Research on Fault Detection of Rolling Bearing Based on CWT-DCCNN-LSTM

Yu Wang, Changfeng Zhu, Qingrong Wang, and Jinhao Fang

Abstract—**As one of the key components in many fields, rolling bearing fault detection is very important. Rolling bearing is in complex and changeable working conditions, so it is challenging to detect its fault. Because the traditional method has weak adaptability in complex and changeable situations, it needs to rely on the opinions of experts more often. Deep learning methods can make up for the shortcomings of traditional methods. Therefore, this paper proposes a method combining continuous wavelet transform (CWT), dual-channel convolutional neural network(DCCNN), and long short-term memory network (LSTM), mainly for fault detection of vibration signals of rolling bearings. Firstly, the vibration signal is denoised by CWT, then the feature of the vibration signal is extracted by DCCNN, and finally, the time series of the vibration signal is extracted by LSTM. Compared with CNN, CWT-CNN, CNN-LSTM, and CWT-CNN-LSTM four models, and analyzed the parameters of the model. The results show that the accuracy of CWT-DCCNN-LSTM model detection is better than other models, and the accuracy rate reaches 99.98 %.**

Index Terms—**Deep Learning, Rolling bearings, Continuous Wavelet Transform(CWT), Dual-Channel Convolutional Neural Network(DCCNN), Long Short-Term Memory Network(LSTM)**

I. INTRODUCTION

Rolling bearing plays a key role in the field of aviation and other transportation, workshop large equipment, and precision instruments. However, due to the impact of the relevant conditions, rolling bearings become one of the most vulnerable mechanical components, and early weak fault detection and fault condition monitoring is particularly important. The real-time monitoring of rolling bearing faults is mainly based on the analysis of vibration signals. Because rolling bearings often work in harsh environments such as high speed, high temperature, and high pressure, the obtained vibration signals often contain interference problems such as artificial noise, which seriously affects the accuracy and timeliness of detection. Therefore, how to effectively obtain the fault frequency characteristics and monitor the fault situation in real time is of great significance for improving the detection accuracy, reducing the maintenance cost of rolling bearings, and prolonging their service life.

Early, [1] proposed a rolling bearing fault detection theory based on the vibration signal detection method, which laid the foundation for this problem. After that, [2][3] proposed a stochastic resonance denoising method, where, [2] is based on the stochastic resonance method of the coupled bistable system to deal with the noise of the vibration signal, mainly using its noise to enhance the characteristics of the vibration signal, [3] proposed a single well model, which combines the piecewise function with the classical bistable stochastic resonance method. The model can detect the vibration signal of the rolling bearing defect. [4][5][6][7][8][9], the entropy function is used to deal with the noise of rolling bearing vibration signal, whereas, in [4] proposes a model of cyclic entropy function to deal with the vibration signal with noise, in [5] proposed an improved multi-scale fuzzy entropy model, which mainly solves the time series problem of rolling bearings, in [6][7][8] used the method of multi-scale discrete entropy, where, in [6][8], multi-scale normalized discrete entropy is used to solve the time series problem of nonlinear rolling bearing, in [7], fault identification of weak signals of rolling bearings by combining cuckoo search algorithm with multi-scale discrete entropy. [10][11][12] used the Empirical Mode Decomposition (EMD) method, where, in [10] combined matching pursuit with EWD to filter the original signal to generate components, in [11] proposed an ensemble empirical mode decomposition to reduce its error, in [12] combined EMD with quantile permutation entropy, and calculated by quantile permutation entropy algorithm.

However, the above method cannot simultaneously obtain the time frequency of the vibration signal. Fourier Transforms (FT) [13] laid the foundation for obtaining the time frequency at the same time, and then Morlet [14] proposed Wavelet Transformation (WT) on this basis, which further improved FT and overcame the problem that the window size did not change with the frequency transformation. [15] considered the Q factor contained in the vibration signal while using WT denoising, and combined with the characteristic scale decomposition, the model is better for weak signal extraction. After that, Wavelet-Packet transform (WPT) was proposed in [16][17], where, in [16] proposed a Dual-Tree Complex Wavelet Packet Transform (DTCWPT) to preprocess the vibration signal. The practicability and accuracy of this method are good, in [17] used three layers of WPT so that the model can improve the fault classification performance under sensitive analysis.

Manuscript received January 17, 2023; revised May 22, 2023.

This work was supported in part by the National Natural Science Foundation of China (No.72161024), "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No.GSSYLXM-04) and Gansu Provincial Science and Technology Plan Project (22ZY1QA005).

Yu Wang is a postgraduate student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 962078842@qq. com).

Changfeng Zhu is a Professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, phone: +86 18919891566, e-mail: cfzhu003@163. com).

Qingrong Wang is a Professor at School of Electronic and Information Engineering,Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail:wangqr003@163.com)

Jinhao Fang is a doctoral candidate at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: fangjin_hao@163. com).

[18][19] used method of Empirical Wavelet Transform (EWT), whereas, in [18] combines EWT with adaptive Kurtogram. The model optimizes the method of dividing the boundary in the frequency domain, in [19] improves the traditional EWT and combines the improved EWT with a Support Vector Machine (SVM). The improved EWT effectively separates the vibration signal from the noise. Wavelet Transform (WT) mainly includes Continuous wavelet Transformation (CWT) [20] and Discrete Wavelet Transformation (DWT) [21]. Because the vibration signal of the rolling bearing is a non-stationary signal, some scholars choose to use CWT to denoise the vibration signal. [22][23], CWT is used to detect the fault of rolling bearing, whereas, in [22] analyzes the bearing fault by scale and time wavelet spectrum. The two methods can detect the fault and identify the fault mode at the same time.

For the feature extraction of rolling bearing vibration signal, the traditional method has poor adaptability and low stability in complex and changeable situations, and more often needs to rely on expert opinions. Deep learning methods can make up for the shortcomings of traditional diagnostic methods, and Convolutional Neural Networks (CNN) can automatically extract features and is widely used in image classification and other fields. [24] proposed a deep learning method for convolutional neural networks. In 2006, [25] proposed to reduce the data dimension by the neural network, allowing deep autoencoder networks to learn low-dimensional code, as a tool to reduce dimension and provide a guarantee for principal component analysis. [26][27] used CNN to detect the faults of rolling bearings, whereas, in [27] improved the traditional LeNet-5 network to a two-dimensional LeNet-5 network, and the improved model improved the fault classification ability. [28] proposed a physics-based CNN model, which can simultaneously monitor the vibration signals of multiple rolling bearings and detect their faults. [29] converted the vibration signal into a spectrum image and then input it into CNN, the model has good robustness. [30] proposed a model of noise reduction of vibration signals by multi-kernel maximum mean difference and then combined it with an adaptive Deep Belief Network (DA-DBN) to solve the problem of the lack of a large number of labeled samples under new working conditions.

Due to the significant characteristics of CNN, some scholars have combined CNN with CWT to diagnose rolling bearings after CWT has processed the vibration signal noise. [31] combined CWT with CNN to detect sensor fault diagnosis of an aero-engine control system. [32] proposed a model combining WT-CNN, which introduces WT into the adversarial network to ensure that CNN can be translated unchanged, this model improves the quality of data generation and balances the data set. [33][34] is a model that combines CWT-CNN to detect its faults, whereas, in [33] combines two-dimensional CWT-CNN, extracts its features through multi-layer convolution and pooling, and detects gearbox faults in [34] not only combines CWT-CNN but also combines the Variational Mode Decomposition (VMD) method to effectively detect the fault of helicopter bearings. However, only combining CWT-CNN ignores the timing characteristics of rolling bearing failure, resulting in a certain one-sidedness.

Therefore, for the time series feature extraction of rolling bearing vibration signals, some scholars have proposed Long Short-Term Memory (LSTM), which can solve the time series problem, among them, [35] proposed a batch normalized LSTM model to reflect the mapping relationship of the data set to generate auxiliary samples, and then align the auxiliary samples with unlabeled data, which shows that the model is effective under a small amount of labeling. [36], a WT-EMD-EEMD-LSTM-based model is proposed to extract features from EMD-EEMD with LSTM as classifier to accomplish fault prediction. [37] implemented end-to-end detection using a model that combines CNN, LSTM, and Attention Mechanism (AM).

However, single-channel CNN will lose some information. To make up for this problem, some scholars have proposed multi-channel CNN. Among them, [38] proposed a Dual-Channel CNN (DCCNN) method for feature fusion, the first channel extracts the time domain and the second channel extracts the time-frequency domain, after, the fused features are extracted by CWT to detect rolling bearing faults. [39] proposed the Markov Transfer Field (MTF) method to convert the vibration signal into a two-dimensional image, and input two different data sets into the MultiDimensional Convolutional Neural Network (MDCNN) , which makes the model more robust. [40] proposed a model combining MCNN with Multi-Scale Cropping Fusion (MSCF), after, using MSCF to enhance the vibration signal, the image is then fused into MCNN, which reduces the complexity. However, in the fault detection of rolling bearings, some scholars only use CWT and CNN, without considering the characteristics of time series, some scholars use CNN and LSTM, but ignore the noise problem of vibration signal, which will produce the error of diagnosis results.

Based on this, this paper comprehensively considers the noise, fault characteristics, and time series characteristics of rolling bearing vibration signals, and introduces CWT theory, DCCNN theory, and LSTM theory. These three theories can reduce the noise of vibration signals and extract the fault characteristics and time series characteristics of vibration signals. Considering the three elements of rolling bearings, a model based on CWT-DCCNN-LSTM is finally constructed to detect and classify rolling bearing faults.

The rest of this article is introduced as follows: Section Ⅱ elaborates the theories of CWT, DCCNN, and LSTM respectively, and constructs the CWT-DCCNN-LSTM model. Section Ⅲ, taking the bearing data center of Case Western Reserve University as an example, the fault of the rolling bearing is detected based on the CWT-DCCNN-LSTM model and compared with other models to test its accuracy. Section Ⅳ gives the conclusion for the above discussion.

