Fault Diagnosis of Gearbox Based on CNN and LSTM with Attention Mechanism

Kefan Cheng, Longsheng Cheng *, Ting Mao, Mengdie Wu

School of Economics and Management, Nanjing University of Science and Technolog, Nanjing, China

* Corresponding Author Email: cheng_longsheng@163.com

Abstract. Gearbox is a key component of mechanical equipment, which has a complex structure, harsh working conditions, and a higher probability of failure. Therefore, gearbox fault diagnosis is vital to ensure the efficient operation of the whole mechanical equipment. However, traditional gearbox fault diagnosis methods mainly rely on manual feature extraction. To address this issue, a novel end-to-end fault diagnosis model that combines convolutional neural network (CNN), long short-term memory network (LSTM) and attention mechanism (AM) is proposed. Firstly, Spatial features are extracted from the original input data using CNNs, and then temporal features are extracted even further from the spatial features using LSTMs. The attention mechanism improves the network's attention to the global key features by assigning weights, and finally, the fault diagnosis results are derived from the fully connected layer and Softmax classifier. The experimental results show that the method is able to adaptively extract features and ensure higher diagnostic accuracy, validating the effectiveness of the proposed method.

Keywords: Gearbox, fault diagnosis, feature extraction, CNN, LSTM.

1. Introduction

As a core component of rotating mechanical equipment, gearboxes play a crucial role in wind power, aerospace, transportation and other fields. Modern gearboxes are developing in the direction of complexity and high reliability, and the realization of accurate and efficient fault diagnosis is of great significance for improving their operational reliability and maintenance economy [1]. The research and improvement of gearbox fault diagnosis methods are not only for the long-term safe and stable operation of mechanical equipment, but also for ensuring production efficiency and reducing casualties have very necessary practical significance and application value [2].

Fault diagnosis serves an important role in pursuing the relationship between the monitoring data and the health states of machines, which has been a widely concerned issue in machine health management [3]. Shallow neural networks based on artificial feature extraction, such as random forest (RF), support vector machine (SVM), extreme learning machine (ELM), etc., have been widely used in the fault diagnosis of machinery and equipment [4]. Roy et al [5] proposed an autocorrelation-assisted feature extraction method, in which the dimensionality-decreased feature vector is input into a random forest classifier to classify faults in the vibration signals of rolling bearings. Chen et al [6] proposed a rolling bearing early fault diagnosis method based on orthogonal neighborhood preserving embedding and Adaboost_SVM algorithm. Du et al [7] proposed a discriminative manifold ELM self-encoder which utilizes the geometric features and labeling information of the edge distributions of the training samples, and then uses a kernel ELM instead of the traditional ELM classifier for the final classification of the fault patterns. However, there is a certain amount of artificial feature extraction in traditional fault diagnosis methods, and it is difficult to adapt to the trend of intelligent manufacturing in terms of fault pattern recognition due to the limitation of the classifier's own shallow structure, which makes it have certain limitations in learning complex nonlinear relationships and processing large-scale data, especially for mechanical equipment with complex internal structure and dynamic characteristics.

In recent years, deep neural networks such as autoencoders and convolutional neural networks have been widely used to construct end-to-end intelligent diagnostic models, which reduces the dependence on manual labor and expert knowledge and greatly promotes the development of intelligent fault diagnosis [8]. The end-to-end diagnostic model directly inputs raw data into the constructed machine learning model for training and combines it with classification methods to realize fault diagnosis without relying on any a priori knowledge [9]. Shao et al [10] proposed a new method for intelligent diagnosis of rolling bearing based on the integrated deep autoencoder effectively overcomes the limitations of individual deep learning models. In [11] the gearbox signal was first transformed into a time-frequency spectrogram using wavelet transform, and then a fault health classification model based on a convolution neural network (CNN) was proposed. Yin et al [12] proposed a fault diagnosis method for wind turbine gearboxes based on cosine loss optimized LSTM neural network (Cos-LSTM), which eliminates the influence of the signal strength and improves the diagnostic accuracy.

