ISSN 1330-3651(Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20220512135354
Original scientific paper

Bearing Fault Diagnosis Based on Wide Deep Convolutional Neural Network and Long Short Term Memory

Zijian CHEN, Ji ZHAO*

Abstract: Mechanical fault can cause economic loss of different degrees, even casualties. Timely fault diagnosis is an essential condition for ensuring safe production in modern industry. With the growth of intelligent manufacturing, more and more attention is paid to fault diagnosis methods that are based on deep learning. However, the diagnostic accuracy of existing diagnostic methods has still to be improved. Therefore, a fault diagnosis method called WDCNN-LSTM is proposed by combining Wide First-layer Deep Convolutional Neural Network with Long and Short Term Memory. Feature information is extracted adaptively from one-dimensional original vibration signals by Convolutional Neural Network. The extracted features are further extracted by Long and Short Term Memory, so that the fault feature information can be fully obtained. Experiments are performed on CWRU datasets to verify our proposed method. By analyzing the experimental results, we find that the average accuracy of the proposed WDCNN-LSTM model is 99.65%.

Keywords: bearing fault diagnosis; deep learning; mechanical equipment; vibration signal

1 INTRODUCTION

Rotating machinery and equipment often operate in complex environments with high temperature, humidity, and heavy load. Their parts are easily damaged and cause equipment failure, which will cause serious economic losses and even loss of life [1, 2]. Therefore, an effective failure diagnosis of key components of rotating machines and equipment is an important condition for ensuring safe production in modern industry. The rolling bearing, the most widely used part of rotating machinery, is one of the most vulnerable. Bearing fault is the most common type through the mechanical equipment failure survey [3, 4], accounting for 30%~40% of the total number of failures. How to accurately diagnose the abnormal state of rolling bearings and accurately locate the fault location has attracted widespread attention from researchers and engineers.

In the early days, based on the vibration signal processing method there is the mainstream bearing fault diagnosis method, which is generally split into two parts: extraction characteristic and classification [5, 6]. The original vibration signal is picked up by the sensors located on the bearing equipment, which contains information about bearing state as well as a large amount of useless noise. Based on the signal processing methods, the aim is to extract information about the fault characteristics of the original signal using advanced signal denoising and filtering techniques [7-9]. The extracted features cannot all be used for diagnostics. It will not only increase the computational burden, but also reduce the accuracy of fault classification. Therefore, usually after feature extraction, useless features are removed by feature selection, and a suitable feature training classifier is selected to improve the computational efficiency.

In recent years, the equipment in the industrial field has gradually become large-scale, intelligent, and complex, and the amount of mechanical fault data and fault factors that can be collected has greatly increased. However, most of these traditional fault diagnosis methods use manual extraction of bearing vibration signals. Bearing fault diagnosis cannot meet the requirements of social development due to its long time, high cost and strong randomness of results. At present, deep learning-based methods [10-13] have become popular in the field of bearing fault diagnosis, which can adaptively extract representative features and overcome the limitations of manual feature extraction.

A large number of papers attempt to use convolutional neural network (CNN) [14] in deep learning technology for mechanical fault diagnosis. Although these approaches can comfortably extract fault characteristics and diagnose faults, the error rate of diagnosis is high, which cannot achieve satisfactory results. There are three main reasons for the above situation:

1) The error or unrelated features in the original signal are extracted;

2) CNN can only learn spacial information in the receiving field, but not time series information in vibration signal;

3) As the network structure deepens, the features that are learned by the CNN become more and more abstract. And it is easy to over-fitting, which will affect the test results.

To tackle the above issues, this paper suggests a novel fault diagnosis model WDCNN-LSTM. It cannot only extract fault feature information effectively, but also prevent over-fitting.

The rest of this article is in the following order. The work that is relevant to this article is talked about in section 2. Section 3 introduces the components of WDCNN-LSTM model in detail. Experimental Settings are described in detail in Section 4, such as fault diagnosis data sets, baseline models, implementation details, and evaluation metrics. In Section 5, the experimental results of the model on the CWRU dataset are shown and compared with some innovative baseline models. Finally, Section 6 summarizes the paper and makes recommendations for future research.

2 RELATED WORK

Deep learning technology has developed rapidly, and the trend of bearing fault detection has shifted from traditional methods to methods based on deep learning. Since the original signal collected from the accelerator is

one-dimensional data, CNN is widely used by many researchers in bearing fault diagnosis. Ince et al. [15] propounded a one-dimensional CNN that was directly applicable to the original vibration signal for real-time motor fault diagnosis. Guo et al. [16] proposed a hierarchical adaptive deep convolutional neural network (ADCNN) based on the conventional DCNN model. Zhang et al. designed a deep convolutional neural network with a wide first-layer kernel (WDCNN) [17] to repress high-frequency noise. In this method, the wide kernel of the first convolution layer can repress the high frequency noise while abstracting the features from the collected original vibration signals. Experimental results show that high accuracy can be achieved even in a noisy environment. Zhuang et al. [18] proposed a one-dimensional multi-scale deep convolutional neural network (MS-DCNN), which is directly used to process the original vibration signal. First, different sizes of convolution kernels are used to extract different sizes of features in parallel. Then, the neurons after the multi-scale convolution layer are connected and input to the output layer for classification.

