitates the integration of multiple technologies to realize the state monitoring and fault diagnosis of motor bearings. Because the data of motor running state shows the characteristics of big data. Therefore, higher requirements are put forward for the robustness, generalization ability and real-time performance of diagnostic technology.

The intelligent model is mainly divided into signal processing and artificial intelligence. The method of signal processing focuses on the effective extraction of artificial characteristic parameters and

depends on the numerical calculation method and signal processing technology. In the signal processing method, the signal is extracted, transformed and analyzed, and the characteristics of mechanical faults are obtained through numerical calculation. The characteristic values commonly used in motor bearing fault diagnosis include the characteristics of time domain, frequency domain and time-frequency domain. Envelope analysis(Tsao et al.2012), Spectrum Kurtosis, Wigner-Ville(WVD) distribution(Choy et al.2009), wavelet transform(Siddiqui et al.2016; Zhang et al.2013), empirical mode decomposition(Mohanty et al.2015) and Variational Mode Decomposition(Jinde et al.2016) are widely used in rolling bearing state monitoring and fault diagnosis. Signal processing methods are often combined with artificial intelligence methods to play a role in data processing and feature optimization.

Compared with the signal processing method, the method based on artificial intelligence focuses on learning historical and empirical data, does not rely too much on the calculation and analysis of signals, and shows a good prospect in the aspect of fault diagnosis. There is no requirement for the relevant motor professional background of the researcher, which is more suitable for data-driven research. For example, (Wen and Gao et al.2017) minimized the maximum mean difference (MMD) between the source problem and the target problem by using the Transfer Learning method, and realized the domain adaptive cross-domain fault diagnosis. In this research background, the most common motor fault diagnosis methods are CNN, DBN and SAE. CNN is a supervised neural network model that is good at image processing. At present, it also shows strong application ability in fault diagnosis. Hoang D T et al. proposed a new method based on CNN to diagnose the fault of rolling bearings(Hoang et al.2018). By converting 1-D vibration signals into 2-D images and utilizing the effectiveness of CNN in image classification, this method achieves very high accuracy and robustness in noisy environment without any feature extraction technology. Wen L et al. converted the signal into a two-dimensional (2-D) image and proposed a new CNN method based on Lenet-5 for fault diagnosis, which improved the diagnosis effect(Wen et al.2017). As an unsupervised learning method, DBN is one of the earliest methods proposed in the field of deep learning. Tamilselvan et al. proposed a new multi-sensor fault diagnosis method using deep belief network (DBN) and successfully applied it to the fault diagnosis of aircraft engines(Tamilselvan and Wang 2013). Shao H et al. used DBN to realize the fault diagnosis of rolling bearings under variable working conditions and high noise(Shao et al.2015). They also made a detailed comparison between the traditional fault detection method and DBN deep learning method, and proved that the fault diagnosis model using DBN deep learning method has better accuracy and robustness. Shao H et al. also used the SAE network optimized by fish swarm algorithm to diagnose motor faults and improve the accuracy of classification(Shao et al.2017).

CNN can capture sensitive fault information without the need of expert knowledge, so it is becoming more and more popular in various methods of motor bearing fault diagnosis. CNN can extract the invariant features from the original vibration data, but it cannot take into account the timing characteristics of the motor bearing vibration signal itself. Therefore, considering the effectiveness of LSTM in processing the timing data, this study combined CNN and LSTM to form a CLSTM network.

CLSTM network takes into account both feature learning and data time correlation, but CNN is still a black box and cannot automatically identify which features are more important. Therefore, this paper uses hierarchical attention mechanism to explore important features. The hierarchical attention

mechanism is first applied to text classification, which can improve the classification performance(Yang et al.2016). In this paper, the hierarchical Attention mechanism can connect features extracted after CNN model processing with fault diagnosis results, so as to intuitively see the effect of CNN feature extraction. Then, without changing the timing sequence of CNN output features, weighted results are successively sent to LSTM network for further learning. We call this method CALSTM.

This paper discusses the application of CALSTM method in motor bearing fault diagnosis, which does not require complex data preprocessing or signal processing of motor bearing vibration signal. Researchers without professional background of electrical engineering or signal processing can easily obtain rolling bearing fault diagnosis results through this model. And compared with the results of other methods, the accuracy of fault diagnosis obtained by the method proposed in this paper has been improved to some extent.

The rest of this paper is arranged as follows: Section 2 introduces the method, then Section 3 conducts experiments and discusses the experimental results, and Section 4 draws conclusions.

2、Methodology

In this section, the proposed CALSTM method will be introduced. First, we introduce the convolutional neural network CNN, then the hierarchical attention mechanism and LSTM model, and finally the specific architecture of the CALSTM network combined with the first three parts.

2.1 Convolutional neural network (CNN)

The convolutional neural network (CNN) is inspired by the receptive field mechanism in biology, and it is an artificial neural network (ANN) with a special structure. Different from the traditional fully connected network, each neuron in the feature map of each layer in the convolutional neural network is only sparsely connected to a small part of the neurons in the upper layer. CNN has the characteristics of local receiving field, Shared weight and spatial sub-sampling, and its hidden layer is divided into convolutional layer, activation layer and pooling layer.

The convolution operation, which extracts features by translating the original image, can be defined as the multiplication operation between the input information I and the filter (convolution kernel) w .Input information $I \in R^{M \times N}$, filter $w \in R^{P \times Q}$, in general, $M >> P$, $N >> Q$ .Convolution operation is defined as:

$$Z = \varphi\left(w * I + b\right) \tag{1}$$

Where, b stands for bias and φ represents activation function which can carry out nonlinear mapping and increase nonlinear segmentation capability. After convolution and activation, several feature maps can be obtained, and the overall feature mapping group is $Z \in R^{H \times L \times D}$.

Pooling can not only extract the most important local information in each feature map, but also significantly reduce the feature size. Therefore, pooling layer can compress the amount of data and parameters, reduce the overfitting and reduce the complexity of the network. Each characteristic

graph $Z^d \in R^{H\times L}, 1\le d\le D$ in the characteristic graph group $Z \in R^{H\times L\times D}$ is divided into several regions $R_{h,l}^d$ .These regions may or may not overlap, depending on the step size of the sliding window at the time of sampling, so $1\le h\le H^{'}, 1\le l\le L^{'}$. Each region is subjected to Down Sampling to obtain a value as a generalization of the region. The pooling layer selects the maximum pooling, that is, the maximum value of all neurons in each divided region $R_{h,l}^d$ is selected as the representative of this region, which is defined as follows:

$$y_{h,l}^d = \max_{i\in R_{h,l}^d} z_i \tag{2}$$

Where z_i is the value of each neuron in the R_j^d region.

For each feature map $Z^d \in R^{H\times L}$, the output of the feature map of the pooling layer Y^d can be expressed as follows:

$$Y^d = \left\{ y_{h,l}^d \right\} \tag{3}$$

In the convolutional neural network, both the sliding of the filter in the convolutional layer and the sliding window in the pooling layer need to use the filling method to control the size of the feature. That is, the structure of the input data is assumed to be $H_I \times L_I \times D_I$, and the structure of the output data is $H_O \times L_O \times D_O$ .Because the calculation of the data structure of the convolutional layer and the pooling layer is similar, the calculation formula under the option of "not filling" can be expressed as:

$$H_O = \left(H_I - F\right)/S + 1 \tag{4}$$

$$L_O = \left(L_I - F\right)/S + 1 \tag{5}$$

$$D_O = K \tag{6}$$

Where, F is the size of filter (also refers to the size of pooling window); S refers to the size of the slide step and K to the number of filters.

2.2Hierarchical attention mechanism

Yang Z et al. proposed the concept of hierarchical Attention. Hierarchical "Attention" network was first used for text classification. The network is divided into two parts, the first part is "word attention" part, the other part is "sentence attention" part(Yang et al.2016). The whole network divides a sentence into several parts, and then maps each part into a vector by using the bidirectional RNN combined with the attention mechanism. Finally, the sequence vector obtained by the mapping goes through a layer of bidirectional RNN combined with the "attention" mechanism, so that text classification can be realized. The classification effect of this method is obviously better than other

methods.

In this study, a similar hierarchical attention network is designed, that is, a layer of hierarchical attention is added between CLSTM models. First, hierarchical Attention replicates the output of CNN, and then maps it to a set of vectors that reflect the importance of each feature in the feature map. For each CNN feature map, hierarchical Attention can be expressed as:

$$X_{vec} = soft\max(Y) \tag{7}$$

$$X = Y * X_{vec} \tag{8}$$

Where, Y is the set of all Y^d, that is, the characteristic map output of characteristic map group Z of CNN network. A set of vectors obtained by softmax are multiplied by the matrix of the output of CNN network to obtain the weighted feature X.

Hierarchical attention assigns different weights to each output vector of CNN, so that the model can focus attention on key features and reduce the role of other features. The use of the Attention mechanism can make the results expressed by CNN more reasonable in the current task.

2.3 Long Short-Term Memory (LSTM)

The Recurrent Neural Network (RNN) has the ability to process sequential information and can represent the relationship between the current output of a sequence and the previous information.