In this paper, a novel end-to-end fault diagnosis model combining convolutional neural network, long short-term memory network and attention mechanism is proposed. Firstly, CNN is used to extract spatial local features from the original input data, then LSTM is used to further extract temporal global features, and then more attention is focused on the key feature information through the attention mechanism, so as to enhance the utilization of effective features and improve the model generalization. Finally, the fault diagnosis results are derived through the fully connected layer and Softmax classifier.

The rest of the paper is organized as follows. Section 2 briefly introduces the fundamentals of CNNs, LSTMs and attention mechanisms; Section 3 describes the structure of the proposed model; Section 4 verifies the effectiveness of the proposed method through experiments; and Section 5 gives conclusions.

2. Methodology

2.1. Convolutional Neural Network

CNN is a kind of deep feedforward neural network with weight sharing and local connectivity features [13], which has very powerful deep feature extraction and pattern recognition capabilities, and has been widely used in many fields such as image recognition. A typical convolutional neural network mainly consists of a convolutional layer, a pooling layer and a fully connected layer.

The convolutional layer is mainly used to extract the implicit feature information of the input signal, using the convolutional kernel to perform the convolution operation on the local region of the input data and use the activation function to further generate the output features, the convolutional layer operates as follows:

$$x_i^{k+1} = f(\sum_j x_j^k * w_{ij}^{k+1} + b_i^{k+1}) \tag{1}$$

Where: x_i^{k+1} is the ith feature tensor output from layer k+1, x_j^k is the jth feature tensor output from layer k, w_{ij}^{k+1} is the convolution kernel with j inputs connecting the ith feature tensor, b_i^{k+1} is the k+1 layer bias term, f is the activation function, and * is the convolution operation.

The activation function of CNN can get the nonlinear expression of the input signal and finally decide whether to send the signal and what to send to the next neuron. In this paper, we use ReLU activation function, ReLU can add some nonlinear factors in the network, which can better deal with and solve complex problems, in addition, ReLU arithmetic is fast and does not have gradient problems, its expression is:

$$z_i^{k+1}(j) = \left[y_i^{k+1}(j)\right] = max\,\{0, y_i^{k+1}(j)\} \tag{2}$$

Where: $y_i^{k+1}(j)$ is the convolution operation output value and $z_i^{k+1}(j)$ is the activation value.

The pooling layer is generally connected to the convolutional layer, which reduces the spatial size of the features and network parameters using down sampling operations to reduce the feature dimensions and avoid overfitting. In order to reduce the parameters while retaining the fault characteristics of the periodic time-domain signal as much as possible, the maximum pooling method is chosen in this paper, which is calculated as:

$$p_i^{k+1}(j) = \max_{(j-1)W+1 \le t \le jW} \{q_i^k(j)\} \tag{3}$$

Where: $q_i^k(j)$ is the value of the tth neuron of the ith feature tensor in layer k, W is the width of the pooling area, and $p_i^{k+1}(j)$ is the jth value of the ith feature tensor in layer k+1.

After convolution and pooling operations will connect a fully connected layer, through the form of fully connected using different classification algorithms to classify the input data, in this paper, the fully connected layer is classified using the softmax function, whose expression is:

$$\alpha_i = softmax(\varphi_i) = \frac{e^{\varphi_i}}{\sum_{i=1}^{C} e^{\varphi_i}} \tag{4}$$

Where: C is the number of categories; φ_i is the output value of the ith category; α_i is the probability of the ith category.

2.2. Long and Short Term Memory Network

LSTM evolved from recurrent neural networks, which not only consider the current information, but also make full use of the previous information to complete the prediction of the upcoming state, and its neurons can keep the memory in its channel, accumulate the experience in training, and can better handle long-term dependent data, which can effectively solve the problems such as gradient explosion [14]. Its structure is shown in Fig 1.