In some cases, one-dimensional data are convertible into two dimensional representations by preprocessing. Chong et al. [19] put an effective method forward to extract vibration signal features in a two-dimensional domain. Firstly, the amplitude of each vibration signal sample is normalized to obtain a two-dimensional gray image. Finally, a scale invariant feature transform algorithm (SIFT) is used to extract local features. Kang and Kim [20] proposed a two-dimensional gray image representation method based on Shannon Wavelet, and used SVM to identify faults of induction machines to better adapt to the requirements of the multi-layer greedy learning based diagnosis model. Lu et al. [21] mapped the original signal to a series of feature maps by using the image transformation method. Wen et al. [22] proposed a novel signal-image conversion approach without predefined parameters, and suggested an improved fault diagnosis model based on LeNet-5.

RNN is suitable for time series data has been concerned by scholars in the domain of fault detection. In 2018, Tang et al. [23] proposed a bearing fault identification method based on a long short-term memory network (LSTM). Yu et al. [24] proposed a hierarchical algorithm based on stacked LSTM, including three layers of LSTM, and the dropout strategy [25] was adopted after each LSTM layer to avoid overfitting. The experimental results show the great potential of RNN. Pan et al. [26] focused on both CNN and LSTM, and proposed a modified bearing fault diagnosis method based on CNN and LSTM. Subsequently, in Khorrama et al. [27] the original time-domain feature is reconstructed in the shape of time-series, and the Convolutional Long-Short-Term-Memory Recurrent Neural Network (CRNN) is proposed. In 2020, Chen et al. [28] proposed a multi-scale CNN and LSTM combined model (MCNN-LSTM). Signal features of different frequencies in the original data are extracted automatically by two CNN with different kernel sizes. Then, LSTM are used to identify the type of fault. You et al. [29] proposed CNN-BLSTM in 2021, the original one-dimensional vibration signals collected were input into the model, and the feature information was extracted by CNN adaptively, and Bi-LSTM was used to fuse the extracted features. It fully obtains fault information and prevents the overfitting of the model.

3 MODEL

3.1 Convolutional Neural Network

Convolutional neural networks (CNN), namely feedforward neural networks with convolutional computation, were first employed in computer vision and other fields. CNN consists of three parts. The first part is the convolution layer, which extracts features through the convolution kernel. In this paper, one-dimensional convolution method is applied to time series vibration signal. The formula for one-dimensional convolution is as follows:

$$y_m^{L+1}(i) = W_m^L * x^L(i) + b_m^L, \tag{1}$$

where W_m^L and b_m^L indicate the weights and deviations of the m-th convolutional kernel in the L-th layer, respectively. $x^L(i)$ represents the i-th local input in L-th layer.

The second part is the activation function layer, which is the key because it is convenient for the network to obtain the nonlinear characteristics of the input signal. In recent years, the usually adopted activation functions are Sigmoid, Tanh, Relu, etc. When the parameters are adjusted by back propagation, the Relu function makes the shallow weights more trainable. Therefore, this paper uses the Relu activation function, its equation is formed as follows:

$$a_m^{L+1}(i) = f\left(y_m^{L+1}(i)\right) = \max\left(0, y_m^{L+1}(i)\right), \tag{2}$$

where $y_m^{L+1}(i)$ is the output after convolution operation, $a_m^{L+1}(i)$ is the activation value of $y_m^{L+1}(i)$.

The pooling layer further selects and filters the features obtained by convolution operation. In this paper, the pooling layer adopts the maximum pooling operation. The results of the m-th channel in the (L + 1)-th layer after pooling can be defined as:

$$P_m^{L+1}(j) = \max_{(j-1)w+1 \le t \le jw} \left\{q_m^L(t)\right\}, \tag{3}$$

where $q_m^L(t)$ represents the t-th neurons of the m-th channel in the L-th layer , w is the width of the pooling area.

3.2 Batch Normalization

Batch normalization layer (BN) [30] is devised to diminish the deviation of internal covariance [31]. It cannot only speed up the convergence of model, but also alleviate the "gradient dispersion" caused by the deep network structure to a certain extent. Therefore, the BN layer is generally added after the convolution layer [32]. Given q-dimensional input of a BN layer $y^{L(i)} = \left(y^{L(i,1)}, ..., y^{L(i,q)}\right)$. After the BN layer transformation, $z^{L(i,j)}$ is obtained, the specific calculation process is as follows:

$$\hat{y}^{L(i,j)} = \frac{y^{L(i,j)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \tag{4}$$

$$z^{L(i,j)} = \gamma \hat{y}^{L(i,j)} + \beta, \tag{5}$$

where $\mu = E\left[y^{L(i,j)} \right]$ stands for the mean value of the output data at the L-th layer and $\sigma^2 = Var\left[y^{L(i,j)} \right]$ expresses the standard deviation of the output data at the L-th layer. ε is a constant, which is added to prevent the denominator being 0. γ and β are learning parameters.

3.3 Long Short Term Memory

Fig. 1 illustrates the unit structure of the LSTM. It consists of a memory cell and three gated units, which are forgetting gate, input gate, and output gate. Specifically, the memory cell controls the data to be remembered through these three gates. The forgetting gate selectively forgets the useless information. The input gate filters the input information and updates the memory status. The output gate plays a decisive role; it outputs information according to the memory cell. The specific calculation process is revealed in Eq. (6) ~ Eq. (11):

$$f_t = \sigma(W_f \left[h_{t-1}; x_t \right] + b_f), \tag{6}$$

$$i_t = \sigma(W_i \left[h_{t-1}; x_t \right] + b_i), \tag{7}$$

$$o_t = \sigma(W_o \left[h_{t-1}; x_t \right] + b_o), \tag{8}$$

$$\tilde{c}_t = tanh\left(W_c \left[h_{t-1}; x_t \right] + b_c \right), \tag{9}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{10}$$

$$h_t = o_t \odot \tanh\left(c_t\right), \tag{11}$$

where σ and tanh represent sigmoid and tanh activation functions, respectively. f_t is the forgetting gate, W_f are the weights of the forgetting gate and b_f are the bias vector of the forgetting gate. i_t is the input gate, W_i are the weights of the input gate and b_i are the bias vector of the input gate. o_t is the output gate, W_o are the weights of the forgetting gate and b_o are the bias vector of the forgetting gate. c_t is the memory cell, and h_t represents the hidden layer state. W_c and b_c are the weights of the memory cell, the bias vector of the memory cell, respectively. [;] represents the splicing operation of two vectors. $\odot$ represents the multiplication operation of two vectors.