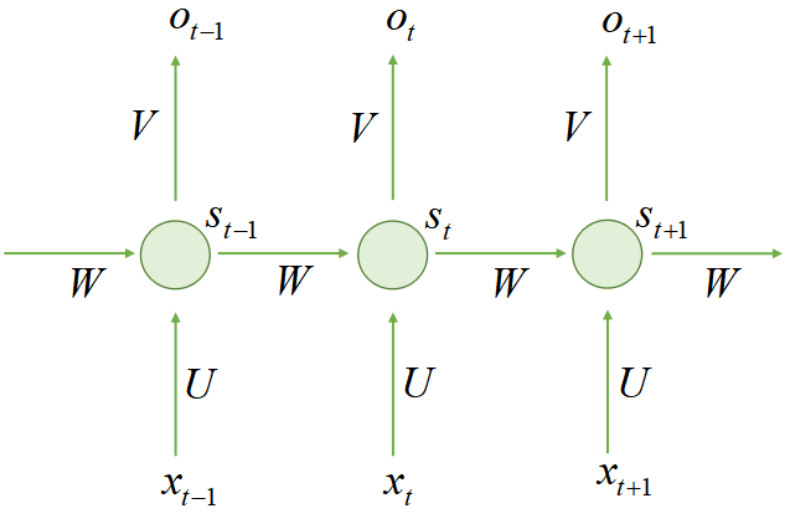


Fig. 1 Structure diagram of circulating neural network

As shown in Fig. 1, there are connections between each neuron in the RNN hidden layer. After the network receives the input x_t at time t, the value of the hidden layer is s_t and the output value is o_t.The point is, the value of st depends not only on x_t, but also on s_{t-1}.

The following formula is used to represent the calculation method of Recurrent Neural Network:

$$o_t = g(V \cdot s_t) \tag{9}$$

$$s_t = f(U \cdot X_t + W \cdot s_{t-1}) \tag{10}$$

Where V and U respectively represent the weight matrix from hidden layer to output layer and input

layer to hidden layer. g and f are nonlinear activation functions.

However, there is a phenomenon of gradient disappearance in the back propagation process of RNN, so RNN may not be able to effectively obtain the long-term dependence of data. Therefore, Long Short-Term Memory (LSTM), a special RNN, is proposed to solve the problem of gradient disappearance and gradient explosion in the long sequence training process(Hochreiter and Schmidhuber 1997). LSTM performs better in longer sequences than normal RNN. Its structure is shown in Fig. 2:

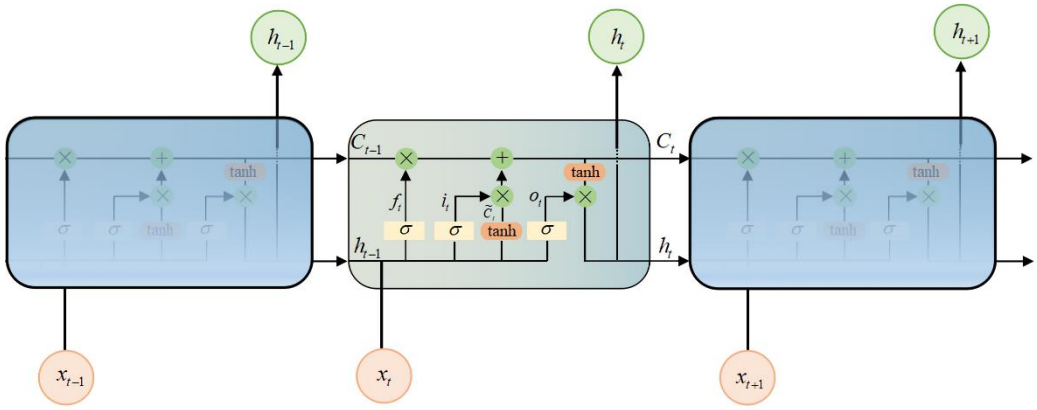


Fig. 2 LSTM network structure diagram

The LSTM network can delete or add cell state information through a structure called a gate. Generally, the LSTM is controlled by three gates, which are called forgetting gate, input gate and output gate respectively. At each time step, cell units are accessed, written, and cleared through several gates to control the flow of information along the data sequence, thereby enhancing the ability to learn long-term dependencies.

First, the LSTM has to forget some information. This part is handled through a sigmoid unit called the forget gate. As shown in formula (11):

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{11}$$

After that, the LSTM network adds new information. h_{t-1} and x_t are used to determine which information to update through an operation called an input gate. Then h_{t-1} and x_t are used to obtain new candidate cell information $\tilde{C}_t$ through a $\tanh$ layer, which may be updated into the cell information. The update process is shown in formula(12)(13):

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{12}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{13}$$

Update cell status information C_{t-1} to new cell status information C_t. Formula (14) can be expressed as:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{14}$$

The output gate judges and decides the output characteristics of the final RNN unit, as shown in formula (15)(16):

$$o_t = \sigma\left(W_O \cdot \left[h_{t-1}, x_t\right] + b_O\right) \tag{15}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{16}$$

In the above formulas, W_f, W_i, W_C, W_O are the weight matrix of the corresponding operation, b_f, b_i, b_C, b_O are the bias of the corresponding operation.

2.4 Combining method (CALSTM)

The basic network architecture of the CALSTM model proposed in this study is shown in Fig. 3. The data of vibration signals enter the network through CNN, and the convolutional layer and pooling layer of CNN reduce the data dimension and learn the fault characteristics of the data adaptively. Then, the feature weight coefficients are calculated adaptively at the hierarchical attention layer. As shown in Fig. 3, the results of CNN module are processed in two lines. One line is to retain the features learned by CNN directly, and the other line is to obtain the adaptive weight coefficient matrix after Softmax processing of the learned features. The results of the two lines are then computed in Multiply, that is, the weighted characteristic matrix is obtained by multiplying the feature by the weight coefficient of the corresponding position. Effective features usually have a greater impact on the classification results, while invalid or ineffective features have less impact on the classification results. Weighted features can enhance the effectiveness features and suppress the importance of invalid or inefficient features. After CNN and the hierarchical attention layer, the weighted features are sent into the LSTM network to further learn the temporal correlation of features. And then sends the learning results to the Softmax layer through the fully connected network to obtain the final classification results. The CALSTM model takes into account not only the fault characteristics of the original vibration signal but also the time characteristics, and emphasizes that the importance of different characteristics is different. The method can learn features efficiently from high-dimensional raw data and improve the accuracy of fault diagnosis.

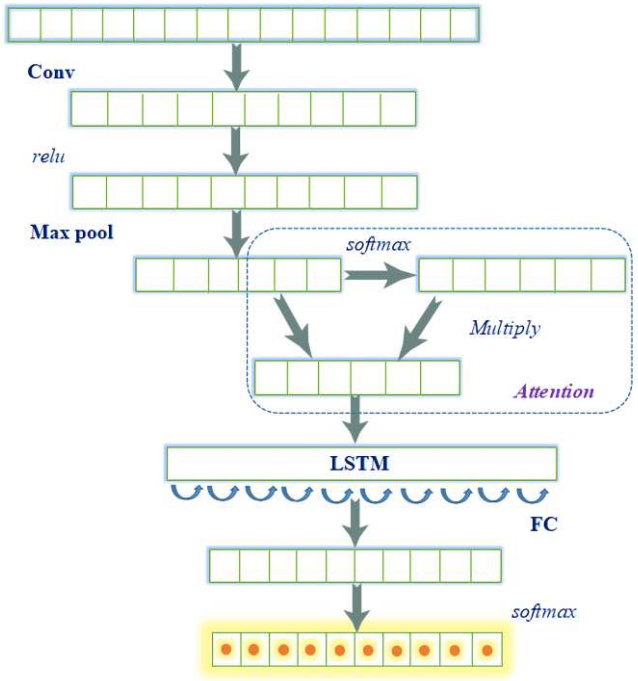


Fig 3. The basic framework of the CALSTM model

3、Experiments

3.1Dataset and Data preprocessing

The CWRU bearing data set provided by the Bearing Data Center of Case Western Reserve University is widely used in the study of bearing fault diagnosis. Its acquisition platform is shown in Fig. 4, which is mainly composed of motor, torque sensor, power meter and electronic control equipment. The motor is set to run at the speeds of 1797, 1772, 1750, and 1730rpm respectively, and each speed corresponds to the load of 0, 1, 2, and 3HP.The bearing wear fault of 7, 14, 21 and 28 mils is manually set, and the sampling frequency of vibration signal is 12K and 48K.

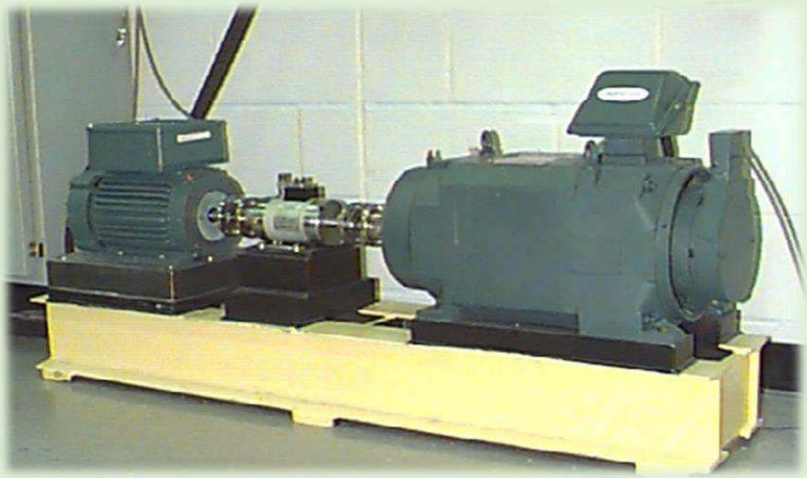


Fig. 4 CWRU rolling bearing data set acquisition platform

(Picture from https://csegroups.case.edu/bearingdatacenter/pages/apparatus-procedures)

In this study, drive end signals with a sampling frequency of 12K, bearing speed of 1797rpm, and 0HP were used. The bearing state included normal state, inner raceway fault, outer raceway fault and ball fault, with minor(7mils), moderate(14mils), and severe(21mils) wear degrees. Therefore, it can be divided into 10 types of bearing states, namely normal state(Normal), minor fault in the inner raceway (IR007), moderate fault in the inner raceway (IR014), severe fault in the inner raceway (IR021), minor fault in the outer raceway (OR007), moderate fault in the outer raceway (OR014), severe fault in the outer raceway (OR021), minor fault in the ball(B007), moderate fault in the ball(B014) and severe fault in the ball(B021).