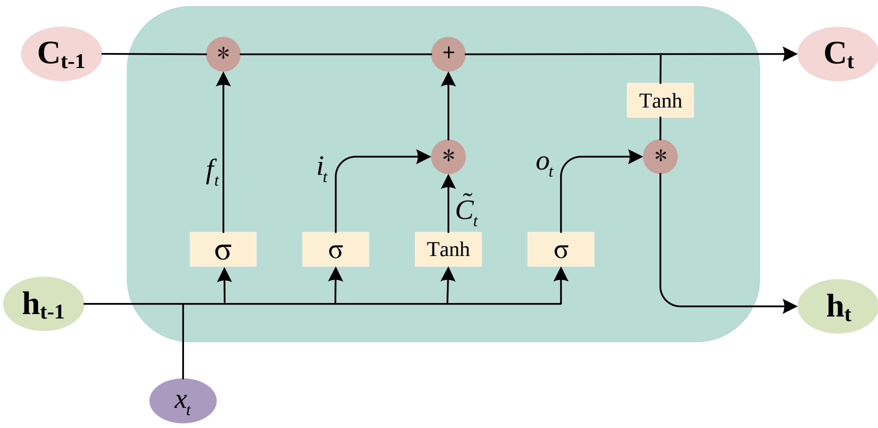

Figure 1. Structure of LSTM.

LSTM controls the long-term memory of the network by adding and removing information from the cell state through 3 control gates: forgetting gate, input gate, and output gate.

The role of the forgetting gate is to decide what information should be discarded or retained and to selectively filter the features, which is calculated as:

$$f_t = \sigma(w_{L1}[h_{t-1}, x_t] + b_{L1}) \tag{5}$$

Where: σ is the sigmoid function, w_{L1} is the weight matrix of the oblivious gate, b_{L1} is the bias term, x_t is the current input, and h_{t-1} is the hidden state in the previous moment.

The role of the input gate is to decide which information is updated, which is calculated as:

$$i_t = \sigma(w_{L2}[h_{t-1}, x_t] + b_{L2}) \tag{6}$$

$$\tilde{C}_t = tanh(w_{L3}[h_{t-1}, x_t] + b_{L3}) \tag{7}$$

Where: tanh is the activation function, w_{L2} and w_{L3} are the weight matrices of the input gates, b_{L2} and b_{L3} are their bias terms, and $(\widetilde{C_t})$ is the memory cell updated at the current moment.

According to the vectors f_t and i_t, the storage unit C_t at the current moment can be obtained, and the combination of the two can realize the fusion of short-term memory and long-term memory, as shown in Eq:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{8}$$

The role of the output gate is to control the output of information from the internal state to the external state at the current moment, which is calculated as follows:

$$O_t = \sigma(w_{L4}[h_{t-1}, x_t] + b_{L4}) \tag{9}$$

$$h_t = O_t * tanh(C_t) \tag{10}$$

Where: w_{L4} and b_{L4} are the weight matrix and bias term of the outputs, respectively; O_t is the output of the output gate; h_t is the hidden state at the current moment.

2.3. Attention Mechanism

AM is a resource allocation mechanism that simulates the visual signal processing system of the human brain, which, by rapidly scanning a large amount of information to be observed and utilizing limited computational resources, skillfully and reasonably shifts the attention to the focus of attention and the target information, ignores irrelevant information and amplifies important information, so as to enable the model to efficiently extract more important and more useful feature information.

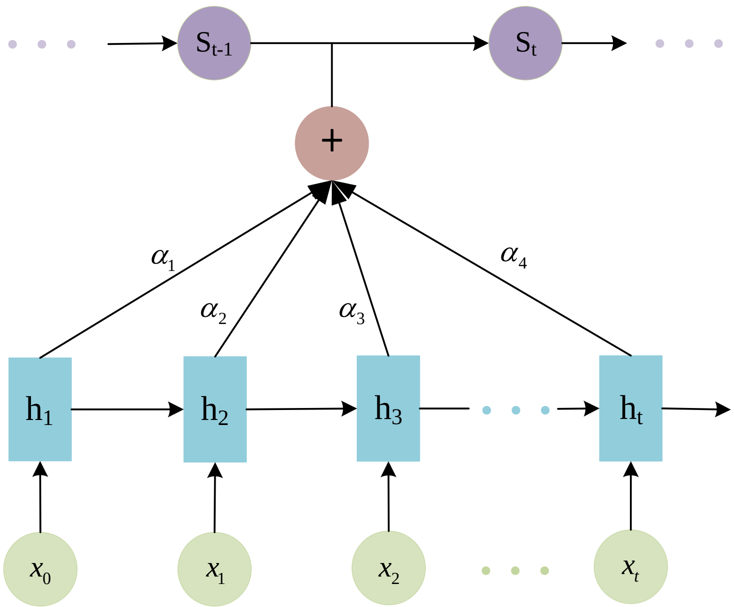

Figure 2. Structure of AM.