3.4 Proposed WDCNN-LSTM Model

This paper presents a model combining a Deep Convolutional Neural Networks with Wide First-layer Kernels [17] and LSTM (WDCNN_LSTM). Specifically, the WDCNN-LSTM consists of the WDCNN module, the LSTM module, and a fully connected layer. The WDCNN module has a 15-layer network structure, which consists of five alternating convolutional layers, batch normalization layers and pooling layers. The LSTM module consists of two LSTM layers. The parameters of the WDCNN-LSTM architecture are as shown in Tab. 1.

The original signal is fed directly into the first convolution layer without any other transformation, and the convolution layer extracts features from the input. The first convolution layer in the model adopts the wide convolution kernel because the wide convolution kernel can suppress the high frequency noise better than the small one. To decrease the deviation of internal covariance, batch normalization is performed after each convolutional layer and before the pooling layer [28]. Then the features obtained after convolutional pooling are input into the LSTM, which extracts the temporal features. The processed features are input into the full connection layer, and the classification results are output.

Table 1 The parameters of the WDCNN-LSTM

Layer Name	Filters	Kernel Size	Stride	Trainable parameters
Convolution1	16	64 × 1	1 × 1	1040
BatchNormalization1				64
Pooling1	16	2 × 1	2 × 1	0
Convolution2	32	3 × 1	1 × 1	1568
BatchNormalization2				128
Pooling2	32	2 × 1	2 × 1	0
Convolution3	64	3 × 1	1 × 1	6208
BatchNormalization3				256
Pooling3	64	2 × 1	2 × 1	0
Convolution4	64	3 × 1	1 × 1	12352
BatchNormalization4				256
Pooling4	64	2 × 1	2 × 1	0
Convolution5	64	3 × 1	1 × 1	12352
BatchNormalization5				256
Pooling5	64	2 × 1	2 × 1	0
LSTM1				12416
LSTM2				8320
Dense				330

4 EXPERIMENTAL STUDY

4.1 Dataset

This paper uses the rolling bearing fault CWRU dataset that is offered by [24] to evaluate the propounded method. The rolling bearing fault simulation experimental platform is exhibited in Fig. 1. The experimenter used electric spark to manually create faults on deep groove ball bearings (6205-2RSJEMSKF), the fault data are collected by an acceleration sensor mounted above the bearing seat of the motor's fan/drive end.

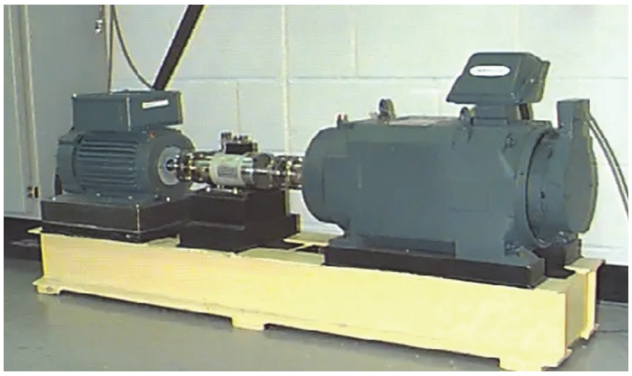

Figure 1 Rolling bearing experiment platform

In this paper, only the driving end bearing fault data sampled at a frequency of 12 KHZ is used, including three fault states: inner race fault, outer race fault and ball fault. The outer race fault is located at 3 o'clock, 6 o'clock and 12 o'clock respectively. Each fault has three distinct damage diameters of 0.007 inch, 0.014 inch and 021 inch (1 inch = 25.4 mm) respectively. In this paper, only the outer ring fault at 6 o'clock position is used. Therefore, there are 10 fault types, so it is a 10 classification problem.

We plotted the time-domain waveform of vibration acceleration signals, as shown in Fig. 2. It is observable from the picture that each bearing the characteristics of different bearing states are different, it is hard to recognize the fault type directly from the time domain waveform by the eye.