For the convenience of calculation, the signal is normalized in this study. Then, these 10 kinds of signals are divided into several samples, and each sample is a continuous 400 data points. Each sample should meet the input requirements of CNN network, so the sample should be reshaped. Data preprocessing of vibration signal is shown in Fig. 5.

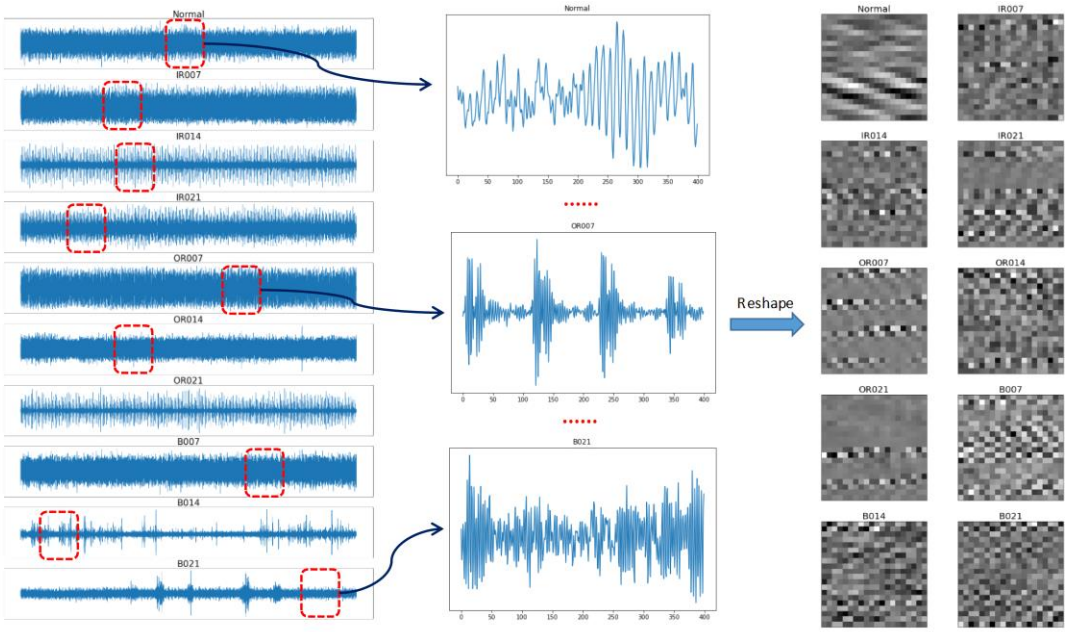


Fig. 5 Data preprocessing process diagram of vibration signal

3.2 Data preprocessing

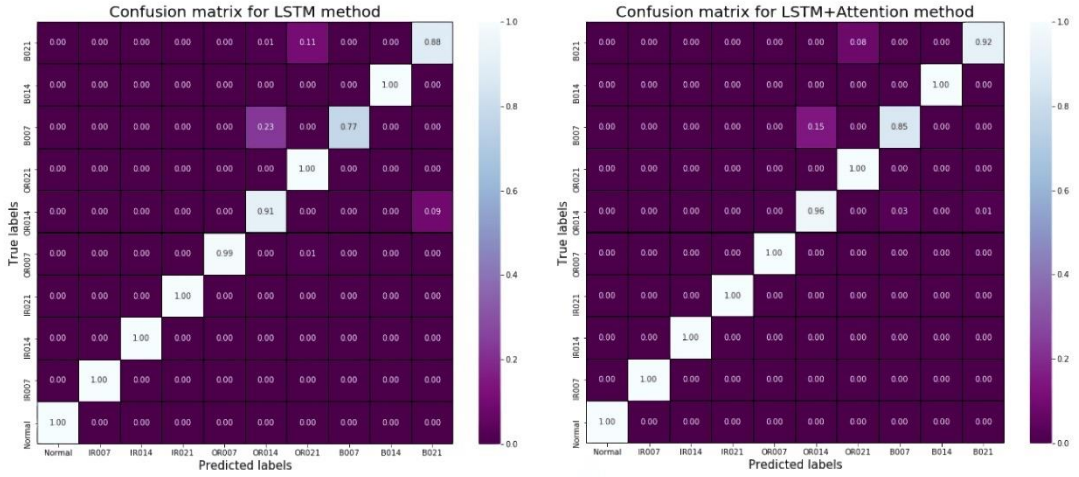
3.2.1 Experimental network

The input layer of the CALSTM model has 400 neurons in the shape of 20×20.The input information is first entered into the CNN network for two-dimensional convolution. The size of the convolution kernel is 2×2, and the slide step is 2×2.The sliding window of the pooling layer is 2×2, and the sliding step size is 2×2.In order to prevent external filling from destroying the distribution of the data itself, both the convolutional layer and the pooling layer are unfilled. Assuming that the volume of the input data is $H_1\times L_1\times D_1$, the convolutional layer has a size of $H_2\times L_2\times D_2$. The activation layer only changes the mapping relationship of each neuron, but does not change the structure of the data. Therefore, the data structure through the activation layer is still $H_2\times L_2\times D_2$, and finally $H_3\times L_3\times D_3$ is obtained after pooling. According to formula (4) - (6), the output data structure of CNN is $8\times8\times32$.Then, the results of CNN are copied and processed by the hierarchical attention layer to obtain the weighted features without changing the data dimension, that is, after the attention mechanism, the output data dimension remained at $8\times8\times32$.The results are sent into the LSTM network, where the input data size of each time step is 8×8 and the time step is 32. In order to achieve the effect of diagnosis classification, the output of LSTM network needs to add a full connection layer and change through softmax to get 10 kinds of fault diagnosis results.

3.2.2 Experimental results

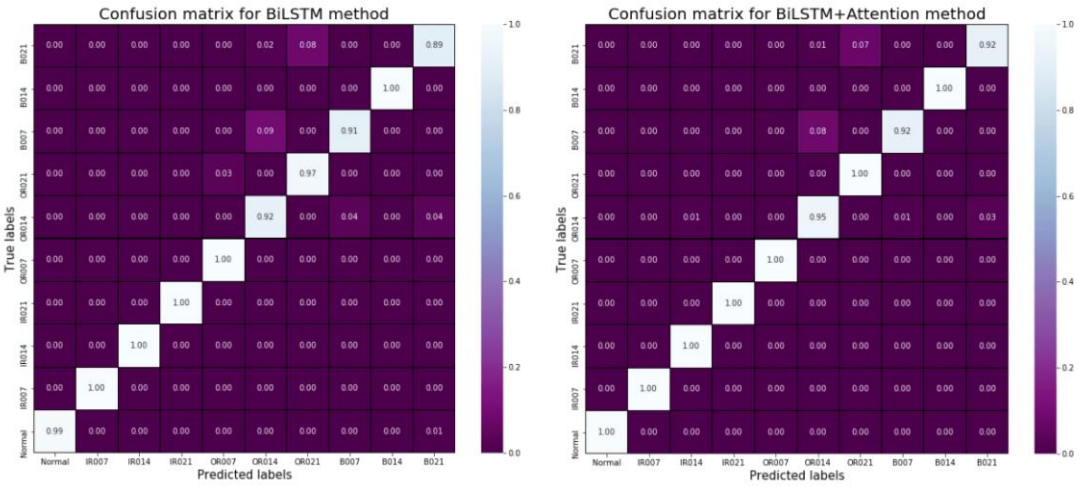
Firstly, the improvement effect of hierarchical Attention was experimented. In Fig. 6 and Fig. 7, the

abscissa represents the predictive labels for each health state, and the ordinate represents the real labels for each state. By adding the hierarchical Attention mechanism on both LSTM and BiLSTM networks, it can be found that the overall diagnostic effect has been improved to some extent. In Fig. 6, the method of "LSTM+Attention" has a significantly improved effect than LSTM network in the diagnosis of moderate fault in the outer raceway(OR021), minor fault in the ball(B007) and severe fault in the ball(B021). In Fig. 7, the BiLSTM+Attention method also achieved better results than the simple BiLSTM network, and achieved higher accuracy in the categories of normal state(Normal), moderate fault in the outer raceway(OR014), severe fault in the outer raceway(OR021), minor fault in the ball(B007) and severe fault in the ball(B021).