II. Construction of Rolling Bearing Fault Detection Model Based on CWT-DCCNN-LSTM

Rolling bearings work at high speed and long time under complex conditions such as tension, compression, and alternating, which leads to friction and damage of the inner ring, outer ring, and rolling element of rolling bearings. The peak spectrum of the acceleration signal in the time-frequency domain is mainly used to detect the fault. But the actual detection of the vibration signal will contain noise,

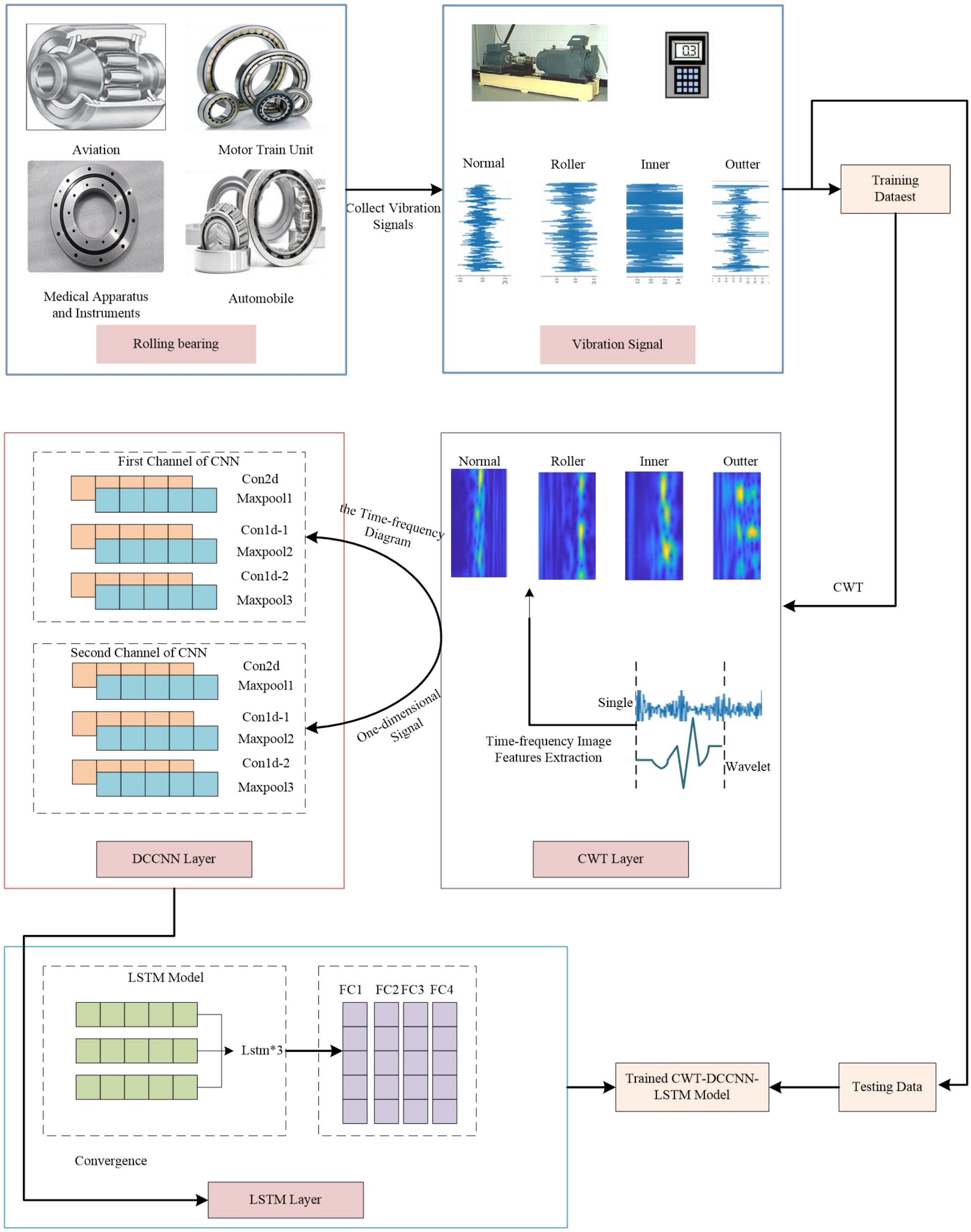


Fig. 1 Rolling bearing detection structure based on CWT-DCCNN-LSTM model.

which will lead to early some of the weaker features will be ignored, resulting in errors for bearing fault detection having a certain challenge.

Due to the error caused by the noise contained in the initially detected rolling bearing signal, CWT is used for noise reduction to obtain a time-frequency map. Then, the time-frequency diagram is input into DCCNN to extract the features of rolling bearings. Then the signal of the rolling bearing after dimension reduction is input into LSTM for further filtering, and then the signal is input into the fully connected layer, and finally, the fault diagnosis of the rolling bearing is completed. The rolling bearing detection structure based on the CWT-DCCNN-LSTM model is shown in Fig.1.

A. Data Noise Reduction Based on CWT

Firstly, the noise of the vibration signal of the rolling bearing should be considered. To reduce the error of detecting the fault of the rolling bearing, the vibration signal should be denoised first. CWT can divide the vibration signal into multiple time intervals to complete the initial vibration signal denoising. Because CWT can obtain both the time domain and the frequency domain, CWT is selected to denoise the vibration signal, that is, the signal is input into the CWT layer first. The noise reduction process of rolling bearing vibration signal based on CWT is shown in Fig.2.

To better simulate the waveform of a rolling bearing vibration signal, considering that the waveform generated by

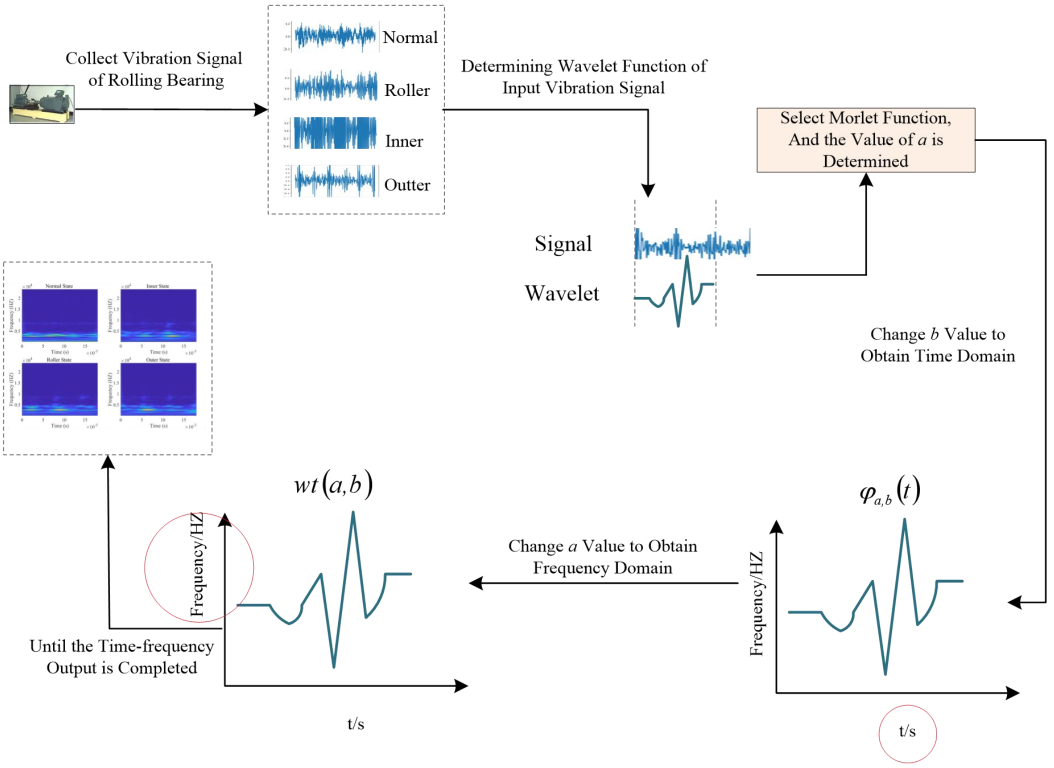


Fig. 2 Signal de-noising process of rolling bearing based on CWT.

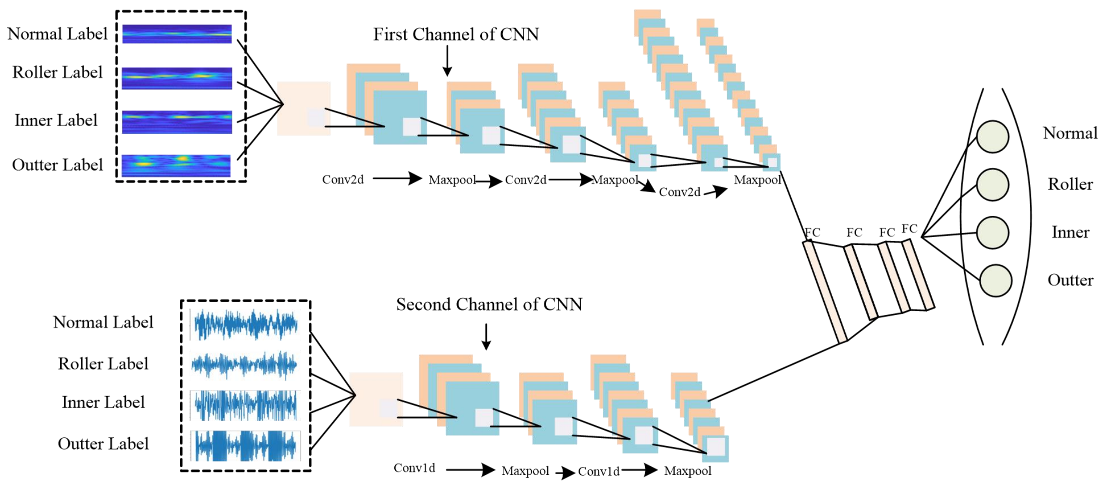


Fig. 3 Structure diagram of DCCNN processing vibration signal.