The structure of the attention mechanism is shown in Fig. 2, and its essence is to target the importance of certain feature information by using weights of different sizes for each part of the output feature, and using the weights to indicate the size of the attention given. Its calculation formula is:

$$e_t = utanh(wh_t + b) \tag{11}$$

$$\alpha_t = softman(e_t) = \frac{exp(e_t)}{\sum_t exp(e_t)} \tag{12}$$

$$s_t = \alpha_t c_t \tag{13}$$

Where: u and w are the weight coefficients, b is the bias coefficient, e_t is the probability distribution of attention at the tth moment, α_t is the weight term in the attention model, and s_t denotes the output of the attention mechanism layer at the tth moment.

3. The Proposed Method

Gearboxes operate in a complex environment, and the effective components of the acquired signals are often overwhelmed by noise. Considering that the CNN network can adaptively extract spatial features in the signal, and the LSTM network can effectively learn the temporal correlation of vibration signals, this paper organically integrates the two kinds of networks and introduces the attention mechanism to obtain rich information about the fault features, so as to improve the utilization of the effective features, and constructs a deep feature extraction network model based on the CNN-LSTM-Attention, and its structure is shown in Figure 3.

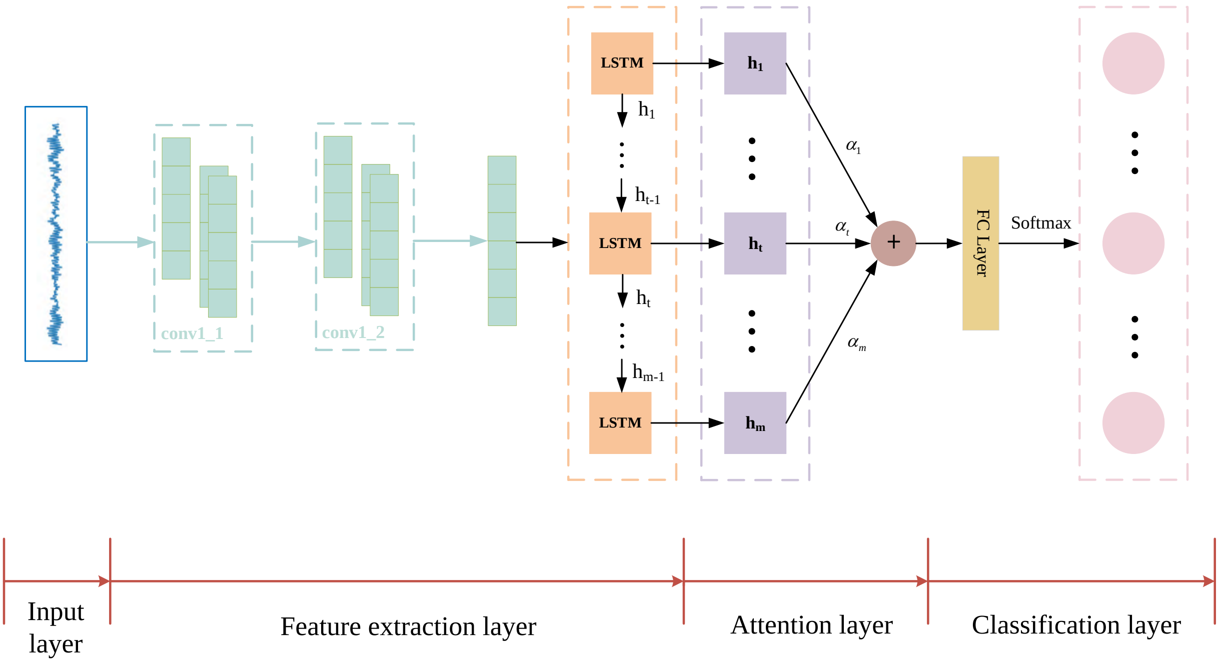

Figure 3. Network structure of CNN-LSTM-Attention.