(a) (b)

Fig. 6 LSTM and LSTM+Attention fault classification confusion matrix



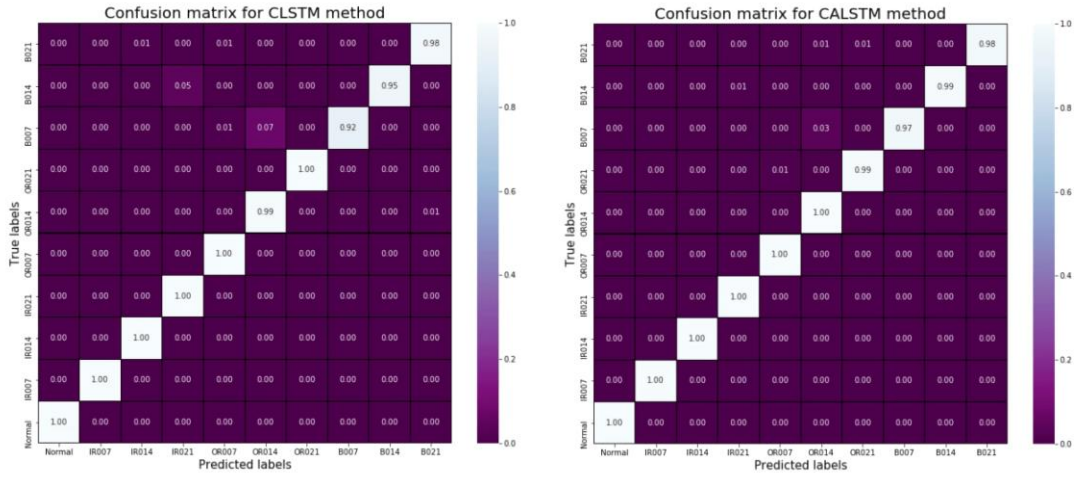
(a) (b)

Fig. 7 BiLSTM and BiLSTM+Attention fault classification confusion matrix

Basic models added hierarchical attention layer have been shown to improve the performance of diagnostic models. Therefore, the hierarchical attention mechanism is considered to be added to the CNN-LSTM network(CLSTM), which has a good performance in temporal data classification, to observe whether the performance of the model is further improved.

Fig. 8 (a) represents the classification results on the CLSTM network, and Fig. 8 (b) represents the classification results of the CALSTM method proposed in this paper. Only by comparing Fig. 8(a)

with Fig. 6(a) and Fig. 7(a), it can be found that CLSTM model is indeed superior to LSTM and BiLSTM in classification effect. As can be seen from the comparison in the Fig. 8, CALSTM has a better classification effect on the whole, achieving 100% accuracy in normal state(Normal), minor fault in the inner raceway(IR007), moderate fault in the inner raceway(IR014), severe fault in the inner raceway(IR021), minor fault in the outer raceway(OR007) and moderate fault in the outer raceway(OR014). The CALSTM method is more effective than the CLSTM method in the categories of moderate fault in the outer raceway(OR014), minor fault in the ball(B007) and moderate fault in the ball(B014). Among the classification results of the CALSTM method, minor fault in the ball(B007) has the worst effect, with an accuracy rate of about 97%. There are also a few misclassifications for severe fault in the outer raceway(OR021) and moderate fault in the ball(B014) and severe fault in the ball(B021). The comparison between Fig. 8 (a) and Fig. 8 (b) once again proves that hierarchical attention mechanism has a certain improvement effect on the performance of the diagnostic model.



(a) (b)

Fig. 8 CLSTM and CALSTM fault classification obfuscation matrix

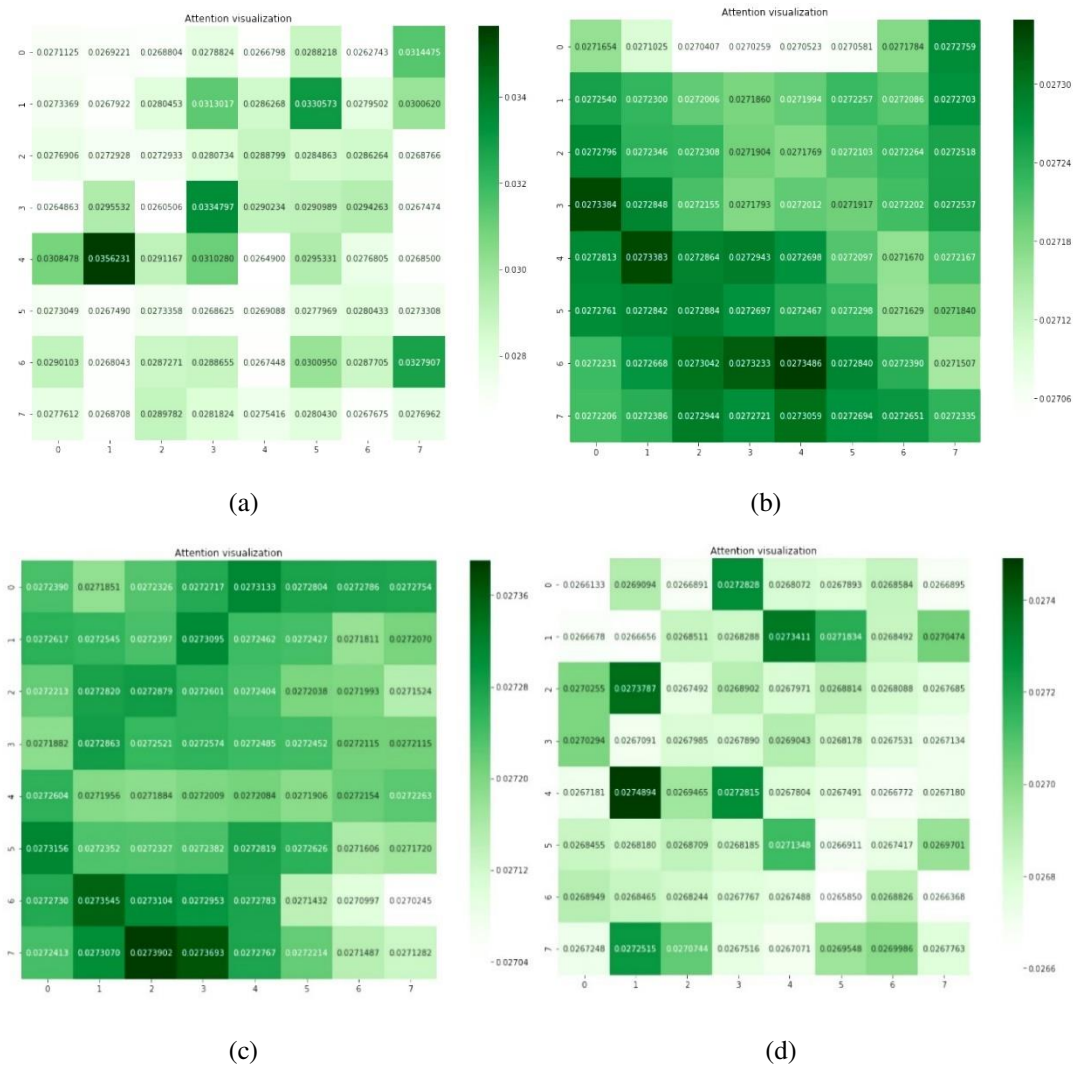
Table 1 shows the experimental results of different methods on the CWRU dataset. It can be seen from the results that the CALSTM method proposed in this study achieves a better accuracy, reaching 99.22%, which is better than other methods. The predicted results of traditional machine learning methods, such as decision tree, SVM and KNN are 52.67%, 59.44% and 68.89% respectively. And the diagnostic results of common neural network models such as DFCN(Deep Fully Connected network, FCN_3Layer), LSTM, BiLSTM and TCN(Temporal Convolutional Network) are 77.78%, 95.44% and 96.77%, respectively. From the data in Table 1, it can also be visually found that the diagnostic performance of LSTM, BiLSTM and TCN, which have been improved by hierarchical attention, have been improved to some extent. Similarly, the CALSTM method proposed in this study has been improved to a certain extent compared with the CLSTM method, and has a better diagnostic effect than various methods. Through a variety of comparative experiments, CALSTM shows its superiority in both the overall accuracy and the classification accuracy of each motor bearing state.

Table 1 Fault diagnosis accuracy of different algorithms on CWRU data set

Methods	Accuracy
FCN_1Layer	68.66%
FCN_2Layer	75.00%
FCN_3Layer	77.78%
SVM	59.44%
KNN	68.89%
Decision Tree	52.67%
LSTM	95.44%
LSTM+Attention	97.22%
BiLSTM	96.77%
BiLSTM+Attention	97.77%
TCN	97.66%
TCN+Attention	98.11%
CLSTM	98.33%
CALSTM	99.22%

3.2.3 Visualization of hierarchical attention

Fig. 9 shows the visualization of the Attention layer corresponding to any four feature maps in the test set. It can be intuitively seen that hierarchical attention mechanism gives different weights to different feature points. In the Fig. 9, the four hierarchical attention mechanism layers correspond to four randomly selected samples , which means that the proposed CALSTM method does not give a fixed weighting coefficient to the learned features of each sample, but adaptively weights the coefficients of the learning features of each sample. Although this approach poses a challenge to computation, it makes the model more flexible and ensures that the learning features of each sample can obtain a unique weight coefficient matrix.



(a) (b)

(c) (d)

Fig 9 The visualization of hierarchical attention on a test set

In Fig. 9, the darker the grid color is, the higher the weight coefficient of the corresponding feature point is, and the more important the feature learned by the feature map obtained by CNN network is. For example, as shown in Fig. 9 (a), the weight coefficient corresponding to the darkest grid is 0.0356231, which is the maximum value in the matrix. Therefore, after learning feature weighting, feature points at this position have a greater impact on the classification results. Hierarchical attention mechanism increases the weight expression for the feature map learned by CNN, and increases the relationship between the importance of the feature and the result. Moreover, the learning effect of CNN can be intuitively seen through the visualization of hierarchical attention mechanism, which enhances the interpretability of the neural network to a certain extent.

3.2.4 Related work

Table 2 lists different fault diagnosis methods, including improved SVM methods, popular SAE and DBN methods, and methods based on the combination of signal processing and machine learning, such as EEMD+AR+SVM, EMD+BP+LSFLA, etc., with different diagnosis effects. Compared with these methods, the CALSTM method proposed in this paper has better diagnosis effect and great development potential.

