



---

*Research article*

## **Bearing fault diagnosis with parallel CNN and LSTM**

**Guanghua Fu\*, Qingjuan Wei and Yongsheng Yang**

Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China

\* **Correspondence:** Email: [ghfu@shmtu.edu.cn](mailto:ghfu@shmtu.edu.cn).

**Abstract:** Intelligent diagnosis of bearing faults is fundamental to machinery automation and their intelligent operation. Deep learning-based analysis of bearing vibration data has emerged as one research mainstream for fault diagnosis. To enhance the quality of feature extraction from bearing vibration signals and the robustness of the model, we construct a fault diagnostic model based on convolutional neural network (CNN) and long short-term memory (LSTM) parallel network to extract their temporal and spatial features from two perspectives. First, via resampling, vibration signal is split into equal-sized slices which are then converted into time-frequency images by continuous wavelet transform (CWT). Second, LSTM extracts the time-correlation features of 1D signals as one path, and 2D-CNN extracts the local frequency distribution features of time-frequency images as another path. Third, 1D-CNN further extracts integrated features from the fusion features yielded by former parallel paths. Finally, these categories are calculated through the softmax function. According to experimental results, the proposed model has satisfactory diagnostic accuracy and robustness in different contexts on two different datasets.

**Keywords:** bearing fault diagnosis; CWT; CNN; LSTM; parallel path

---

### **1. Introduction**

Machinery operation and maintenance tend to be automated and intelligent, which challenges the diagnosis of their faults. As key and essential parts in rotating device, bearings are prone to failure since their operation tend to be affected by severe working conditions [1]. Once the bearing fails, it causes serious system collapse and equipment damage, even cause great economic losses and casualties. Therefore, effective fault diagnosis and operating state analysis of bearings can help to detect mechanical equipment failures in time, which is the pre-requisite of reliable and stable operation

of mechanical systems. The faults of bearings could be reflected by their vibration signals, which are time series data of periodic oscillation and contain various information about operational state [2]. In different working conditions, the mapping relationship between the vibration signal of bearings and their fault state becomes more complicated, which increases the difficulty of fault identification [3]. Over the last decades, based on vibration signal analysis, research on bearing fault diagnosis have mainly gone through three phases: (i) Fault diagnosis based on signal modal analysis; (ii) fault diagnosis based on traditional shallow machine learning; (iii) fault diagnosis based on deep learning.

Research on fault diagnosis based on signal modal analysis gradually develops into a mature methodology. Time-frequency methods are widely used for the pre-processing of bearing vibration signals, including empirical mode decomposition (EMD) [4], variational mode decomposition (VMD) [5], short-time Fourier transform (STFT) [6], and wavelet transform (WT) [7], etc. Focusing on the time domain, STFT can extract vibration components of the signal distribution in the frequency domain, but it is not applicable to time-varying signals. Based on STFT, WT can analyze signals at multiple scales, capturing details and features in different frequency ranges, and has been used extensively in signal analysis [8,9].

The development of machine learning push fault diagnosis towards automated and intelligent diagnosis. Generally, traditional shallow machine learning methods in fault diagnosis are usually utilized for signal processing, feature extraction, and final fault classification by a classifier [10–13]. Commonly-used methods in fault diagnosis include support vector machines (SVM) [14], Bayesian classifiers [15], artificial neural networks (ANN) [16], etc. Li et al. [17] used EEMD to decompose signal and extracted fault feature frequency from optimized signal by power spectrum to identify the fault. ANN could utilize the features extracted from envelope analysis to realize cross-domain fault diagnosis [18]. Although great improvement has been achieved by these traditional machine learning methods in fault diagnosis, their diagnostic accuracy depends on manual feature selection and extraction. Thus, some scholars put methods for adaptive feature extraction. Li et al. [16] decomposed the signal into sub-signals of multiple scales, extracted their local features with BPNN, and identified fault features by SVM. However, the diagnostic performance of traditional shallow machine learning models (e.g., SVM, ANN) is limited by the quality of extracted fault features. Moreover, data from sensors in real time are so diverse and ever-increasing that these methods cannot fully utilize all information, and the accuracy goes down with the increasing quantity of the data.