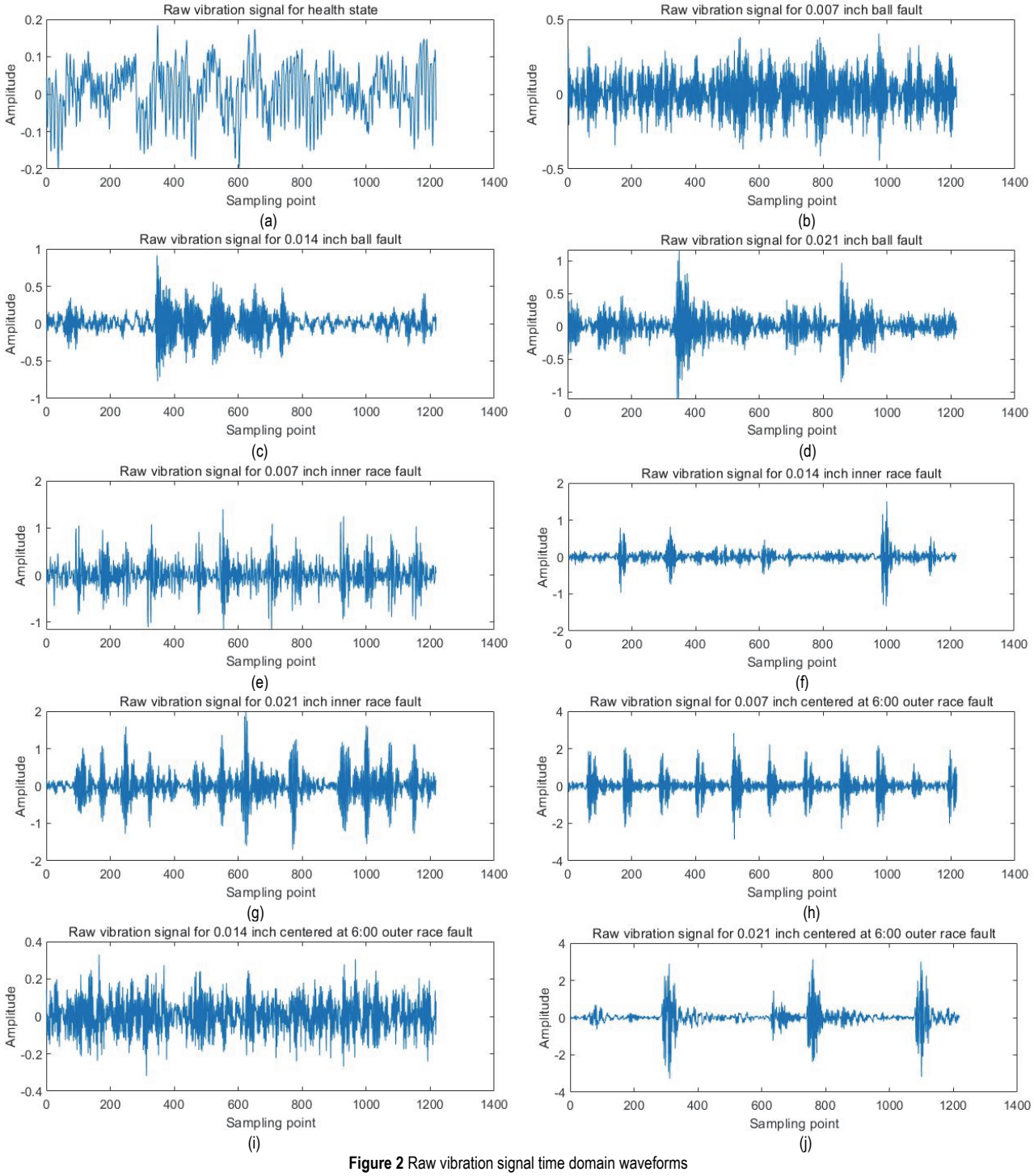

Figure 2 Raw vibration signal time domain waveforms

The number of sample points in each cycle can be inferred by acquisition frequency and speed (sample points/each cycle = sampling frequency × 60/speed). This experiment mainly studies the bearing fault situation under 1hp load, and the motor speed is 1797/minute, so the number of sample points in each cycle is 406.32 (12000 × 60/1772 = 406.32). Therefore, the sample points length was set to 406. For the training samples, the raw data were intercepted using overlapping sampling, with a sliding stride of 80, as shown in Fig. 3.

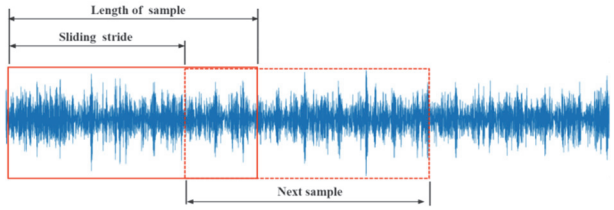

Figure 3 Overlapping sampling operation

For the test samples, a no-overlap intercept will be used. We collected 10 classes of fault data at 1hp to construct the dataset for experiments, each class containing 660 training samples and 100 test samples. Tab. 2 shows the basic information of the dataset.

Table 2 Description of rolling bearing datasets

Location	Diameter / inch	Labels	Train	Test
None	0	0	660	100
Ball	0.007	1	660	100
	0.014	2	660	100
	0.021	3	660	100
Inner race	0.007	4	660	100
	0.014	5	660	100
	0.021	6	660	100
Outer race	0.007	7	660	100
	0.014	8	660	100
	0.021	9	660	100

4.2 Baseline

To confirm the performance of the proposed model, we compare it with the following innovative model:

1) WDCNN: Zhang et al. [17] proposed Convolutional Neural Networks with Wide First-Layer Kernels model.

2) MS-DCNN: Zhuang et al. [18] proposed one-dimensional multi-scale deep convolutional neural network.

3) MCNN-LSTM: Chen et al. [28] proposed Multi-scale Convolutional Neural Network and Long short-term Memory in 2020.

4) CNN-BLSTM: You et al. [29] proposed CNN-BLSTM in 2021.

To reduce the impact of randomness on classification accuracy, 20 independent experiments were run for each baseline model. The models were written using tensor flow = 1.14, and all experiments were run on GTX3060Ti machine. The learning rate is set to 0.001 and the batch size and epochs are set to 128, 300 respectively. The cross entropy loss function is used to calculate the loss. To prevent over-fitting, we used the Early Stopping method throughout the training process. When the loss value of the verification set did not decrease for 10 consecutive times, the training was ended.

4.3 Evaluating Metric

We use accuracy to assess the performance of the fault diagnostic model, and the formula is defined as follows [33]:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}, \tag{12}$$

TP (True Positives) indicates the number of cases classified as positive correctly; *FP* (False Positives) shows the number of cases classified as positive wrongly; *FN* (False Negatives) refers to the number of cases wrongly divided into negative cases; The number of samples correctly classified as negative cases is represented by *TN* (True Negatives).