Table 2 Comparison of results of relevant work

	Method	Accuracy
Shao et al. 2015	DBN	87.45%
Wang et al.,2012	Hyper-sphere-structured multi-class support vector machine	91.99%
Li et al.,2020	SAE+DBN+a binary processor	98.53%
Yuwono et al.,2016	Particle swarm clustering, HMM	98.02%
Liang et al. ,2020	GDCCN(a novel capsule network with gate-structure dilated convolutions)	94.62%
Goyal et al. 2020	DWT+Mahalanobis distance criteria+SVM	95.8%
Qi et al. 2017	EEMD+AR+SVM	98.65%
	SAE	92.20%
Chen et al. 2020	MCNN+LSTM	98.46%
Zhang et al. 2015	SVM with EEMD	97.91%
Chen et al. 2017	hierarchical CNN	92.60%
Zhao et al. 2016	EMD+BP based on an improved shuffled frog leaping algorithm (ISFLA)	89.33%
Proposed method	CALSTM	99.22%

4、Conclusion

This study proposes a CALSTM fault diagnosis method. This method does not need to go through complex data preprocessing, but only normalizes the original signal and divides the data segment sample for the convenience of calculation. Such simple processing does not require the researcher to have the professional background of motor, or even need the researcher to know the relevant knowledge of signal processing. This method improves the CNN-LSTM method by adding a hierarchical attention mechanism layer to adaptively assign weights to the features of each sample learned by the neural network. This improvement enhances the performance of the model and achieves an accuracy of 99.2%. The diagnostic effect of this method surpasses common machine learning and deep learning methods, and has great potential in the field of motor bearing fault diagnosis. In the future work, we will consider applying the model to the embedded terminal to improve the computing efficiency of the model through model pruning.

Acknowledgements This study was supported by the National Key Research and Development Program 2017YFD0701702 and 2019YFD0901000

Compliance with ethical standards

Conflict of interest We declare that we have no conflict of interests.

Authors' contributions Yaoguang wei and Yisha Jiao conceived and designed the study. Yaoguang wei and Dong An proposed critical revisions to the experimental design. Dong An, Yaoguang Wei, Wenshu Li and Qiong Wei reviewed and edited the manuscript. All authors read and approved the manuscript.

References

Chen X, Zhang B, & Gao D (2020) Bearing fault diagnosis base on multi-scale cnn and lstm model. Journal of Intelligent Manufacturing (1). https://doi.org/10.1007/s10845-020-01600-2.

Choy FK, Jia W, & Wu R (2009) Identification of bearing and gear tooth damage in a transmission system. Tribology Transactions 52(3): 303-309. https://doi.org/10.1080/10402000802441611.

Cooley JW, Turkey JW (1965) An Algorithm for the Machine Calculation of the Complex Fourier Series. Mathematics of Computation 19(90): 297-301. https://doi.org/10.1090/S0025-5718-1965-0178586-1.

Goyal D, Choudhary A, Pabla BS, & Dhami SS (2020) Support vector machines based non-contact fault diagnosis system for bearings. Journal of Intelligent Manufacturing 31. https://doi.org/10.1007/s10845-019-01511-x.

Hoang DT, Kang HJ (2018) Rolling element bearing fault diagnosis using convolutional neural network and vibration image. Cognitive Systems Research 53(JAN.): 42-50. https://doi.org/10.1016/j.cogsys.2018.03.002.

Hochreiter S, & Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Jinde Z, Zhanwei J, Ziwei P, & Kang Z (2016) VMD based adaptive multiscale fuzzy entropy and its application to rolling bearing fault diagnosis. International Conference on Sensing Technology. IEEE. https://doi.org/10.1109/ICSensT.2016.7796267.

Lau E C C, Ngan H W (2010) Detection of Motor Bearing Outer Raceway Defect by Wavelet Packet Transformed Motor Current Signature Analysis. IEEE Transactions on Instrumentation & Measurement 59(10):2683-2690. https://doi.org/10.1109/TIM.2010.2045927

Li J, Li X, He D, & Qu Y (2020) Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor. Journal of Intelligent Manufacturing 1-18. https://doi.org/10.1007/s10845-020-01543-8.

Liang Y, Li B, & Jiao B (2020) A deep learning method for motor fault diagnosis based on a capsule network with gate-structure dilated convolutions. Neural Computing and Applications (6). https://doi.org/10.1007/s00521-020-04999-0.

Lu C, Wang Z, & Zhou B (2017) Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Elsevier Science Publishers B. V. https://doi.org/10.1016/j.aei.2017.02.005.

Mohanty S, Gupta KK, Raju KS (2015) Comparative study between VMD and EMD in bearing fault diagnosis. International Conference on Industrial & Information Systems. IEEE. https://doi.org/10.1109/ICIINFS.2014.7036515.

Qi Y, Shen C, Wang D, Shi J, & Zhu Z (2017) Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. IEEE Access PP(99):1-1. https://doi.org/10.1109/ACCESS.2017.2728010.

Shao H, Jiang H, Zhang X, Niu M (2015) Rolling bearing fault diagnosis using an optimization deep belief network. Measurement ence and Technology 26(11): 115002. https://doi.org/10.1088/0957-0233/26/11/115002.

Shao HD, Jiang HK, Zhao HW (2017) A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. Mechanical Systems and Signal Processing 95: 187-204. https://doi.org/10.1016/j.ymssp.2017.03.034.

Siddiqui K M, Sahay K., & Giri VK (2016) Early, diagnosis of bearing fault in the inverter driven induction motor by wavelet transform. 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT). IEEE. https://doi.org/10.1109/ICCPCT.2016.7530233.

Tamilselvan P, Wang P (2013) Failure diagnosis using deep belief learning based health state classification. Reliability Engineering & System Safety 115(jul.): 124-135. https://doi.org/ 10.1016/j.ress.2013.02.022.

Tsao W C, Li YF, Le DD, & Pan M C (2012) An insight concept to select appropriate imfs for envelope analysis of bearing fault diagnosis. Measurement 45(6):1489-1498. https://doi.org/10.1016/j.measurement.2012.02.030.

Wang Y, Kang S, Jiang Y, Yang G, Song L, & Mikulovich VI (2012) Classification of fault location and the degree of performance degradation of a rolling bearing based on an improved hyper-sphere-structured multi-class support vector machine. Mechanical Systems & Signal Processing 29(none): 404-414. https://doi.org/10.1016/j.ymssp.2011.11.015.

Wen L, Gao L, & Li X (2017) A new deep transfer learning based on sparse auto-encoder for fault diagnosis. IEEE Transactions on Systems, Man, and Cybernetics: Systems PP(99):1-9. https://doi.org/10.1109/TSMC.2017.2754287.

Wen L, Li X, Gao L, & Zhang Y (2017) A new convolutional neural network based data-driven fault diagnosis method. IEEE Transactions on Industrial Electronics PP(99): 1-1. https://doi.org/10.1109/TIE.2017.2774777.

Yang Z, Yang D, Dyer C, He X, & Hovy E (2016) Hierarchical Attention Networks for Document Classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. https://doi.org/10.18653/v1/N16-1174.

Yuwono M, Qin Y, Zhou J, Guo Y, Celler BG, & Su SW (2016) Automatic bearing fault diagnosis using particle swarm clustering and hidden markov model. Engineering Applications of Artificial Intelligence 47(JAN.): 88-100. https://doi.org/10.1016/j.engappai.2015.03.007.

Zhang X, Liang Y, Zhou J, & Zhang Y (2015) A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized svm. Measurement 69: 164-179. https://doi.org/10.1016/j.measurement.2015.03.017.

Zhang Z, Wang Y, & Wang K (2013) Fault diagnosis and prognosis using wavelet packet decomposition, fourier transform and artificial neural network. Journal of Intelligent Manufacturing 24(6):1213-1227. https://doi.org/10.1007/s10845-012-0657-2.

Zhao Z, Xu Q, & Jia M (2016) Improved shuffled frog leaping algorithm-based bp neural network and its application in bearing early fault diagnosis. Neural Computing and Applications 27(2): 375-385. https://doi.org/10.1007/s00521-015-1850-y.