Deep learning network (e.g., CNN, RNN, DBN, GAN, AE) [19–23], as an extension of traditional machine learning based on neural networks, has shown significant superiority in image classification, natural language processing, and target detection [24–26]. It has excellent capability of feature representation, and could extract deep features of the data by constructing a deep network and achieve the complex nonlinear mapping between the pre-processing data and the fault states. In fault diagnosis, CNN, as a typical deep learning network, has achieved a great deal of excellence [27–31]. Many stacked CNNs have been proposed which could directly extract information from the input of time domain vibration signal without priori expert knowledge, and achieve good performance under variable loads [3,32]. Ye et al. [33] used adaptive variational modal decomposition to preprocess the signal to obtain the desired components, and a modified 1D-CNN is used to extract features and identify faults. Vibration signals of the faults belong to time series and contain temporal features. Although CNN could extract their spatial features, it cannot fully mine the fault features from the perspective of local space. While RNN could effectively deal with time series data, and better handle vibration signals with time-series correlation [34,35]. An et al. [36] proposed a diagnostic method

based on periodic sparse attention and LSTM, where LSTM is used to extract time-correlated features of the signal. Some scholars also proposed a combination of 2D-CNN and LSTM for fault diagnosis [37]. Although neural network-based diagnosis models already have a high diagnostic accuracy, some fails to fully consider the feature specialty of vibration signals, and is unable to fully explore the time domain and frequency domain features. These inadequacies are reflected in the following two aspects: (i) The vibration signals belong to periodic vibration signals, if the signal analysis merely from the time domain or the frequency domain cannot adequately extract the fault features of the signals. (ii) The combination of CNN and LSTM for fault diagnosis is often utilized in a single-path and series way. In the series situation, the output of one network will be taken as the input of the following network, and extracted features of the latter are inevitably influenced by the former, thus negatively affecting performance of feature extraction and fault classification of the model. Moreover, as the deep network continue to be deeper, these diagnostic models tend to be overfitting [38,39]. Extra optimization needs to be done to solve this problem [33,40].

To avoid over-stacked and dilated CNNs, attention mechanism is introduced to dynamically focus on different parts of the input, especially when sequential data or images are subject to multiple target processing with multi-stages [41,42]. Wang et al. [43] embedded the attention mechanism into a neural network to achieve high diagnostic performance even if noises are added into the signal. Some scholars suggested converting original signal into two-dimensional (2D) images to facilitate feature extraction via CNN [37,44,45]. Wen et al. [46] designed an improved LeNet-5 network with high diagnostic accuracy which could convert the original 1D signal into 2D images and then directly extract features from the latter.

In sum, although existing research provides many intelligent solutions for fault diagnosis, shortcomings exist, especially in the following aspects:

1) Most fault diagnosis models based on traditional shallow machine learning relies on manually extracted features, which cannot be applied into equipment fault diagnosis in the era of big data due to their structural limitations.

2) The mere model input of time-series data or frequency-domain data would lead to insufficient extraction of spatiotemporal fault features.

3) Existing fault diagnosis model combining CNN and LSTM utilizes the output of one network as the input of the following network, features extracted from CNN will cause negative effects on that extracted from LSTM, and CNN cannot adjust the deviation resulting from the output of LSTM under the cascade network.

In terms of the limitations above, to adequately extract the failure features and improve the performance while avoiding overfitting, this paper constructs a shallow neural network with parallel feature extraction from time and frequency domains, which can extract local frequency distribution features and time-correlated features from the vibration signal. The shallow structure has fewer network parameters, which could quickly stabilize and effectively converge. The major contributions of this paper are summarized as follows.

1) A dual-input fault diagnosis model with shallow structure is built based on CNN and LSTM with the inputs of vibration signals and time-frequency images simultaneously. The model could extract time-correlated and frequency distribution features of the signals in parallel from time and frequency domains to improve the quality of the features.

2) The features obtained from the parallel network are fused to form more comprehensive features that can further enhance the representation ability of the model, thus improving its robustness

and generalization.

3) Experiments demonstrate that the model proposed provides excellent robustness and generalizability.

The rest of this paper is organized as follows: Section 2 briefly summarizes the related theoretical methods; Section 3 introduces the diagnostic process and model framework; Section 4 describes the experimental results; and conclusions are made in Section 5.

## 2. Theoretical methods

### 2.1. CNN

Convolutional neural network (CNN) is a feed-forward neural network with distinguishing features of weight sharing, local connectivity, and pooling. The core idea of CNN is to extract features through convolutional and pooling layers, then input the them into a fully connected layer for classification or regression tasks. A typical CNN usually includes input, convolutional, pooling, fully connected, and output layers.