5 EXPERIMENTAL RESULTS

5.1 Case Study I: Effect of the Number of Hidden Layers of the LSTM and the Number of Hidden Units

The more complex the network structure, the more training parameters. The error of such network structure will become smaller and smaller during training, but there may be "overfitting". In the following experiments, we explore the influence of the number of hidden layers and the number of hidden units in LSTM on the model performance. We train the two network structures, namely stacking one layer or two layers LSTM after WDCNN, and the experimental results are corresponding to Tab. 3 and Tab. 4. WDCNN-LSTM-*i*-*j* indicates that there are *i* hidden units in the first layer LSTM and *j* hidden units in the second layer LSTM. For example, WDCNN-LSTM-32 indicates that there are 32 units in the first layer LSTM and there is no the second layer LSTM. All models were run 20 times, and the average accuracy and standard deviation of the test set were the result measures. By analyzing Tab. 3, we find that the average accuracy increases first and then decreases as the number of hidden layer units increases. The number of hidden units is set to 64, the standard deviation is the smallest. In practice, the standard deviation is often understood as stability, and the smaller the standard deviation, the more stable it is. Therefore, The WDCNN-LSTM-64 is the best and has an average accuracy of 99.54%.

Table 3 Results of the WDCNN-LSTM model with one layer of LSTM

Model	Average accuracy/ %	Standard deviation	Execution time / s
WDCNN-LSTM-32	99.38	0.40	62.4658
WDCNN-LSTM-64	**99.54**	**0.21**	**64.1456**
WDCNN-LSTM-128	99.43	0.26	73.9732

To further analyze the influence of LSTM using the different number of hidden units on the classification results, we use T-SNE to visualize test set samples under $i \in \{32,64,128\}$. Bearing fault features of the same fault type will appear obvious aggregation phenomenon. Fig. 4a, Fig. 4b, and Fig. 4c correspond to the feature visualization results of 32, 64, and 128 hidden units in LSTM respectively. It is obvious from Fig. 4 that there are obvious boundaries between the different fault types. Nevertheless, the number of hidden units is set to 32, there is an obvious clustering between label 1 and label 3, label 5 and label 9; the number of hidden units is set to 64, only label 1 and label 3 cluster significantly. While the number of hidden units is 128, label 3 is obviously clustered with label 1 and label 2. Clustering among features of different fault types affects the classification accuracy of the classifier. In contrast, Feature distribution learned automatically from LSTM with 64 hidden units stacked after WDCNN is more separable.

The WDCNN-LSTM model with two-layer LSTM is based on the WDCNN-LSTM-64 experiment, and the number of hidden units of the second-layer LSTM is set to 32, 64, and 128 respectively. The experimental results are demonstrated in Tab. 4. Through the analysis of Tab. 4, we find that the average accuracy increases with the number of layers. This also proves that the number of layers and hidden units has a great influence on the accuracy of the model [34], and an appropriate number of layers and hidden units is conducive to a more accurate and stable classification of the model. When the number of hidden units at the second layer LSTM is 64, the average accuracy of classification is higher than hidden units 32 and 128, and higher than that of the WDCNN-LSTM model with only one layer LSTM.

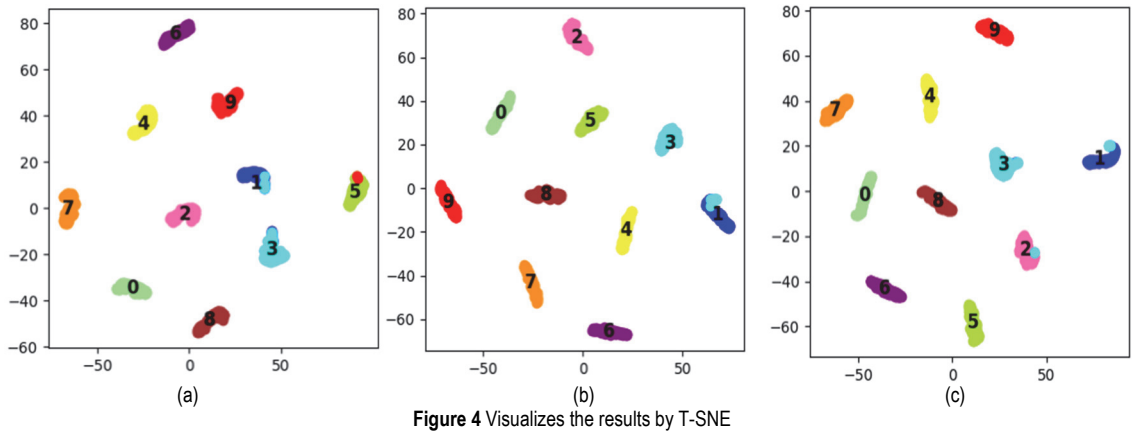

Figure 4 Visualizes the results by T-SNE

Table 4 Results of the WDCNN-LSTM model with two layers of LSTM

Model	Average accuracy / %	Standard deviation	Execution time / s
WDCNN-LSTM-64-32	99.59	0.30	71.1641
WDCNN-LSTM-64-64	**99.65**	**0.35**	**73.7502**
WDCNN-LSTM-64-128	99.59	0.24	79.2877

In the following experiment, the WDCNN-LSTM model of stacked two-layer LSTM was used for training, and the number of hidden units was set to 64.

5.2 Case Study II: Compared With Other Methods

We use the same experimental data to compare the WDCNN-LSTM model with the current advanced intelligent diagnosis model. The experimental results are displayed in Tab. 5.