The CNN-LSTM-Attention model is mainly composed of an input layer, a feature extraction layer, an attention layer and a classification layer. In the input layer, the original signal of the gearbox is in the form of vectors, which can be directly used as the input of the model. In the feature extraction layer, feature extraction is carried out by CNN and LSTM respectively. Firstly, CNN is used to fully extract spatial features from the original vibration signal, and then the extracted spatial features are directly inputted into the LSTM layer, which is utilized to extract the temporal features of the vibration signal, so as to obtain the high-level features that contain both spatial and temporal features. In the attention layer, the output features of each moment will get the corresponding weights through the attention mechanism, which improves the network's attention to the global key features, thus highlighting the features that are more favorable for classification. Finally, in the classification layer, the fault diagnosis results are derived through the fully connected layer and Softmax classifier. The model training process is shown in Fig. 4.

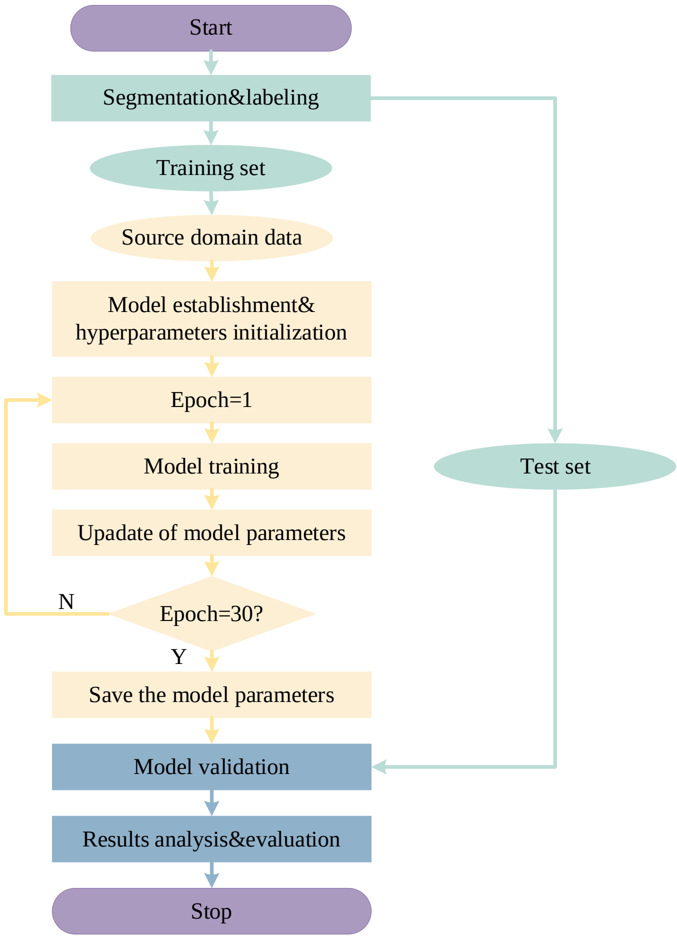

Figure 4. Flow chart of model training.

Step1: Divide the data into training set and test set, use 80% as training set and the remaining 20% as test set;

Step2: Build the network model and initialize the corresponding hyperparameters in the model, and use the training set as the input of the model for training;

Step3: If the specified number of training times is not reached, the model will continue to be trained; after the specified number of training times is reached, the final model will be obtained and the parameters of the trained model will be saved;

Step4: Use the test set to validate the model, compare and evaluate it with different fault diagnosis models, and analyze the results.

4. Experiment and Results Analysis

4.1. Data Description

In order to verify the validity of the proposed model, this paper conducts experiments using the gearbox dataset from Southeast University (SEU), which is provided by SEU and collected from the Dynamic Simulator of Drivetrain (DDS) [15]. As shown in figure 5, the DDS consists of six main components: brake, motor, motor controller, parallel gearbox, planetary gearbox and brake controller.