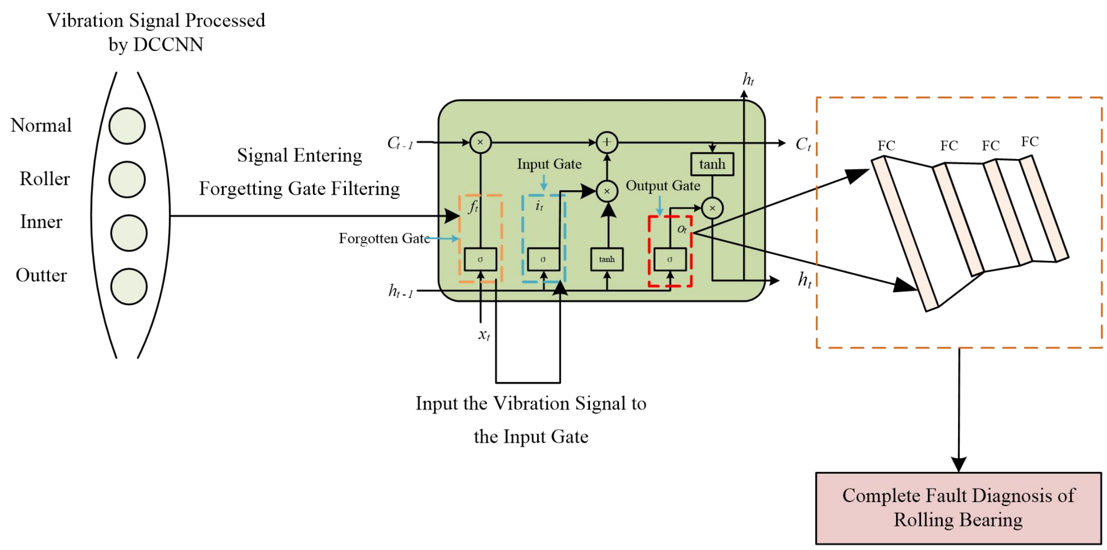


Fig. 4 Structure diagram of LSTM processing vibration signal.

the one-dimensional original signal processed by CWT is similar to the waveform of the wavelet basis function because the Morlet wavelet waveform is similar to the impact characteristics generated by the bearing fault, the Morlet function is selected as the waveform of the simulated rolling bearing vibration signal. The wavelet function is defined as follows

$$\phi(\mathrm{t}) = e^{imt} e^{-\frac{t^2}{2}} \tag{1}$$

Since the Morlet wavelet DC component is not zero, *a* Morlet function considering the correction term is proposed. The final definition of function is as follows :

$$\phi(\mathrm{t}) = e^{imt} e^{-\frac{1}{2}(ct)^2} - \sqrt{2} e^{-\frac{m^2}{4c^2}} e^{imt} e^{-(ct)^2} \tag{2}$$

Where *m* represents the main frequency of the wavelet function, and *c* represents the bandwidth of the modulated wavelet function.

The one-dimensional original signal acquisition time domain considering CWT processing is

$$\phi_{a,b}(t) = |a|^{-\frac{1}{2}} \phi\left(\frac{t-b}{a}\right) \tag{3}$$

Considering that the one-dimensional original signal processed by CWT obtains the frequency domain $wt(a,b)$, that is

$$wt(a,b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \phi^* \left(\frac{t-b}{a}\right) dt \tag{4}$$

Where *a* represents the scale parameter of the expansion or contraction wavelet of the rolling bearing vibration signal in the market function. *b* represents the shift parameter of the wavelet transformed along the time axis in the market function of the rolling bearing vibration signal.

B. Feature Extraction Based on DCCNN

After the vibration signal of the rolling bearing is denoised by CWT, the characteristic information of the signal needs to be extracted. DCCNN mainly extracts features from the time-frequency diagram of rolling bearings after CWT denoising, which can automatically extract features and avoid the influence of other factors, thus improving the recognition accuracy of rolling bearing faults. However, when the single-layer CNN extracts information, some information will be lost, and the optimal classification cannot be achieved. Therefore, a dual-channel convolutional neural network is used for the diagnosis of rolling bearing faults. So, after the signal is denoised by the CWT layer, it is input into the DCCNN layer to extract the feature information of the vibration signal.

The dual-channel CNN model uses the time-frequency map generated by CWT processing the original signal to input into the first CNN channel, and the one-dimensional original signal is input into the second CNN channel. After that, the signals enter their respective convolutional layers and pooling layers to further process the signals. Finally, after the last layer is summarized, it enters the fully connected layer. The DCCNN structure model is shown in Fig.3, where conv1d represents a one-dimensional convolutional layer, conv2d represents a two-dimensional convolutional layer, maxpool represents a pooling layer, and FC represents a fully connected layer.

As shown in Fig.3, the first channel convolution layer mainly processes the time-frequency diagram of the rolling bearing, and the second channel convolution layer mainly processes the one-dimensional original signal of the rolling bearing. Considering that the input of the neuron maps to the output, the real function is selected to operate on it. The convolution layer calculation formula is defined as

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{5}$$

$$f = \sigma\left(w_{L1}\left[h_{t-1}, x_t\right] + b_{L1}\right) \tag{6}$$

Where x_j^l represents the time-frequency diagram of the jth rolling bearing in the *l* layer after processing. *f()* represents the calculation of the activation function, which selects the relu function. M_j represents the set of selected rolling bearing time-frequency diagrams. The * sign indicates that the convolution kernel k is convoluted on all the associated rolling bearing time-frequency diagrams of the *l-1* layer. The purpose of this layer is to reduce the influence of the waveform phase of rolling bearing on diagnosis results.

The pooling layer is to reduce the dimension of the vibration signal of the rolling bearing. The pooling layer can be calculated by the following formula

$$X_j^l = f\left(\gamma_j^l down\left(x_i^{l-1} + b_j^l\right)\right) \tag{7}$$

Where *down()* denotes the pooling function. γ_j^l represents multiplicative bias. The purpose of this layer is to indicate whether the signal has failed and align the time series of fault events.

The fully connected layer is mainly classified, and The forward propagation formula is defined as

$$z^{l+1(j)} = W_{ij}^l d^{l(i)} + b_j^l \tag{8}$$

Where $z^{l+1(j)}$ represents the logistics value of the jth output neuron in the *l+1* layer. W_{ij}^l represents the weight between neuron *I* and the *l+1*th neuron in the *l* th layer. This layer is mainly to complete the classification of rolling bearing faults.

C. Time Series Extraction Based on LSTM

After the feature extraction of the vibration signal of the rolling bearing by the DCCNN layer, the time series feature of the signal needs to be extracted. LSTM is mainly to solve the time series problem of rolling bearings and further filter the vibration signals processed by DCCNN. Because LSTM is an improvement made to solve the long-term dependence problem of general recurrent neural network (RNN), after the signal is processed in the pooling layer of DCNN, it is selected to be summarized in the LSTM layer.

The CWT-DCCNN-LSTM model is composed of three layers of the LSTM network. Through the three parts of the forgetting gate, input gate, and output gate, the vibration signal is filtered, so that the time series features ignored by the 1D-CNN part can be further extracted to ensure the accuracy of the fault diagnosis model. The process of processing vibration signals by LSTM is shown in Fig.4.The LSTM structure has three steps :

Step1: The vibration signal is processed by the forgetting gate. The forgetting gate processes the convolution layer in

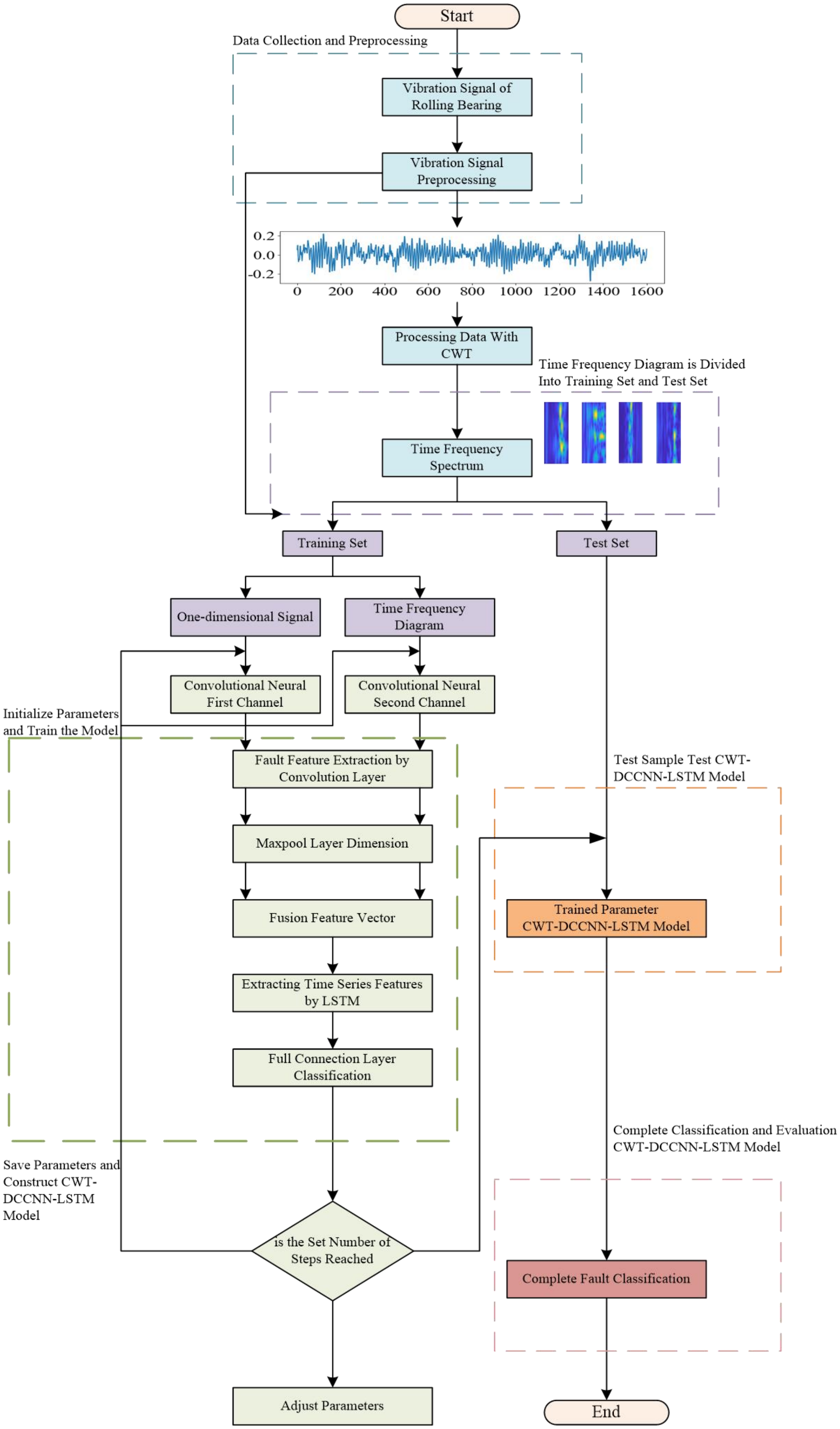


Fig. 5 Flowchart of rolling bearing fault detection model based on CWT-DCCNN-LSTM

the DCCNN layer, and the pooling layer does not filter the information, to forget the information that is considered useless, and further processes the information of the vibration signal. The specific calculation formula for forgetting the gate is as follows

$$Z^{f} = \sigma\left(W_{f} * \left[h_{t-1}, x_{t}\right] + b_{f}\right) \tag{9}$$

Where σ denotes the sigmoid function. W_f represents the weight matrix. [ht-1, xt] represents connecting two vectors to form a long vector. b_f represents the biased term.