The proposed WDCNN-LSTM model has an average accuracy of 99.65% under 1 hp, which is higher than all baseline fault diagnosis methods.

To facilitate the intuitive analysis of the model performance, we drew the test set accuracy line chart for each test of all models, as shown in Fig. 5. It can be seen from Fig. 5 that the accuracy curve of WDCNN-LSTM model has a gentle trend and is located at the top of the figure. It indicates that the test set accuracy for each trial is relatively high. The accuracy of CNN-BLSTM model is slightly higher than that of WDCNN-LSTM by 0.3%~0.4% in the 17th and 18th trials, but it is inferior to our proposed model on the whole. The accuracy curve trend of MCNN-LSTM model is fluctuant and the stability is poor, which is the same as the situation reflected by the standard deviation. The accuracy curve of WDCNN model is at the bottom of the figure and the accuracy is the lowest.

Table 5 Experimental results of the different models

Model	Average accuracy / %	Standard deviation	Execution time / s
WDCNN	97.28	0.57	40.3998
MS-DCNN	98.38	0.24	37.8592
MCNN-LSTM	98.42	0.98	91.0105
CNN-BLSTM	98.80	0.49	130.0010
WDCNN-LSTM	99.65	0.35	73.7502

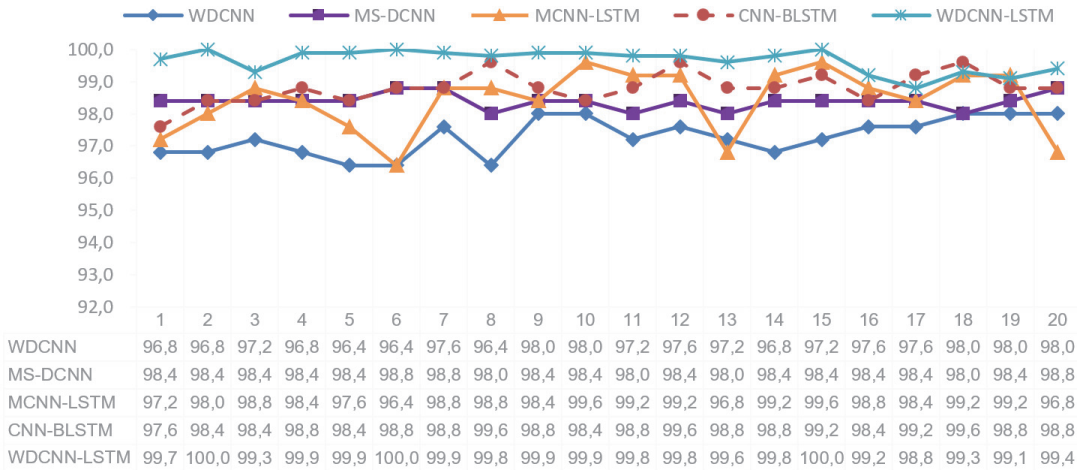

Figure 5 Test set accuracy results of different models

We selected the results of the 5th trial of WDCNN model, the 8th trial of MS-DCNN model, the 6th trial of MCNN-LSTM model, the 1st trial of CNN-BLSTM model and the 17th trial of WDCNN-LSTM to draw the confusion matrix (the number of trials is selected based on the one with the lowest accuracy). Corresponding to (a),(b),(c),(d) and (e) in Fig. 6 respectively. The confusion matrix can intuitively observe the number of true and false for each category of the model, where the true label is represented by the row and the column represents the predicted label. Through the analysis of confusion matrix, it was found that the main reason for the low overall accuracy of WDCNN model is the low classification accuracy of labels 1, 2, and 3. And 8% of label 1 was misclassified as label 3. The MS-DCNN model achieves lower accuracy in label 9, 2, 3, and 5. The reason for the low accuracy of MCNN-LSTM model is the high misclassification of label 5. The classification accuracy of label 2 and 3 in CNN-BLSTM model is low. 12% of label 2 are misclassified as label 0 and 4% are misclassified as label 3. In conclusion, the reason for the lower overall low accuracy of the above models is due to the lower classification rate of the model on labels 1, 2, 3, and 5.

According to Fig. 6e, the error classification of WDCNN-LSTM model proposed in this paper also mainly focuses on label 1, 2, and 3, but the error rate is relatively low. Label 3 was misclassified as label 1 and 9 accounted for 5% and 2%, respectively. Only 1% of label 2 was misclassified as label 3. The similarity between fault signal features leads to misclassification, such as labels 1, 2, 3, which are all fault types with different fault diameters on the ball. By analyzing each category of the models and the overall fault diagnosis results, it is verified that WDCNN-LSTM model can effectively detect bearing faults.