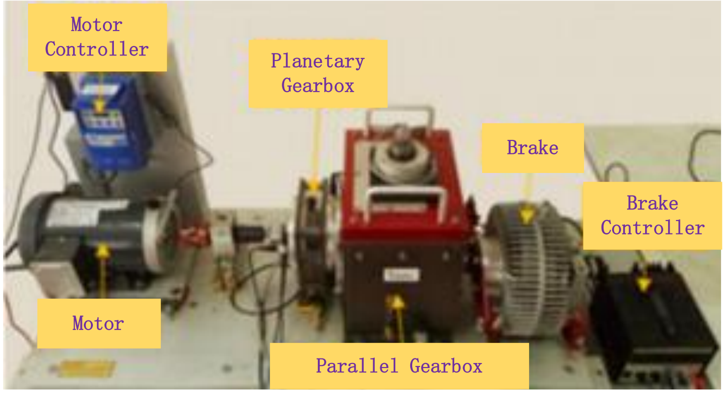

Figure 5. Structure of DDS.

The dataset is set up with a total of two working conditions of speed and load configurations of 20Hz-0V and 30Hz-2V, and four gear failure state signals and one gear health state (NC) signal are collected at the same time. Each state signal includes eight channels of signals, where channel 1 represents the motor vibration signal, channel 2, channel 3 and channel 4 represent the vibration signals of planetary gearboxes in the x, y and z directions, respectively, channel 5 represents the motor torque signal, and channel 6, channel 7 and channel 8 represent the vibration signals of parallel gearboxes in the x, y and z directions, respectively. The four types of gear failure data include chipped, missing, root and surface (chipped, missing, root and surface), and the detailed dataset type divisions for both conditions are shown in Table 1.

Table 1. Classification of dataset types

Fault type	Label	Data length	Training sample size	Test sample size
NC	0	1024	800	200
Chipped	1	1024	800	200
Miss	2	1024	800	200
Root	3	1024	800	200
Surface	4	1024	800	200

4.2. Experimental Settings

In this paper, the data under the speed and load of 20Hz-0V working condition are selected for experimental validation, and the network model is implemented and trained using the deep learning framework Keras and python software. Before the model training, the appropriate hyperparameters must be carefully selected to improve the classification accuracy, accelerate the model convergence and enhance the robustness. Referring to previous research [16], this paper identifies the use of Adam optimizer and finalizes the experimental parameters of the model through grid search, as shown in Table 2.

Table 2. experimental parameters

Setting	Batch size	Maximum epoch	Optimizer	Learning tate
Value	64	30	Adam	0.0001

4.3. Result Analysis

The gearbox dataset with sufficient number of samples is trained using the model proposed in this paper, and the accuracy and loss of the training set and test set during the training process are shown in Fig 6 and Fig 7. It can be seen that the loss values on the training and test sets are rapidly decreasing and the classification accuracy is also rapidly increasing, indicating that the model effectively extracts the fault features and the training is in the right direction. Although there is a small fluctuation in the middle of the training process, this has no effect on the subsequent convergence of the model, and after 30 iterations, the final classification accuracy on the test set is 98.5%.

In order to refine and analyze the diagnostic results of this paper's method, a confusion matrix is plotted, as shown in Fig. 8, with the horizontal axis being the predicted categories, the vertical axis being the true categories, and the diagonal line being the diagnostic accuracy. Among them, the fault diagnosis accuracy of surface faults, root faults and chip faults are almost 100%, and the diagnosis accuracy of the miss faults is above 96%. Therefore, the method in this paper has good feature extraction capability and fault diagnosis performance.

Figure 6. Training and Test Set Accuracy.