Step2:The information of the vibration signal is input into the forgetting gate and then passed into the input gate. The input gate is mainly to process the information left by the forgetting gate, mainly to update the old unit state. The specific calculation formula of the input gate is as follows

$$i_{t} = \sigma\left(W_{i} * \left[h_{t-1}, x_{t}\right] + b_{i}\right) \tag{10}$$

$$\tilde{C}_{t} = \tanh\left(W_{C} * \left[h_{t-1}, x_{t}\right] + b_{c}\right) \tag{11}$$

Where $\tilde{C}_t$ represents the information generated by the repeating module.

Step3:After the information is processed by the forgetting gate and the input gate, the information flows into the output gate and outputs useful information. The specific calculation formula of the output gate is as follows

$$o_{t} = \left(W_{o}\left[h_{t-1}, x_{t}\right] + b_{o}\right) \tag{12}$$

$$h_{t} = o_{t} * \tanh\left(C_{t}\right) \tag{13}$$

Where C_t represents the updated unit state.

For the prediction of rolling bearing faults, the model is mainly considered from three aspects. Firstly, considering the noise contained in the obtained rolling bearing vibration signal, to denoise, the CWT layer is input for processing. Secondly, considering the extraction of vibration signal features, input to the DCCNN layer. Finally, considering the time series characteristics of the vibration signal, the LSTM layer is input for processing. The establishment of the whole model takes into account three aspects to be dealt with in predicting the vibration fault of rolling bearings In summary, the model structure flow chart of rolling bearing fault detection based on CWT-DCCNN-LSTM is shown in Fig.5.

III. Data Analysis

A. Data Sources

The data come from the Bearing Data Center of Case We stern Reserve University [41]. The dimensions of the rollin g bearings are 0.007 in, 0.014 in, and 0.021 in, respectively,

TABLE I

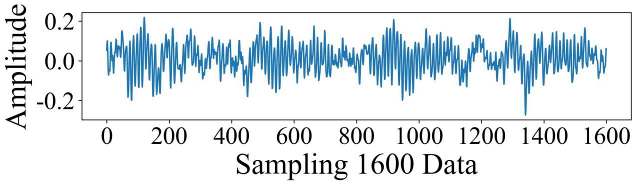
Units for Magnetic Properties Rolling Bearing Data Types

data type	failure mode	Rolling bearing size	sampling set
1	inner fault	0.007 in	1600
2	inner fault	0.014in	1600
3	inner fault	0.021in	1600
4	outer fault	0.007 in	1600
5	outer fault	0.014in	1600
6	outer fault	0.021in	1600
7	roller fault	0.007 in	1600
8	roller fault	0.014in	1600
9	roller fault	0.021in	1600
10	normal		1600

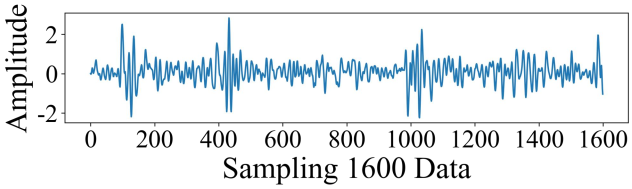
and have four load states. The sampling frequency of each type of data is 48 kHz. Each data set samples 1600 samples, and according to the 7: 3 as the training set and test set. The whole model is run through Python software and runs under the framework of TensorFlow. The specific data types of rolling bearings are shown in Table I.

B. Data Processing

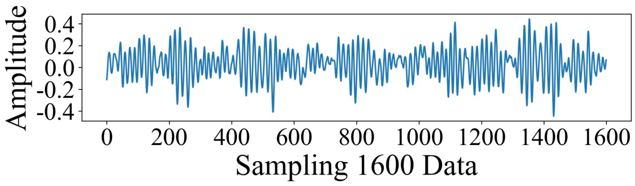
It can be seen from Table I that the rolling bearing is composed of an inner ring, a rolling element, and an outer ring. The waveform diagrams of the inner ring, rolling body, and outer ring of the rolling bearing and the normal state are visualized. Therefore, the six waveforms of rolling bearings are shown in Fig.6 a, b, c, d, e, and f.



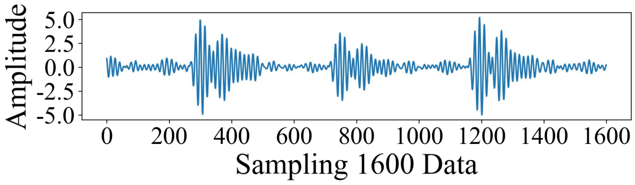
(a)Normal



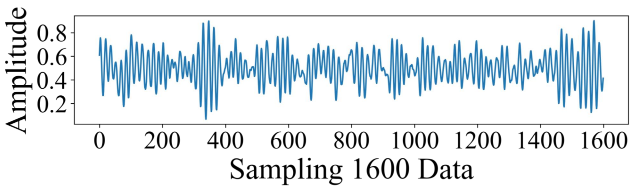
(b)Inner



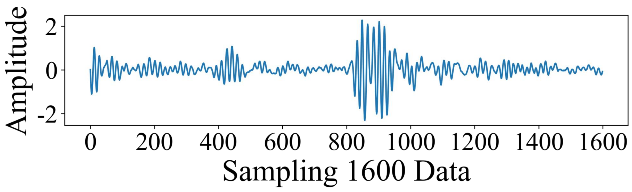
(c)Rolling



(d)Outer



(e)Outer

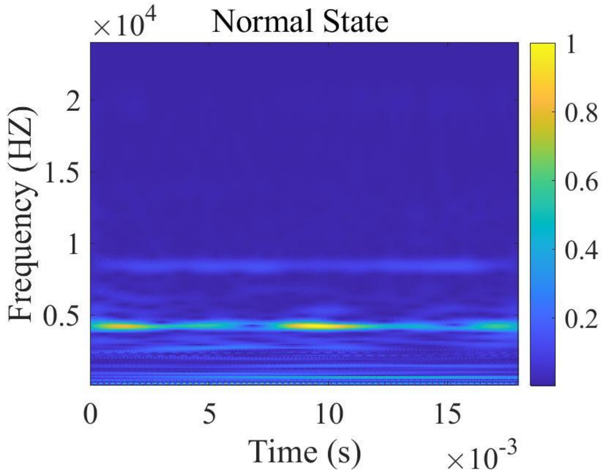


(f)Outer

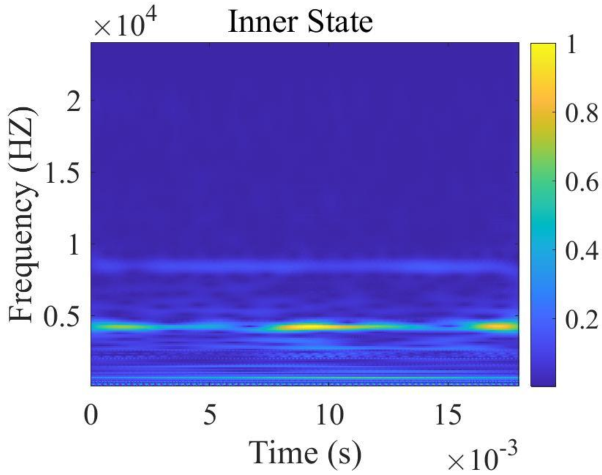
Fig. 6 Data waveform of rolling bearing.

As shown in Fig.6, the horizontal and vertical coordinates represent the amount of data sampled by the rolling bearing, and the vertical coordinates represent the amplitude of the rolling bearing signal. Through the comparative analysis of the three states of the inner ring state, the rolling body state, the outer ring, and the normal state in Fig.6, there are differences in the distribution of sample data. After the noise processing of the vibration signal of the rolling bearing by CWT, the time-frequency diagrams of normal, inner ring, rolling element, outer ring center direction, orthogonal direction, and relative direction are shown in Fig.7 a, b, c, d, e, f.