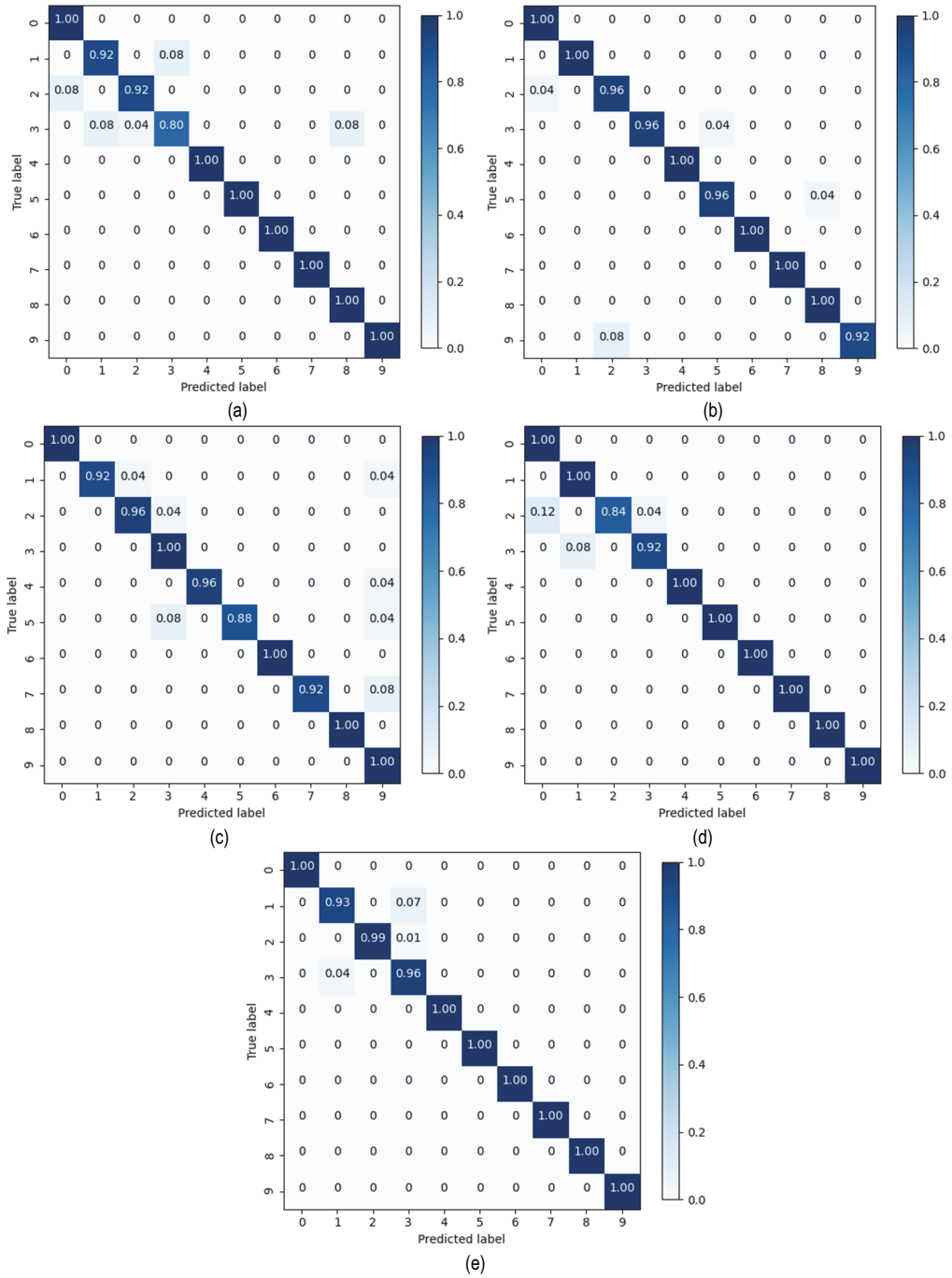

Figure 6 The confusion matrix for each model

6 CONCLUSION

To enhance the stability and diagnostic accuracy and stability of fault diagnosis approach, we propose an intelligent diagnosis model WDCNN-LSTM for rolling bearing faults, which consists of WDCNN module, LSTM module, and a full connection layer. The original vibration signals collected were directly input into the model, and feature information was extracted adaptively through convolution, batch normalization layer, and pooling layer further extracted by stacked LSTM to fully obtain fault feature information. The processed features are input into the full connection layer, and the classification results are output by softmax classifier. We use the CWRU bearing data set to verify the validity of fault model diagnosis WDCNN-LSTM. Experimental results show that the highest accuracy and average accuracy of WDCNN-LSTM model are 100% and 99.65%, respectively, which are higher than other existing fault diagnosis methods. It also proves that the proposed WDCNN-LSTM model can extract features with high discrimination from the original input signal data and is an effective fault diagnosis model. Although the overall accuracy of WDCNN-LSTM model is high, the classification rate of WDCNN-LSTM model for ball fault type needs to be improved. Therefore, in future work, studies can be conducted on the ball fault types of deep groove ball bearings to enhance the accuracy of ball faults, and further improve the overall accuracy of the model.

Acknowledgements

This work was supported by the Natural Science Foundation of Liao Ning Province (20180551048), China.

7 REFERENCES

[1] El Hachemi Benbouzid, M. (2000). A review of induction motors signature analysis as a medium for faults detection. *IEEE Transactions on Industrial Electronics*, *47*(5), 984-993. https://doi.org/10.1109/41.873206

[2] Zhang, P., Du, Y., Habetler, T. G., & Bin, L. (2009). A survey of condition monitoring and protection methods for medium voltage induction motors. 2009 *IEEE Energy Conversion Congress and Exposition*. https://doi.org/10.1109/ecce.2009.5316083

[3] I. C. (1987). Report of large motor reliability survey of industrial and commercial installations: Part 3. *IEEE Transactions on Industry Applications*, *IA-23*(1), 153-158. https://doi.org/10.1109/tia.1987.4504880

[4] JEMA (2000). On recommended interval of updating IMs.