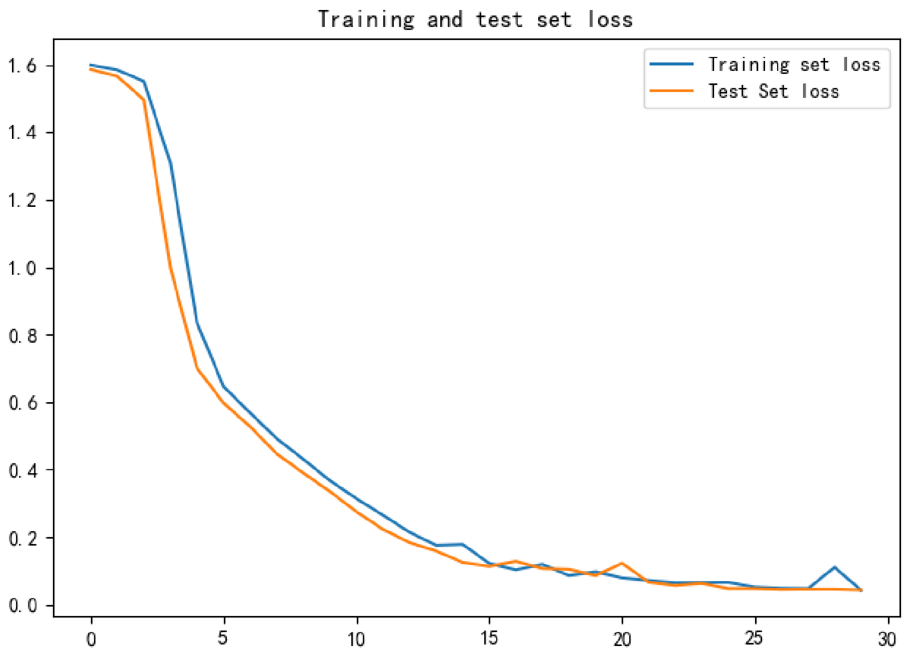

Figure 7. Training and Test Set Loss.

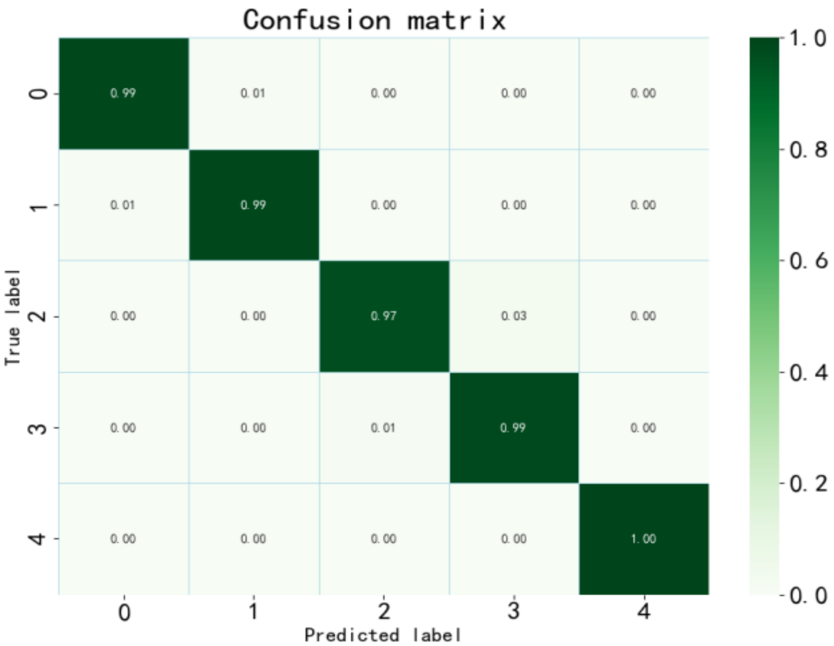

Figure 8. Confusion Matrix.

To further validate the performance of the proposed model, it is compared with four models such as LSTM, CNN, CNN-LSTM and CNN-AM. Ten experiments are conducted simultaneously for each of the four models and the average of the experimental results is taken for model performance evaluation, and the results are shown in Table III. Among them, the LSTM method has the lowest fault diagnosis accuracy, poor diagnostic effect, and the worst generalization performance due to the fact that it is difficult to extract the high-level features from the data using only the LSTM structure, and it lacks the ability to extract the fault features. The other three models using the convolutional structure achieved more than 92% diagnostic accuracy, indicating that the convolutional structure possesses stronger fault feature extraction capability. Compared with the CNN-LSTM model, the method in this paper incorporates the attention mechanism in the model, which improves the average accuracy by 3.68%, indicating that the attention mechanism can effectively improve the performance of the model. Therefore, the proposed method in this paper achieves the best fault diagnosis performance, possesses stronger fault feature extraction capability as well as better generalization.