Figures

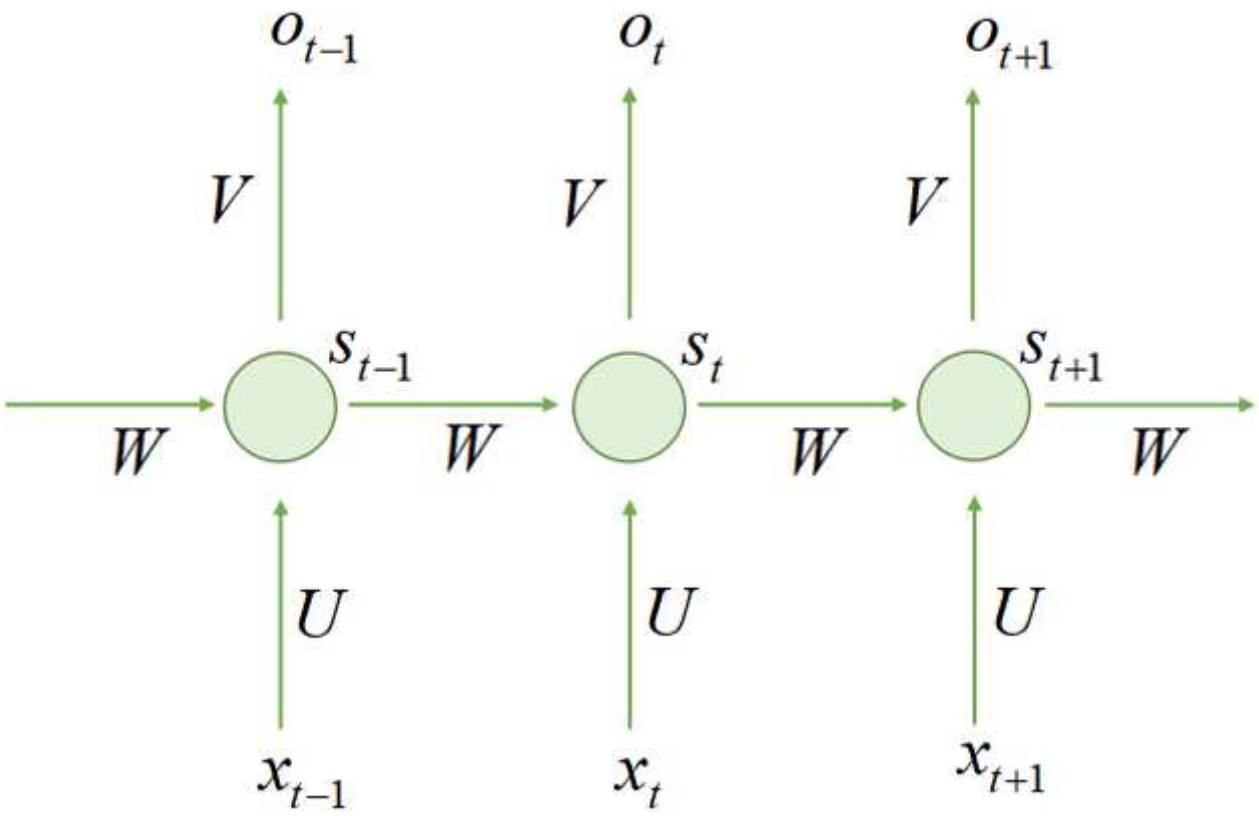


Figure 1

Structure diagram of circulating neural network

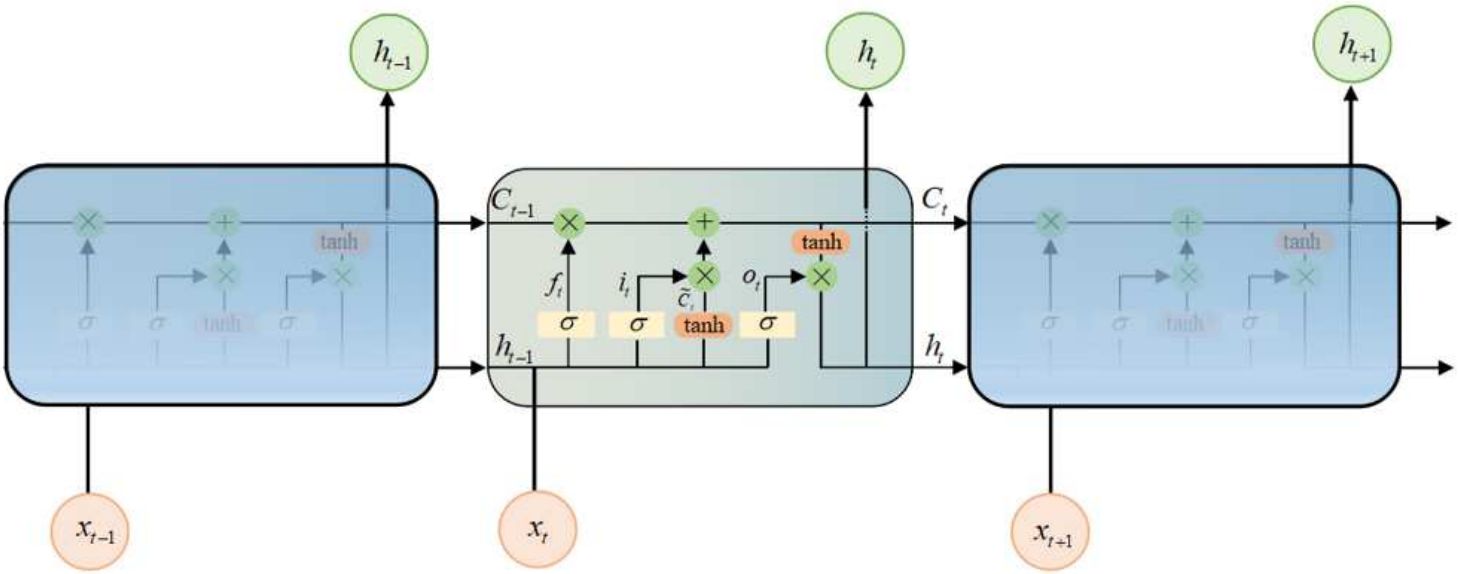


Figure 2

LSTM network structure diagram

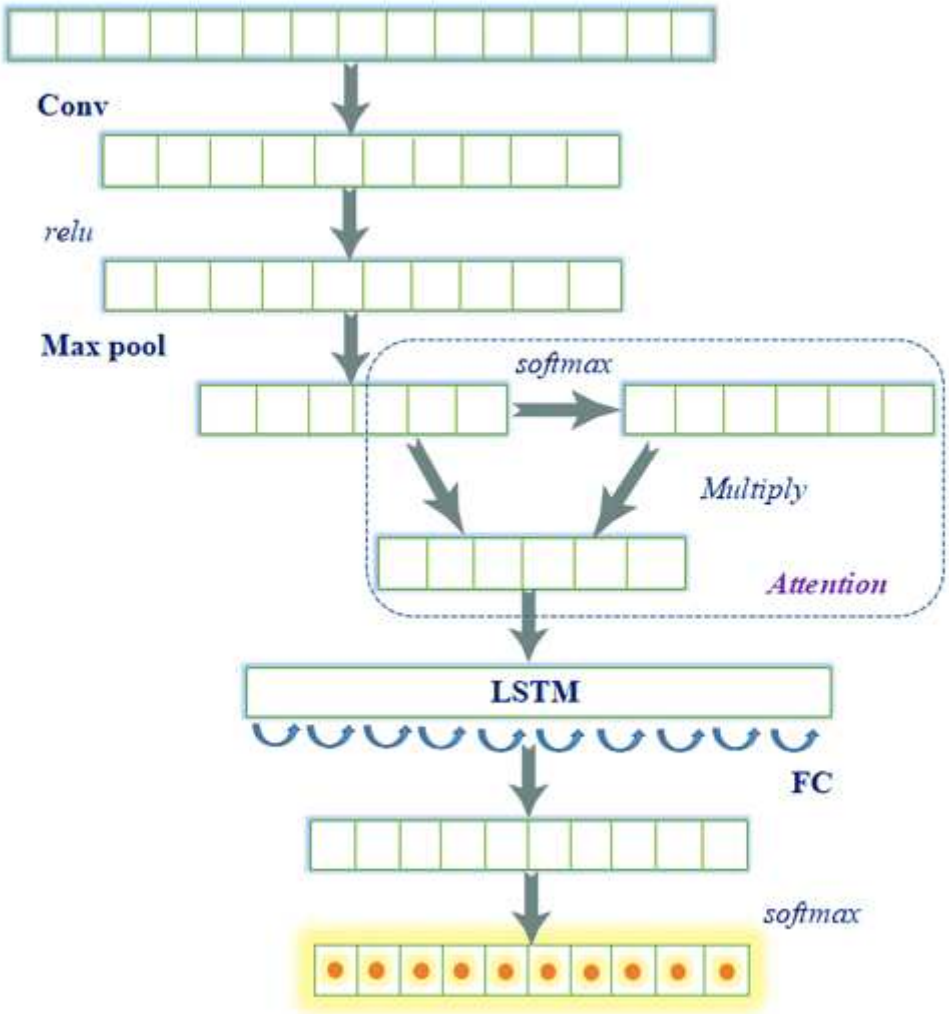


Figure 3

The basic framework of the CALSTM model

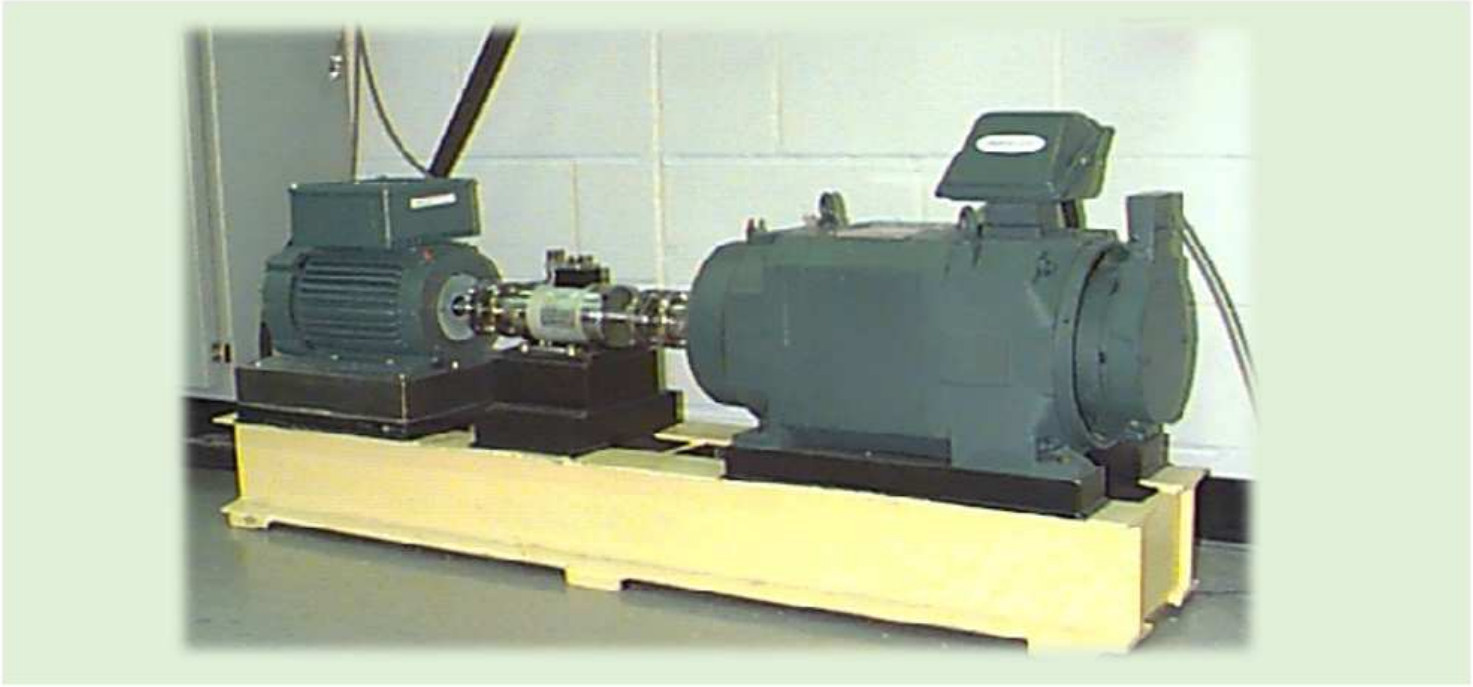