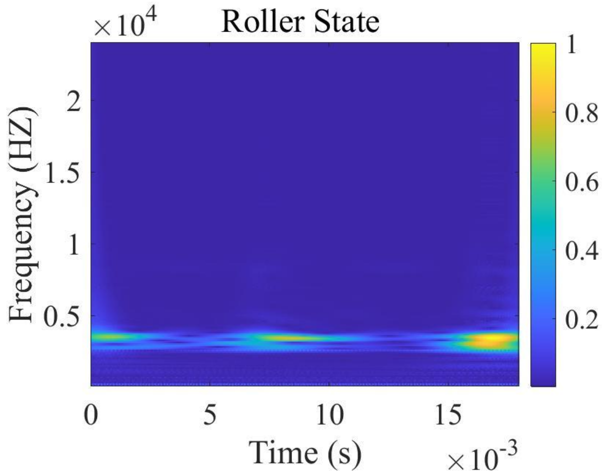
As shown in Fig.7, the time-frequency analysis results of CWT are presented in the form of a time-frequency diagram. Under normal conditions, the vibration signal sampled at a frequency of 0.5 f / Hz for 0.01 s has 1600 data points. When



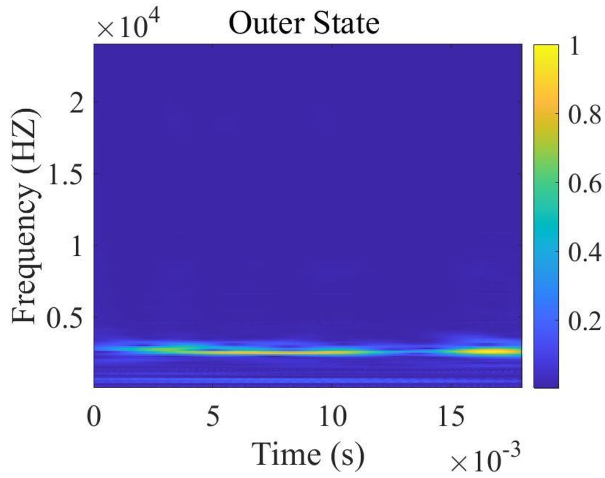
(a)



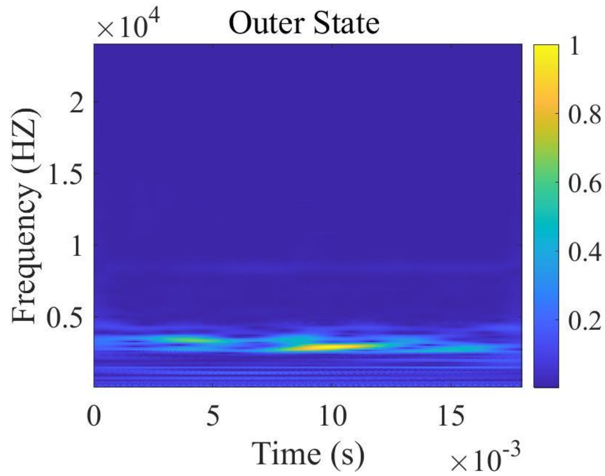
(b)



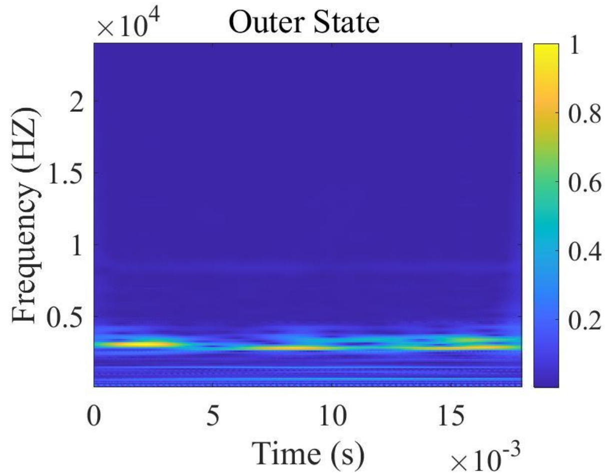
(c)



(d)



(e)



(f)

Fig. 7 Time-frequency diagram of continuous wavelet transform

the bearing is in a fault state, the amplitude of the signal is about -40 to 40, which is larger than normal. In the wavelet power spectrum, the color of each point represents the size of the wavelet coefficients on the time-frequency grid. The yellow representation is well localized in the time and frequency domains. After the vibration signal of the rolling bearing is denoised by CWT, the time-frequency diagram represents the frequency of the vibration signal and the specific position of the time domain, which can detect the fault more accurately.

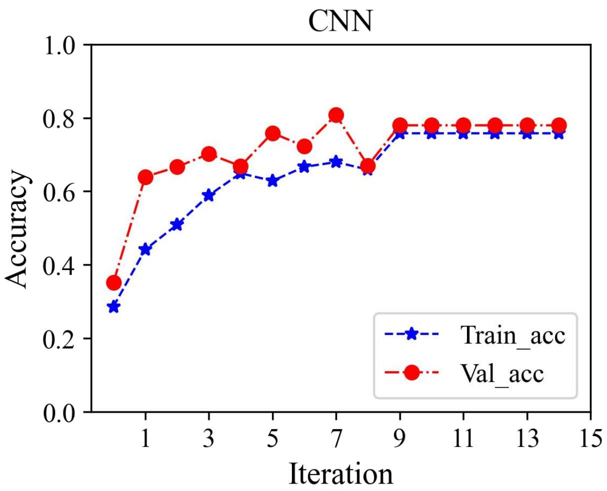
C. Comparative Analysis

Considering the three aspects of noise reduction, feature extraction, and time series extraction of rolling bearing vibration signals based on the CWT-DCCNN-LSTM model, to verify the performance of the three models, the model is divided into CNN, CWT-CNN, CNN-LSTM, and CWT-CNN-LSTM. The five models were evaluated from three aspects: accuracy, recall, and F1-score. Each model epochs 15 times. The experimental results are shown in Table II.

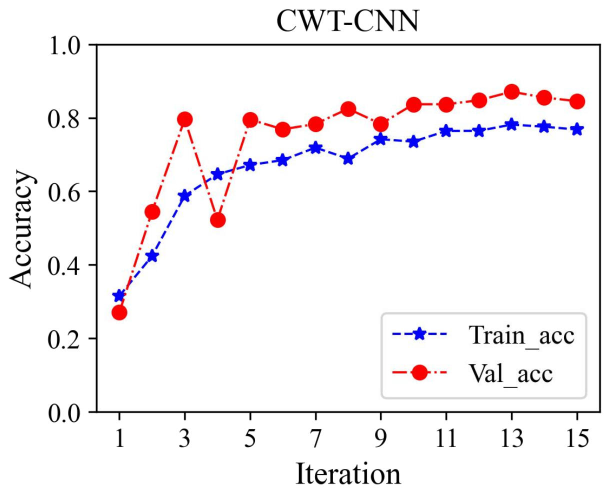
TABLE II
CWT-DCCNN-LSTM MODEL COMPARED TO OTHER MODELS

model	Accuracy	Recall	F1-score
CNN	0.78	0.74	0.73
CWT-CNN	0.85	0.78	0.78
CNN-LSTM	0.90	0.91	0.91
CWT-CNN-LSTM	0.98	0.98	0.98
CWT-DCCNN-LSTM	0.99	1.00	0.99

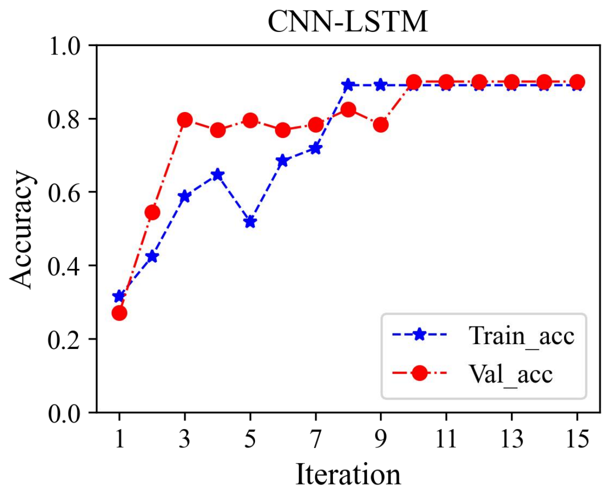
It can be seen from Table II that the CWT-DCCNN-LSTM model performs best in terms of accuracy, recall, and F1-score, which indicates that the model can improve the accuracy of rolling bearing fault diagnosis and better judge the fault of rolling bearing. Considering the vibration signal denoising, time series



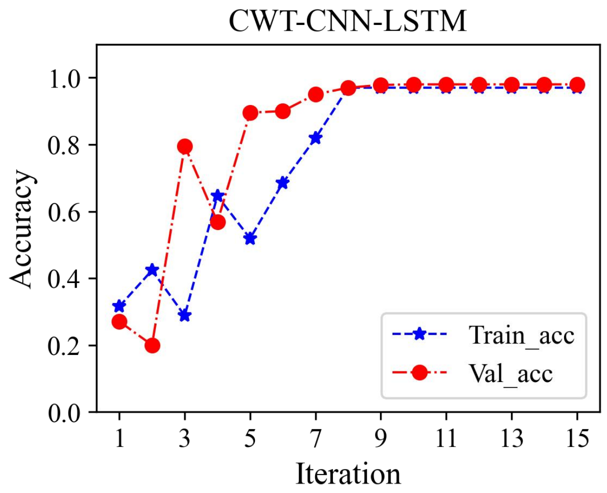
(a)



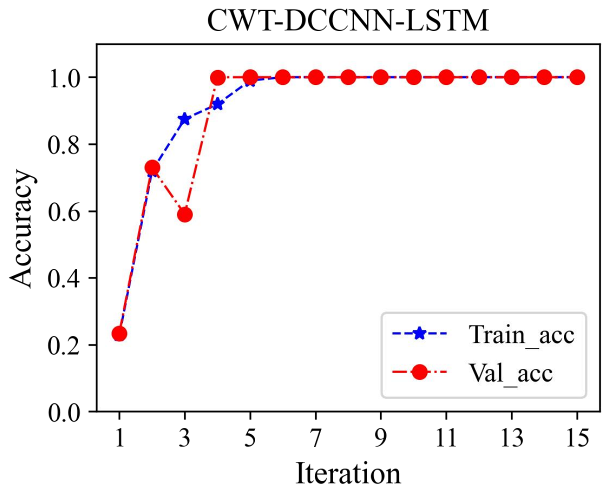
(b)



(c)



(d)



(e)

Fig. 8 The accuracy of CNN, CWT-CNN, CNN-LSTM, CWT-CNN-LSTM and CWT-DCCNN-LSTM models iteration

feature extraction, and Dual-Channel CNN feature extraction of rolling bearings, the above five models are divided into five comparison types: the model of CNN and the model of CWT-CNN comparison, the model of CNN and the model of CNN-LSTM comparison, the model of CNN-LSTM and the model of CWT-CNN-LSTM comparison, the model of CWT-CNN and the model of CWT-CNN-LSTM comparison, the model of

CWT-CNN-LSTM and the model of CWT-DCCNN-LSTM comparison. The accuracy of the five models are shown in Fig.8 a, b, c, d, e.