Table 3. Experimental parameters

Model	LSTM	CNN	CNN-LSTM	CNN-AM	The proposed method
Average accuracy	90.75	92.9	94.82	96.28	98.5
Average loss	0.252	0.2510	0.1218	0.1112	0.0459

5. Conclusion

In this paper, a CNN-LSTM for gearbox fault diagnosis model incorporating the attention mechanism is proposed. Firstly, CNN is utilized to fully extract spatial features from the original vibration signals, and then the extracted spatial features are directly inputted into the LSTM layer, so that the model possesses both spatial and temporal feature extraction capabilities. And the attention mechanism is introduced to focus on important fault features, which enhances the model's ability to select and extract fault features. In this paper, experiments are conducted using the gearbox dataset of Southeast University to verify the effectiveness of the proposed method and realize high-performance fault diagnosis. And it has higher diagnostic accuracy compared with other models.

In the future, the model will continue to be improved to obtain higher accuracy and more stable performance. Meanwhile, the combination of transfer learning method is considered to expand the application of the model in the engineering field, so as to enhance the generality of the model.

References

[1] Liu R, Yang B, Zio E, and Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review [J]. Mechanical Systems and Signal Processing, 2018, 108: 33 - 47.

[2] Senanayaka J S L, Van Khang H, and Robbersmvr K G. CNN based Gearbox Fault Diagnosis and Interpretation of Learning Features [C]//2021 IEEE 30th International Symposium on Industrial Electronics (ISIE). IEEE, 2021: 1 - 6.

[3] Lei Y, Yang B, Jiang X, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap [J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.

[4] Han S, Zhong X, Shao H, Zhao R and Cheng J. Novel multi-scale dilated CNN-LSTM for fault diagnosis of planetary gearbox with unbalanced samples under noisy environment [J]. Measurement Science and Technology, 2021, 32 (12): 124002.

[5] Roy S, Dey S, and Chatterjee S. Autocorrelation aided random forest classifier-based bearing fault detection framework [J]. IEEE Sensors Journal, 2020, 20 (18): 10792 - 10800.

[6] Chen F, Cheng M, Tang B, et al. Pattern recognition of a sensitive feature set based on the orthogonal neighborhood preserving embedding and adaboost_SVM algorithm for rolling bearing early fault diagnosis [J]. Measurement Science and Technology, 2020, 31 (10): 105007.

[7] Du G, Xu Q, and Yang X. Fault diagnosis of rotating machinery components using a deep kernel extreme learning machine under different working conditions [J]. Measurement Science and Technology, 2020, 31 (11): 115901.

[8] Zhang T, Chen J, Li F, et al. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions [J]. ISA transactions, 2022, 119: 152 - 171.

[9] Sun H B, Fan Y G. Fault diagnosis of rolling bearings based on CNN and LSTM networks under mixed load and noise [J]. Multimedia Tools and Applications, 2023: 1 - 25.

[10] Shao H D, Jiang H, Lin Y, and Li X. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto—encoders [J]. Mechanical systems and signal Processing, 2018, 102: 278 — 297.

[11] Sikder N, Bhakta K, Al Nahid A, and Lslam. Fault diagnosis of motor bearing using ensemble learning algorithm with FFT-based preprocessing [C]//2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). IEEE, 2019: 564 - 569.

[12] Yin A, Yan Y, Zhang Z, et al. Fault diagnosis of wind turbine gearbox based on the optimized LSTM neural network with cosine loss [J]. Sensors, 2020, 20 (8): 2339.

[13] Zhang F, Zhu Y, Zhang C, Yu P and Li Q. Abnormality Detection Method for Wind Turbine Bearings Based on CNN-LSTM [J]. Energies, 2023, 16 (7): 3291.

[14] Sun H B, Fan Y G. Fault diagnosis of rolling bearings based on CNN and LSTM networks under mixed load and noise [J]. Multimedia Tools and Applications, 2023: 1 - 25.

[15] Shao S, McAleer S, Yan R, and Baldi P. Highly accurate machine fault diagnosis using deep transfer learning [J]. IEEE Transactions on Industrial Informatics, 2018, 15 (4): 2446 - 2455.

[16] Yang S, Xiang Y, Long Z, et al. Fault Diagnosis of Harmonic Drives Based on an SDP-ConvNeXt Joint Methodology [J]. IEEE Transactions on Instrumentation and Measurement, 2023.