Figure 4

CWRU rolling bearing data set acquisition platform

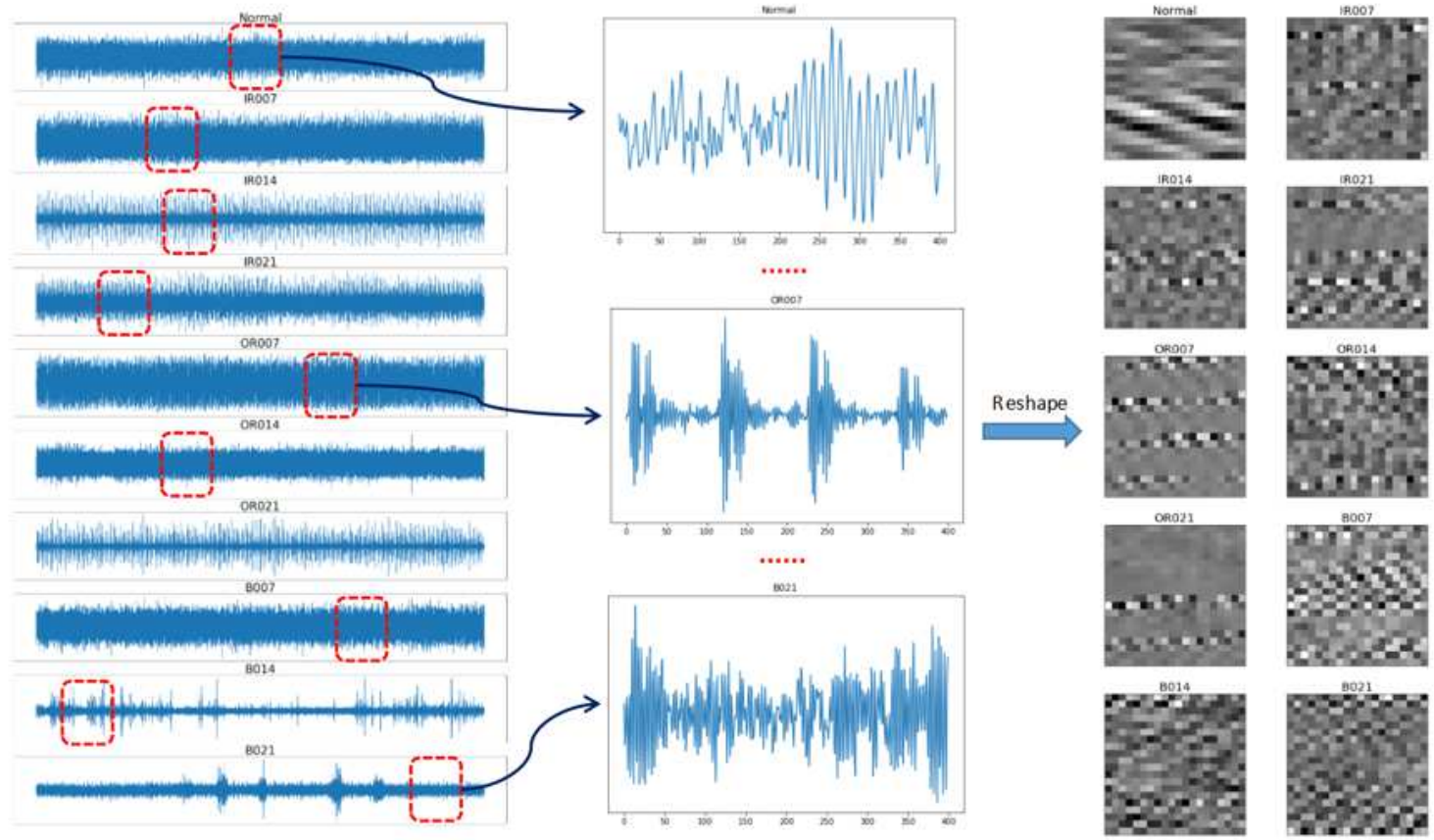


Figure 5

Data preprocessing process diagram of vibration signal

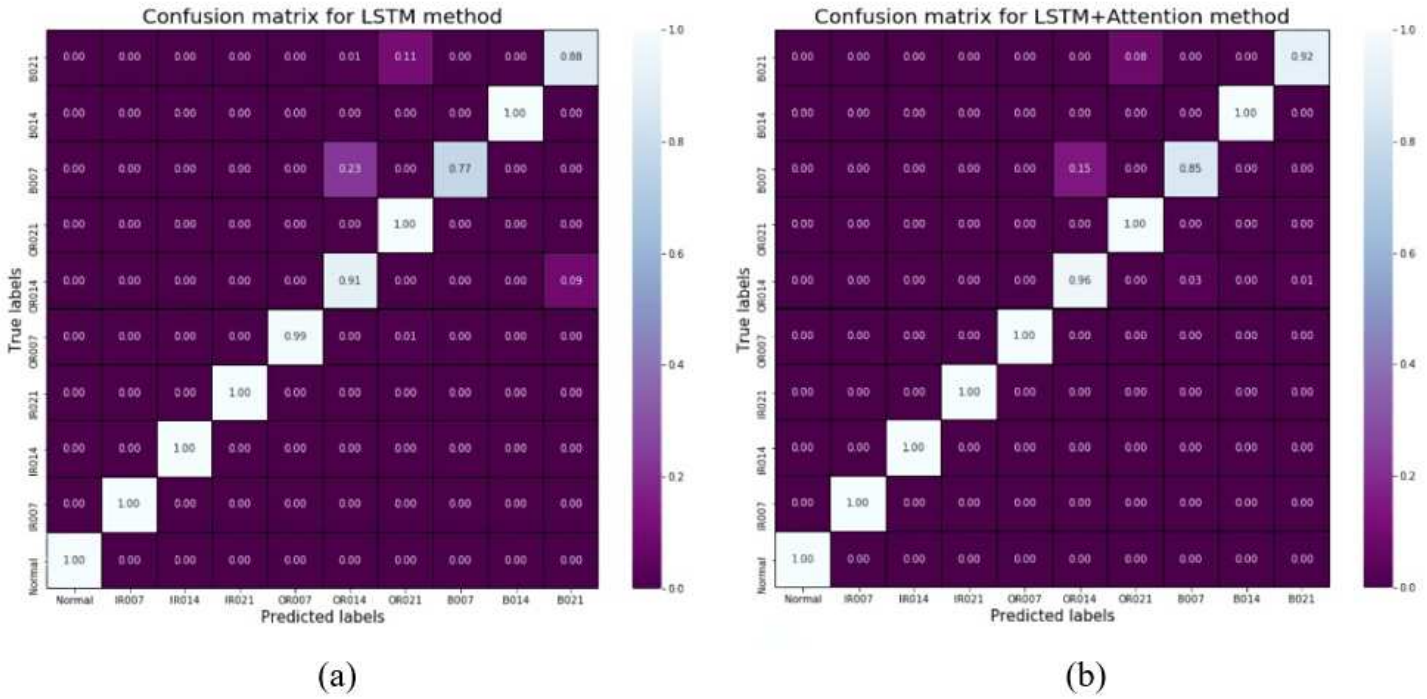


Figure 6

LSTM and LSTM+Attention fault classification confusion matrix