As shown in Fig.8, the above five models are divided into a comparative analysis of whether the vibration signal is denoised, a comparative analysis of whether the time series characteristics are considered, and a comparative analysis of the vibration signal input single-channel CNN or dual-channel CNN.

1) Comparative Analysis of Vibration Signal Denoising

CNN and CWT-CNN comparison analysis. The rolling bearing vibration signal is directly input into the CNN layer and the vibration signal is denoised and then input into the CNN layer for comparison. The accuracy of model CNN and model CWT-CNN are shown in Fig.8(a), and (b), respectively. (a) Ignoring the noise reduction of the vibration signal of the rolling bearing, combined with Table 3, the accuracy of noise reduction by CWT and then input to the CNN layer is 0.85, while the accuracy of only the CNN layer is 0.78. Therefore, when extracting data information, the data noise problem should be given priority. In summary, CNN and CWT-CNN are compared, and the model CWT-CNN should be selected.

CNN-LSTM and CWT-CNN-LSTM comparison analysis. Under the condition that both models consider the time series features of rolling bearing vibration signals, one model is non-noise and one model is denoised. The accuracies of model CNN-LSTM and CWT-CNN-LSTM are shown in Fig.8(c), and (d), respectively. Model CNN and model CWT-CNN were selected for comparison. Combined with Table 3, the accuracy of denoising the vibration signal is higher than that without denoising, the accuracy of model CNN-LSTM is 0.90 and that of model CWT-CNN-LSTM is 0.98, and the denoising reduces a certain error. In summary, the CWT-CNN-LSTM model should be selected.

Combined with the above, denoising should be considered for the vibration signals of rolling bearings, and model CWT-CNN with model CWT-CNN-LSTM has higher accuracy than the model without denoising.

2) Comparative analysis of time series characteristics

CNN and CNN-LSTM comparison analysis. Both models consist of CNN for feature extraction of rolling bearings, and the accuracy of models CNN and CNN-LSTM are shown in Fig.8 (a), and (c), respectively. (a) Without considering the time series problem of vibration signal, the accuracy of the CNN model is not as high as that of combined CNN-LSTM under the condition of simultaneous epoch 15 times, and the accuracy of the CNN model is 0.78 and that of CNN-LSTM model is 0.90. Therefore, for the fault identification of the rolling bearing, the time series problem needs to be considered. In summary, the CNN-LSTM model should be selected for the comparison of these two.

CWT-CNN and CWT-CNN-LSTM comparison analysis. Under the condition that both models denoise the rolling bearing vibration signal, one model does not consider the time series features of the vibration signal, and one model considers the time series features of the vibration signal. The accuracy of models CWT-CNN and CWT-CNN-LSTM are shown in Fig.8(b), and (d), respectively. Model (b) ignores the time series characteristics of the vibration signal by not considering the time series characteristics of the vibration signal, which leads to a lower accuracy of the model (b) than model (d). Combined with Table 3, from the comparison of the three criteria for model evaluation, model (d) outperforms model (b), and the accuracy of the CWT-CNN model is 0.85, and the accuracy of CWT-CNN-LSTM accuracy is In summary, the CWT-CNN-LSTM model should be chosen for the comparison of these two.

From 1), it can be seen that model CWT-CNN and model CWT-CNN-LSTM are more accurate than the model without denoising. However, after the comparative analysis of time series feature extraction, the model of CWT-CNN-LSTM is better than the three models CNN, CWT-CNN, and CNN-LSTM from the three aspects of accuracy, recall, and F1-score, so after these four comparisons, both denoising and time series feature extraction should be considered, so that the prediction accuracy of rolling bearing fault can reach 0.98, and among these four models, CWT-CNN-LSTM should be selected.

3) Comparison analysis of Single Channel and Dual Channel CNN

The model of CWT-CNN-LSTM and the model of CWT-DCCNN-LSTM comparison analysis. The two models simultaneously denoise the rolling bearing vibration signal and consider the time series features, one model inputs the vibration signal into a single-channel CNN and the other model inputs the vibration signal into a DCCNN. the accuracy of models CWT-CNN-LSTM and CWT-DCCNN-LSTM are shown in Fig.8(d), (e), respectively. Model (d) chooses a single-channel CNN which is prone to missing information, resulting in imperfect information and some prediction error. Model (b) selects DCCNN to make up for the deficiency due to single-channel CNN, and the accuracy of the CWT-CNN-LSTM model is 0.98 and the accuracy of the CWT-DCCNN-LSTM model is 99.98%. Therefore, the two-channel CNN is chosen to be superior to the single-channel CNN, and the CWT-DCCNN-LSTM model should be chosen for the comparison of these two.

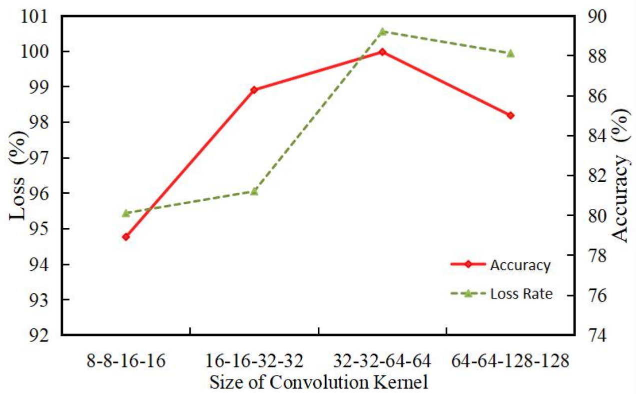
By the comparison of 1) and 2), the CWT-CNN-LSTM model is chosen. However, by the comparison of CWT-CNN-LSTM and CWT-DCCNN-LSTM, the CWT-DCCNN-LSTM model has a higher accuracy. This is because the CWT-DCCNN-LSTM model takes into account both the data noise problem, the time series feature extraction problem, and the problem that a single layer of CNN will have partial information loss when extracting information. The problems that may be encountered in prediction are considered from three aspects, and from a comprehensive comparison of Figure 8 and Table III, the accuracy, recall, and F1-score of the evaluated models for all three methods, the CWT-DCCNN-LSTM outperforms the other models in all three values, has higher prediction accuracy, and achieves optimal classification.

D. Parametric Analysis

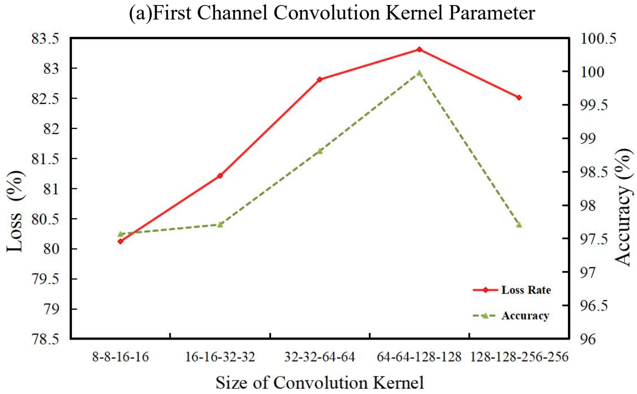
After the models have been compared and analyzed, the parameters of the models also need to be analyzed, because the parameter settings have an important impact on the final experimental results and can affect the performance of the models. To further improve the CWT-DCCNN-LSTM model performance, the convolutional kernel size, and LSTM hidden layer size are further explored.

1) The influence of convolution kernel size on the model

The size of the convolution kernel parameter is an important parameter in the convolution layer. The parameters of the first layer CNN channel convolution kernel and the second layer CNN channel convolution kernel are set to 8-8-16-16, 16-16-32-32, 32-32-64-64, 64-64-128-128, respectively, to test their influence on the accuracy of the model. The two-channel convolution kernel size accuracy and loss rates are shown in Fig.9 (a) and (b) respectively.



(a)First Channel Convolution Kernel Parameter



(b)Second Channel Convolution Kernel Parameter

Fig. 9 Convolution Kernel Parameters

As shown in Fig.9, (a) represents the first channel convolution kernel parameter, and (b) denotes the second channel convolution kernel parameter. When the number of convolution kernels is relatively small, the accuracy of fault classification and degree evaluation is relatively low, and the overall diagnosis result of the model is not ideal. When the number of convolution kernels is too large, the two tasks appear serious over-fitting phenomenon at the same time, and the accuracy of the model test decreases sharply. When the convolution kernel combination of the first channel convolution kernel combination is 32-32-64-64, and the second channel convolution kernel combination is 64-64-128-128, the fault classification accuracy is the highest. Therefore, the CWT-DCCNN-LSTM model finally studies the fault classification for the first channel convolution kernel combination of 32-32-64-64 and the second channel convolution kernel combination of 64-64-128-128.

2) LSTM hidden layer size

The size of the hidden layer parameters of the LSTM layer affects the accuracy based on the CWT-DCCNN-LSTM model. The LSTM layer hidden layer is set to 16-32, 32-64, 64-128, and 128-256 to test its effect on the accuracy of the model. The LSTM layer hidden layer for model accuracy and loss rate are shown in Fig.10.

As shown in Fig.10, when the number of hidden layers is relatively small, the accuracy of fault classification is relatively low, and the overall diagnosis result of the model is not ideal. When the number of hidden layers is too large, the two tasks appear serious over-fitting phenomenon at the

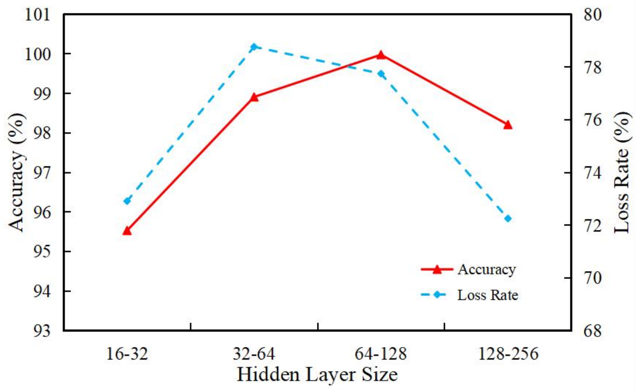


Fig. 10 Hiding Layer Parameter Size

same time, and the model test accuracy drops sharply. The fault classification accuracy is the highest when the hidden layer combination is 64-128. Therefore, the CWT-DCCNN-LSTM model finally studies the fault classification for the LSTM hidden layer of 64-128.