[5] Prieto, M. D., Cirrincione, G., Espinosa, A. G., Ortega, J. A., & Henao, H. (2013). Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and Neural Networks. *IEEE Transactions on Industrial Electronics*, *60*(8), 3398-3407. https://doi.org/10.1109/tie.2012.2219838

[6] Li, K., Chen, P., & Wang, S. (2012). An intelligent diagnosis method for rotating machinery using least squares mapping and a fuzzy neural network. *Sensors*, *12*(5), 5919-5939. https://doi.org/10.3390/s120505919

[7] Wang, X., Zheng, Y., Zhao, Z., & Wang, J. (2015). Bearing fault diagnosis based on statistical locally linear embedding. *Sensors*, *15*(7), 16225-16247. https://doi.org/10.3390/s150716225

[8] Rai, V. & Mohanty, A. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang Transform. *Mechanical Systems and Signal Processing*, *21*(6), 2607-2615. https://doi.org/10.1016/j.ymssp.2006.12.004

[9] Lee, W. & Park, C. (2014). Double fault detection of cone-shaped redundant Imus using wavelet transformation and EPSA. *Sensors*, *14*(2), 3428-3444. https://doi.org/10.3390/s140203428

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

[11] Liu, H., Li, L., & Ma, J. (2016). Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock and Vibration,* 2016, 1-12. https://doi.org/10.1155/2016/6127479

[12] Lu, C., Wang, Z., Qin, W., & Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising auto encoder-based health state identification. *Signal Processing*, 130, 377-388. https://doi.org/10.1016/j.sigpro.2016.07.028

[13] Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards Mechanical Big Data. *IEEE Transactions on Industrial Electronics*, *63*(5), 3137-3147. https://doi.org/10.1109/tie.2016.2519325

[14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84-90. https://doi.org/10.1145/3065386

[15] Ince, T., Kiranyaz S., Eren L., Askar, M., & Gabbouj M. (2016). Real-Time Motor Fault Detection by 1D Convolutional Neural Networks. *IEEE Transactions on Industrial Electronics*, *63*(11). https://doi.org/10.1109/TIE.2016.2582729

[16] Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to Bearing Fault diagnosis. *Measurement*, *93*, 490-502. https://doi:10.1016/j.measurement.2016.07.054

[17] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, *17*(2), 425. https://doi.org/10.3390/s17020425

[18] Zilong, Z. & Wei, Q. (2018). Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based Health State Classification. *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. https://doi.org/10.1109/icnsc.2018.8361296

[19] Do, V. T. & Chong, U. (2011). Signal Model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain. *Strojniški Vestnik - Journal of Mechanical Engineering*, *57*(09), 655-666. https://doi.org/10.5545/sv-jme.2010.162

[20] Kang, M. & Kim, J. (2014). Reliable fault diagnosis of multiple induction motor defects using a 2-D representation of Shannon Wavelets. *IEEE Transactions on Magnetics*, *50*(10), 1-13. https://doi.org/10.1109/tmag.2014.2316474

[21] Lu, C., Wang, Z., & Zhou, B. (2017). Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based Health State Classification. *Advanced Engineering Informatics*, *32*, 139-151. https://doi.org/10.1016/j.aei.2017.02.005

[22] Wen, L., Li, X., Gao, L., & Zhang, Y. (2018). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, *65*(7), 5990-5998. https://doi.org/10.1109/tie.2017.2774777

[23] Tang, S., He, X., Zhang, J., & Yin, A. (2018). Bearing Fault Identification Based on Long Short-term Memory Networks. *Chinese Journal of Automotive Engineering*.
[24] Yu, L., Qu, J., Gao, F., & Tian, Y. (2019). A novel hierarchical algorithm for Bearing Fault diagnosis based on stacked LSTM. *Shock and Vibration*, *2019*, 1-10. https://doi.org/10.1155/2019/2756284
[25] Srivastava, N., Hinton, G., & Krizhevsky, A. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*(1), 1929-1958.
[26] Pan, H., He, X., & Tang, S. (2018) An improved bearing fault diagnosis method using one-dimensional CNN and LSTM. S*trojniški Vestnik - Journal of Mechanical Engineering*, *64*(7-8), 443-452. https://doi.org/10.5545/sv-jme.2018.5249
[27] Khorram, A., Khalooei, M., & Rezghi, M. (2020). End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis. *Applied Intelligence*, *51*(2), 736-751. https://doi.org/10.1007/s10489-020-01859-1
[28] Chen, X., Zhang, B., & Gao, D. (2020). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, *32*(4), 971-987. https://doi.org/10.1007/s10845-020-01600-2
[29] Loffe, S. & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv*. Case Western Reserve University Bearing Data Center.
[30] Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, *100*, 439-453. https://doi.org/10.1016/j.ymssp.2017.06.022
[31] Plakias, S. & Boutalis, Y. S. (2020). Fault detection and identification of rolling element bearings with attentive dense CNN. *Neurocomputing*, *405*, 208-217. https://doi.org/10.1016/j.neucom.2020.04.143
[32] Zeng, Y., Gu, H., Wei, W., & Guo, Y. (2019). Deep-full-range: A deep learning based network encrypted traffic classification and Intrusion Detection Framework. *IEEE Access*, *7*, 45182-45190. https://doi.org/10.1109/access.2019.2908225
[33] Wu, D., Wu, Q., Yin, X., Jiang, B., Wang, H., He, D., & Song, H. (2020). Lameness detection of dairy cows based on the yolov3 deep learning algorithm and a relative step size characteristic vector. *Biosystems Engineering*, *189*, 150-163. https://doi.org/10.1016/j.biosystemseng.2019.11.017

Contact information:

Zijian CHEN
University of Science and Technology Liaoning,
School of Computer Science and Software Engineering,
No. 189, Qianshan Road, Anhsan, Liaoning, China
E-mail: 13998777277@163.com

Ji ZHAO
(Corresponding author)
University of Science and Technology Liaoning,
School of Computer Science and Software Engineering,
No. 189, Qianshan Road, Anhsan, Liaoning, China
E-mail: zhaoji_1974@126.com