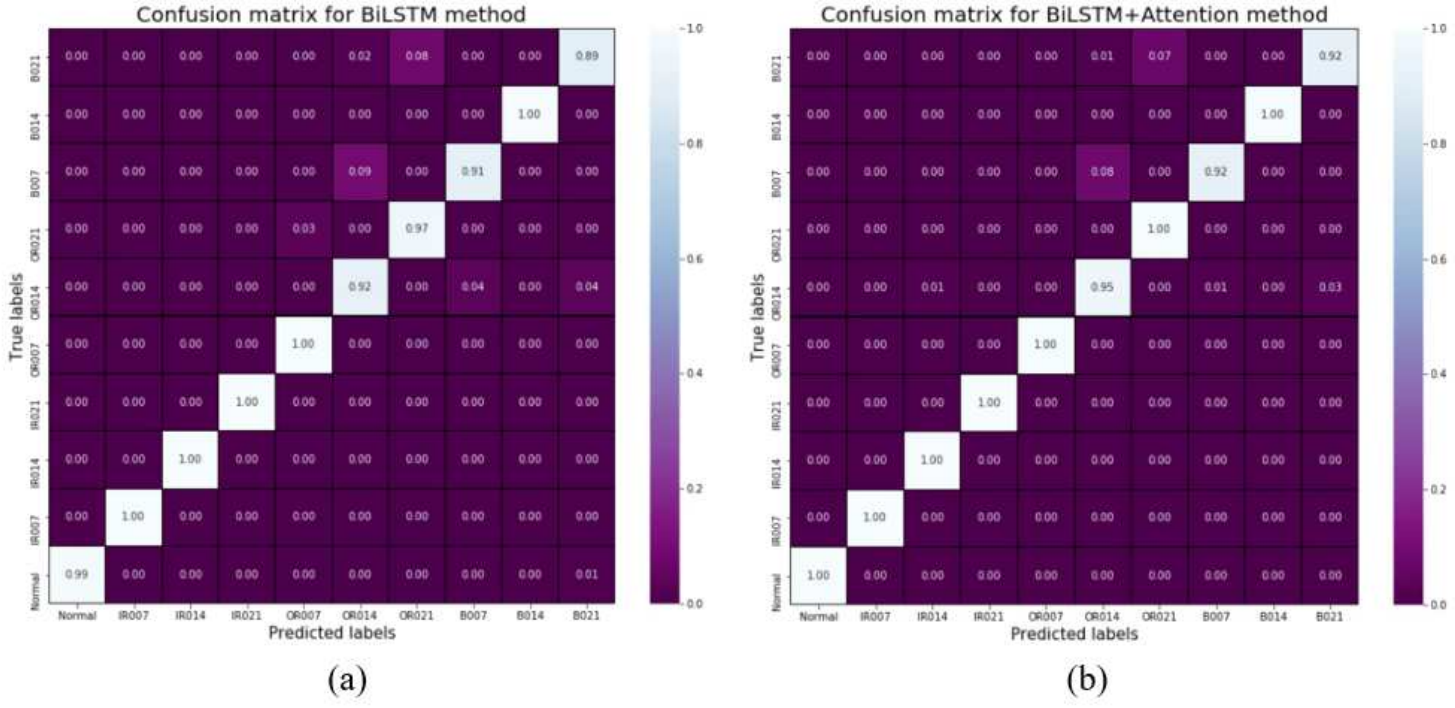


Figure 7

BiLSTM and BiLSTM+Attention fault classification confusion matrix

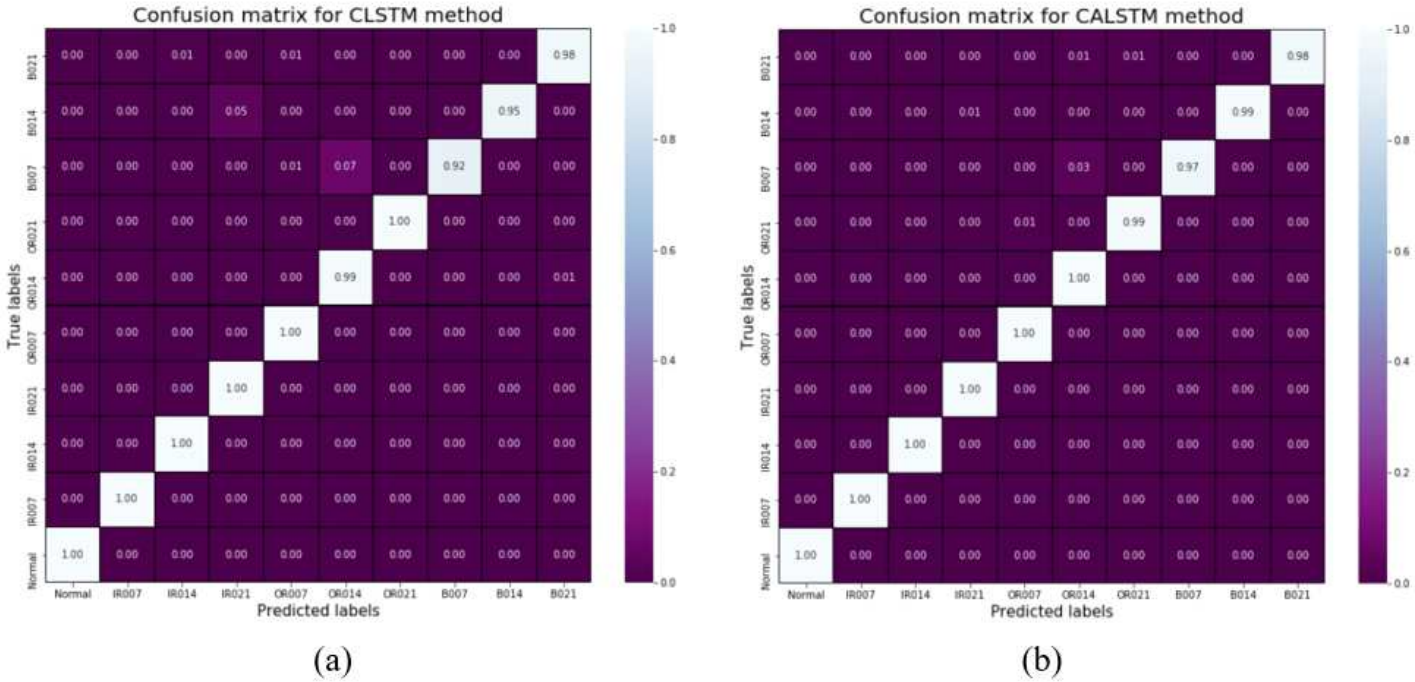


Figure 8

CLSTM and CALSTM fault classification obfuscation matrix

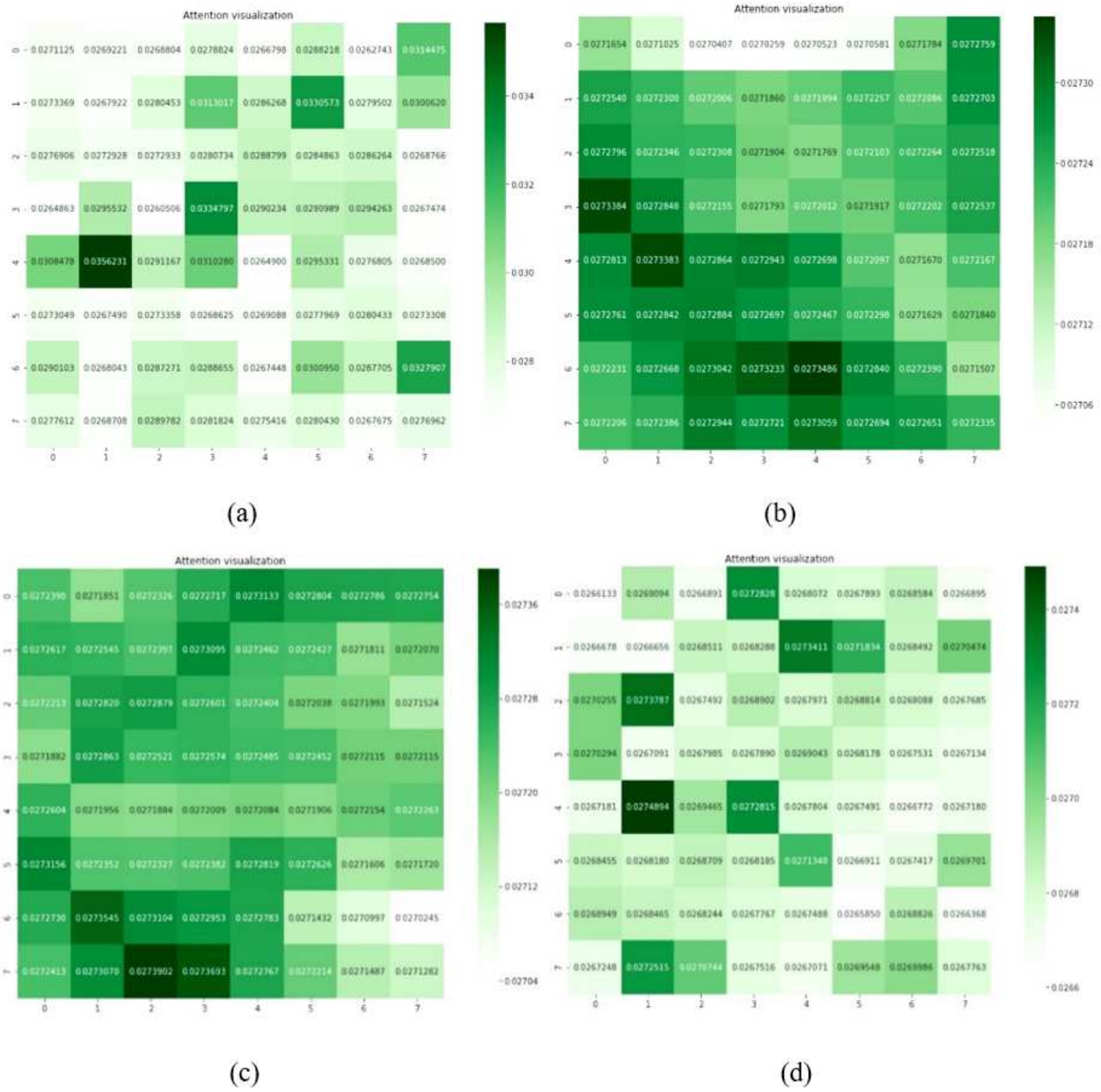


Figure 9

The visualization of hierarchical attention on a test set