3) The influence of CWT time-frequency diagram size on the model

The size of the time-frequency diagram affects the accur acy of the model. The size of the time-frequency diagram is set to 10-10,20-20,30,40-40,50-50,60-60,70-70,80-80,90-9 0,100-100 to detect the influence of these parameters on the accuracy of the model. CWT time-frequency diagram size for model accuracy and loss rate are shown in Fig.11.

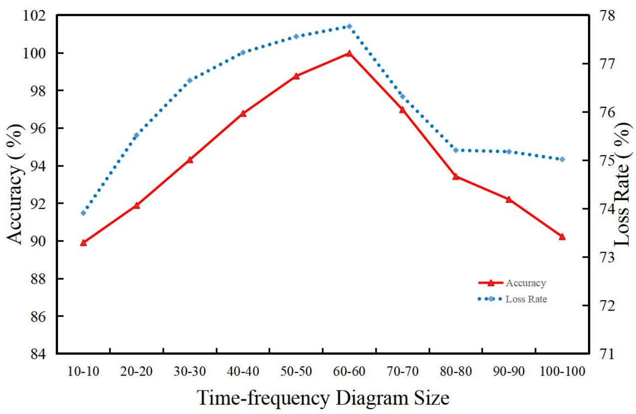


Fig. 11 Parameter Size of Time-frequency Diagram

As shown in Fig.11, when the frequency diagram is relatively small, the accuracy of fault classification is relatively low, and the overall diagnosis result of the model is not ideal. When the frequency map is relatively large, the two tasks have a serious over-fitting phenomenon at the same time, and the accuracy of the model test decreases

TABLE III
FIRST CHANNEL PARAMETERS OF CONVOLUTIONAL NEURAL NETWORK

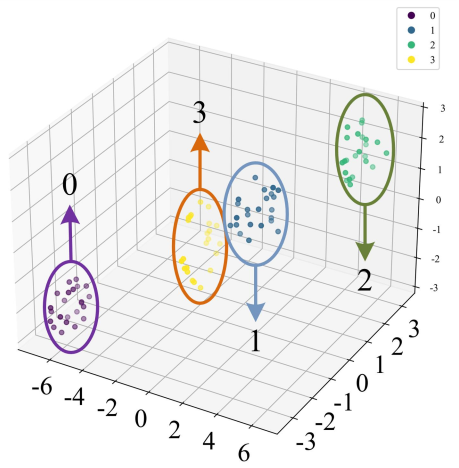
Network layer	Convolution kernel/step size （or other parameters）	Number of convolution kernels	strides
Convolutional layer1	3x3	32	1,1
Pooling layer1	2x2	32	1,1
Convolutional layer2	3x3	64	1,1
Pooling layer2	2x2	64	1,1

TABLE IV
SECOND CHANNEL PARAMETERS OF CONVOLUTIONAL NEURAL NETWORK

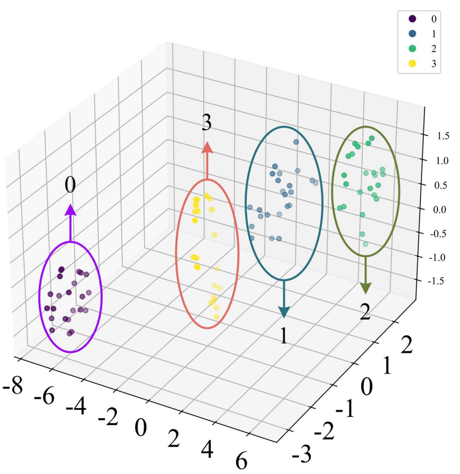
Network layer	Convolution kernel/step size （or other parameters）	Number of convolution kernels	strides
Convolutional layer1	3x2	64	2
Pooling layer1	2x2	64	2
Convolutional layer2	3x3	128	2
Pooling layer2	2x2	128	2

TABLE V
UNITS FOR MAGNETIC PROPERTIES LSTM AND DROPOUT PARAMETERS

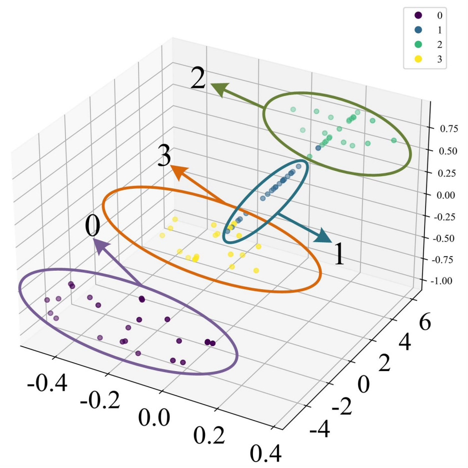
Network layer	Convolution kernel/step size （or other parameters）	Number of convolution kernels
LSTM	hidden_size=10	64-128
Dropout layer	Forgetting rate=0.5	



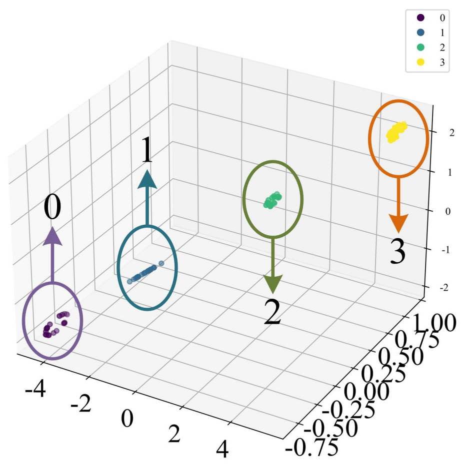
(a)



(b)



(c)



(e)

Fig 12. T-SNE Visualization Scatter Plot sharply. The fault classification accuracy is the highest when the frequency diagram size is 60-60. Therefore, the CWT-DCCNN-LSTM model ultimately selects 60-60 for the time-frequency map size.

E. T-SNE Visualization

The t-SNE is a visualization of the data to be able to compare whether this data set is separable or not, and at the same time to be able to verify the classification ability of the model to complete the prediction of rolling bearing failures. t-SNE validated 3D scatter feature distribution is shown in Fig.12.

In Fig.12, 0 indicates normal data, 1 indicates inner ring fault, 2 indicates rolling body fault, and 3 indicates outer ring fault. (a) indicates the visualization of rolling bearing data input to the first channel convolutional neural network, (b) indicates the visualization of rolling bearing data input to the second channel convolutional neural network, (c) indicates the process of detection, and (d) indicates the final predicted result. The input of data at the very beginning is confusing and does not follow the order of normal, inner ring, rolling body, and outer ring. After the training of the model, the data are stacked from the beginning until the classification of rolling bearings is finally completed, thus completing the prediction. Fig.12 demonstrates the feasibility of the model CWT-DCCNN-LSTM with high prediction accuracy.

In summary, the high accuracy of the model was demonstrated by the comparative and parametric analysis, and the feasibility of the model was verified by the t-SNE visualization. The parameters of the main network layers of the model are shown in Table III, Table IV, and Table V.

IV. Conculsion

To improve the accuracy of detecting rolling bearing faults, a diagnosis method based on the CWT-DCCNN-LSTM model is proposed. The advantages of CWT, CNN, and LSTM networks are combined to extract feature information from the acquired original vibration signals of rolling bearings. The CWT-DCCNN-LSTM model is more accurate than other methods and achieves 99.98% accuracy, which is better than other networks. The model takes into account both the noise of the data and the lack of information in the single-layer convolutional neural network, which is a more comprehensive consideration compared to other models. The model has the following main features.

1) The model selects CWT for the vibration signal of the rolling bearing for noise reduction processing, which weakens the artificial error. DCCNN method is selected for feature extraction of the vibration signal of the rolling bearing, which avoids the error of information loss in a single channel. The LSTM method is selected for time series feature extraction of vibration signals of rolling bearings. Finally, the fault of the rolling bearing is detected based on the CWT-DCCNN-LSTM model, which improves the accuracy of detection.

2) Through the comparison of model CNN and CWT-CNN, the comparison of model CNN, LSTM, and CNN-LSTM, the comparison of model CNN-LSTM and CWT-CNN-LSTM, the comparison of model CWT-CNN-LSTM and CWT-DCCNN-LSTM, four groups of comparative experiments are analyzed.

3) By analyzing the parameters of the CWT-DCCNN-LSTM model, it is finally determined that the first channel convolution kernel combination is 32-32-64-64, the second channel convolution kernel combination is 64-64-128-128, the time-frequency diagram size is 60-60, and the LSTM hidden layer is 64-128. The parameters of the model are verified by t-SNE visualization. Finally, it is shown that the model has the highest accuracy and good classification ability under this parameter.

However, due to the level of the model's more, so for the model although high accuracy, the running speed compared with other methods have no advantage, after further study.

References

[1] M. Xu, Q.N. Zhang, "Application of vibration signal analysis techniques in rolling bearing fault diagnosis," *INTER - NOISE and NOISE - CON Congress and Conference Proceedings*, vol.1983, no.2, pp.1147-1150, 1983.

[2] J.M. Li, J.F. Zhang, M. Li, Y.G. Zhang, "A novel adaptive stochastic resonance method based on coupled bistable systems and its application in rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol.114, pp.128-145, 2019.

[3] W. Cheng, X.M. Xu, Y.P. Ding, K.H. Sun, "Stochastic resonance in a single-well potential and its application in rolling bearing fault diagnosis," *The Review of scientific instruments*, vol.91, no.6, Article ID. 064701, 2020.

[4] X.J. Zhao,Y. Qin, C.B. He, L.M. Jia, L.L. Kou, "Rolling Element Bearing Fault Diagnosis under Impulsive Noise Environment Based on Cyclic Correntropy Spectrum," *Entropy*, vol.21, no.1, pp.50, 2019.

[5] Q.Y. Liu, H.Y. Pan,J.D. Zheng, J.Y. Tong, J.H. Bao, "Composite Interpolation-Based Multiscale Fuzzy Entropy and Its Application to Fault Diagnosis of Rolling Bearing," *Entropy*, vol.21, no.3, pp.50, 2019.

[6] J.D. Zheng, H.Y. Pan, Q.Y. Liu, K.Q. Ding, "Refined time-shift multiscale normalised dispersion entropy and its application to fault diagnosis of rolling bearing," *Physica A: Statistical Mechanics and its Applications*, vol.545, pp.123641-123641, 2020.

[7] Z.Y. Quan and X.L. Zhang, "Rolling bearing fault diagnosis based on CS-optimized multiscale dispersion entropy and ML-KNN," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol.44, no.9, pp.430, 2022.

[8] W.M. Ying, J.Y. Tong, Z.L. Dong, H.Y. Pan, Q.Y. Liu, J.D. Zheng, "Composite Multivariate Multi-Scale Permutation Entropy and Laplacian Score Based Fault Diagnosis of Rolling Bearing," *Entropy*, vol.24, no.2, pp.160-160, 2022.

[9] T.T Xing, Y. Zeng, Z. Meng, X.L. Guo, "A fault diagnosis method of rolling bearing based on VMD Tsallis entropy and FCM clustering," *Multimedia Tools and Applications*, vol.79 pp.1-17, 2020.

[10] Y.F. Peng, J.H. Chen, Y.F. Liu, J.S. Cheng, Y.Yang, K.F. He, et al, "Roller Bearing Fault Diagnosis Based on Adaptive Sparsest Narrow-Band Decomposition and MMC-FCH," *Shock and Vibration*, vol.2019, pp.1-17, 2019.

[11] Y. Cheng, Z.W. Wang, B.Y. Chen, W.H. Zhang, G.H. Huang, "An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis," *ISA Transactions*, vol.91, pp.218-234, 2019.

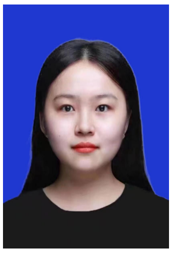
[12] Q.Q. Chen, S.W. Dai and H.D. Dai, "A Rolling Bearing Fault Diagnosis Method Based on EMD and Quantile Permutation Entropy," *Mathematical Problems in Engineering*, vol.2019, pp.1-8, 2019.

[13] X.P Luo, X. Xu and Herschel Rabitz. "On the fundamental conjecture of HDMR: a Fourier analysis approach," *Journal of Mathematical Chemistry*, vol.55, no.2, pp.632-660, 2017.

[14] Bendjama Hocine, "Bearing fault diagnosis based on optimal Morlet wavelet filter and Teager-Kaiser energy operator," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol.44, no.9, pp.392, 2022.

[15] Y.B. Li, X.H. Liang, M.Q. Xu, W.H. Huang, "Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor

wavelet transform," *Mechanical Systems and Signal Processing*, vol.86, pp.204-223, 2017.
[16] S.Y. Liu, X. Yu, Q. Xu, D. Fei, "Rolling Bearing Fault Diagnosis Based on Sensitive Feature Transfer Learning and Local Maximum Margin Criterion under Variable Working Condition," *Shock and Vibration*, vol.2020, pp.1-34, 2020.
[17] Anil Kumar, C.P. Gandhi, Y.Q. Zhou, H.S. Tang, J.W. Xiang, "Fault diagnosis of rolling element bearing based on symmetric cross entropy of neutrosophic sets," *Measurement*, vol.152, no. C, Article ID.107318, 2020.
[18] Y.G. Xu, J.X. Cao, J.Y. Zhao, K. Zhang, W.K. Tian, "Application of fast singular spectrum decomposition method based on order statistic filter in rolling bearing fault diagnosis," *Measurement Science and Technology*, vol.30,no.12, pp.125001-125001, 2019.
[19] Z.C. Qiao, Y.Q. Liu and Y.Y. Liao, "An Improved Method of EWT and Its Application in Rolling Bearings Fault Diagnosis," *Shock and Vibration*, vol.2020, pp.1-13, 2020.
[20] J. Gu, Y.X. Peng, H. Lu, X.D. Chang, G.A. Chen, "A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN," *Measurement*, vol.200, Article ID.111635, 2022.
[21] Bhavsar Keval, Vakharia Vinay, Chaudhari Rakesh, Vora Jay, Pimenov Danil Yurievich, Giasin Khaled, "A Comparative Study to Predict Bearing Degradation Using Discrete Wavelet Transform (DWT), Tabular Generative Adversarial Networks (TGAN) and Machine Learning Models," *Machines*,vol.10,no.3, pp.176-176,2022.
[22] J.S. Cheng, D.J. Yu, Q.W. Deng, S. Yang, "Application of Continuous Wavelet Transform to Fault Diagnosis of Rolling Bearings," *China Mechanical Engineering*, vol. 14,no23, pp.64-67+6, 2003.
[23] J. Lin, "Continuous Wavelet Transform and Its Application for Bearing Diagnosis," *JOURNAL OF XI'AN JIAOTONG UNIVERSITY*, vol.1999, no.11, pp.110-112, 1999.
[24] Le Cun Y, Boser B and Denker JS, "Backpropagation applied to handwritten zip code recognition." *Neural Computation*," vol.1, no.4, pp.541-551, 1989.
[25] Hinton GE and Salakhutdinov RR, "Reducing the dimensionality of data with neural networks," *Science*, vol.313, no.5786, pp.504-507,2006,
[26] X. Li, W. Zhang, Q. Ding, J.Q Sun, "Multi-Layer domain adaptation method for rolling bearing fault diagnosis," *Signal Processing*, vol.157, pp.180-197, 2019.
[27] L.J. Wan, Y.W. Chen, H.Y. Li, C.Y. Li. "Rolling-Element Bearing Fault Diagnosis Using Improved LeNet-5 Network," *Sensors*, vol.20, no.6, Article ID.1693, 2020.
[28] Sadoughi Mohammadkazem and C. Hu, "Physics-Based Convolutional Neural Network for Fault Diagnosis of Rolling Element Bearings," *IEEE Sensors Journal*, vol.19, no.11, pp. 4181-4192, 2019.
[29] Abdelraouf Youcef Khodja, Noureddine Guersi, Mohamed Nacer Saadi, Nadir Boutasseta. "Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks," *The International Journal of Advanced Manufacturing Technology*, vol.106, no.5, pp.1737-1751, 2020.
[30] C.C. Che, H.W. Wang, X.M. Ni, Q. Fu, "Domain Adaptive Deep Belief Network for Rolling Bearing Fault Diagnosis," *Computers & Industrial Engineering*, vol.143, pp. 106427-106427, 2020.
[31] L.F. Gou, H.H. Li, H. Zheng, H.C. Li, X.N. Pei, "Aeroengine Control System Sensor Fault Diagnosis Based on CWT and CNN," *Mathematical Problems in Engineering*, vol.2020, pp.1-12, 2020.
[32] Y.P. Liu, H.K. Jiang, C.Q. Liu, W.F.Yang, W. Sun, "Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis," *Knowledge-Based Systems*, vol.252, Article ID. 109439, 2022.
[33] R.J. Liang , W.F. Ran, C.L. Yu, W.F. Chen, Ni De, "Recognition of gearbox operation fault state based on CWT-CNN," *Journal of Aerospace Power*, vol.36, no.12, pp.2465-2473, 2021.
[34] Z.F. Yu, B.S. Xiong, T.Y. Xiong, Q.F. Ou, X.M. Li, "Fault diagnosis of helicopter based on VMD-CWT and improved CNN," *Journal of Aerospace Power*, vol.36, no.05, pp. 948-958, 2021.
[35] Z.H. Wu, H.K. Jiang, T.F. Lu, K. Zhao, "A deep transfer maximum classifier discrepancy method for rolling bearing fault diagnosis under few labeled data," *Knowledge-Based Systems*, vol. 196, pp. 105814-105814, 2020.
[36] C. Zhong, Y. Liu, J.S. Wang, Z.F. Li, "LSTM Neural Network Fault Diagnosis Method for Rolling Bearings Based on Information Fusion," *IAENG International Journal of Computer Science*, vol.49, no.4, pp.1088-1098, 2022.
[37] J.B.Zheng, J. Liao and Z.B. Chen, "End-to-End Continuous/ Discontinuous Feature Fusion Method with Attention for Rolling Bearing Fault Diagnosis," *Sensors*, vol.22, no.17, pp.6489-6489, 2022.
[38] M.X. Liang, P. Cao and J. Tang, "Rolling bearing fault diagnosis based on feature fusion with parallel convolutional neural network," *The International Journal of Advanced Manufacturing Technology*, vol.112, pp.1-13, 2020.
[39] C.L. Lei, L.L. Xue, M.X. Jiao, H.Q. Zhang, J.S. Shi, "Rolling bearing fault diagnosis by Markov transition field and multi-dimension convolutional neural network," *Measurement Science and Technology*, vol.33, no.11, Article ID.114009, 2022.
[40] R.X. Bai, Q.S. Xu, Z. Meng, L.X. Cao, K.S. Xing, F.J. Fan, "Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation," *Measurement*, vol.184, Article ID.109885, 2021.
[41] Case Western Reserve University Bearing Data Center n.d.[Online]. Available: http://csegroup.case.edu/bearingdatacenter/home



Yu Wang was born in Gansu, China. In 1999. She obtained her Bachelor's degree in Traffic and Transportation from Lanzhou Jiaotong University, Lanzhou, China, in the year 2021. She is currently pursuing her postgraduate student in Traffic and Transportation (Management Science and Engineering) at Lanzhou Jiaotong University. And get a second-class scholarship. Her research interests include transportation system management and deep learning method analysis data.