#### 2.1.1. Convolutional layers

The convolutional layer could extract feature by convolving the input with kernels. The convolutional kernel updates weights by learning, and convolutional kernels with different sizes can obtain different outputs. The computation of the convolutional layer is as follows:

$$x_j^{l+1} = f\left(\sum_{i=1}^N x_i^l * w_{ij}^{l+1} + b_{ij}^{l+1}\right) \quad (1)$$

where  $x_j^l$  denotes the feature map of the  $j$ -th of the  $l$ -th convolutional layer,  $*$  represents the convolution operation,  $w_{ij}^l$  and  $b_{ij}^{l+1}$  are the weights and biases of the  $j$ -th convolutional kernel in the  $l$ -th convolutional layer,  $f$  is the activation function which is commonly ReLU function.

#### 2.1.2. Pooling layers

The pooling layer can reduce the size of the feature map, extract salient features, and reduce the sensitivity of the model to the input data. Maximum pooling is applied in this paper, whose operation can be expressed as follows:

$$\hat{x}_j^l = \max(x_j^l: x_{j+r}^l) \quad (2)$$

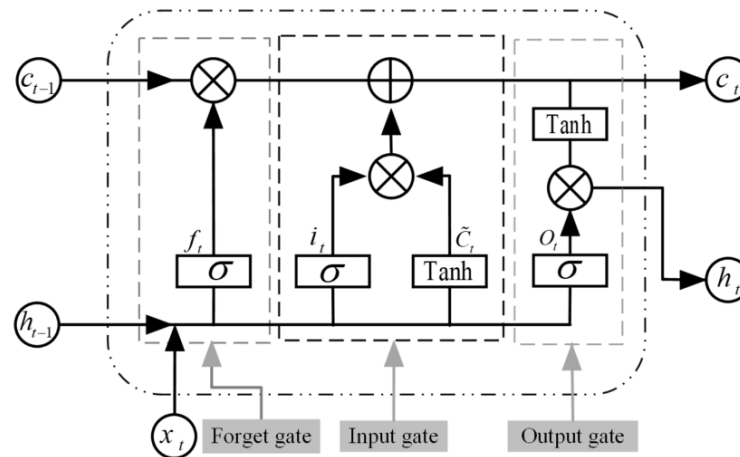
#### 2.1.3. Fully connected layers

The fully connected layer could effectively integrate and transform the features extracted from the previous layer to improve expressive ability and performance of the network. The classification results are obtained by the softmax function.



## 2.2. LSTM

Long short-term memory network (LSTM) introduces memory block and gate mechanisms to improve feature extraction of the input. The gate unit in LSTM is responsible for controlling and regulating the flow of information, so as to realize selective memorizing and forgetting of important information in the sequence, which addresses the gradient vanishing or explosion issues in traditional RNNs. Figure 1 illustrates a common schematic of an LSTM unit.



**Figure 1.** Common schematic of an LSTM unit.

Forget gate determines which information would be forgotten from the memory state of the previous time step, and is implemented by a sigmoid function. It is calculated as follows:

$$f_t = \sigma(w_{fx}x_t + w_{fh}h_{t-1} + b_f) \quad (3)$$

The input gate serves to control the extent to which new input information updates the current time-step memory state, and it is calculated as follows:

$$i_t = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \quad (4)$$

$$\tilde{c}_t = \tanh(w_{cx}x_t + w_{ch}h_{t-1} + b_c) \quad (5)$$

The output gate is a gating unit which controls the flowing of information from the current state to the output layer. Selectively outputting information in the hidden state at current time step could pass useful, long-term dependencies to the final layer. Its calculation is as follows:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (6)$$

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \quad (7)$$

$$h_t = o_t \times \tanh(c_t) \quad (8)$$

where  $\sigma$  is a sigmoid function,  $w$  is a weight matrix,  $x_t$  is the input of current time step,  $h_{t-1}$  is the short-term state of the last cell, and  $b$  is the bias vector.

### 2.3. CWT

Continuous wavelet transform (CWT) is a technique for analyzing the time-frequency features of non-smooth signal, which can capture transient changes and local structure in the signal. It can provide a time window that changes with frequency and perform adaptive multiscale analysis on signals through scaling and translation transformation. Its mathematical model is formulated as follows:

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x_t \varphi\left(\frac{t-b}{a}\right) dt \quad (9)$$

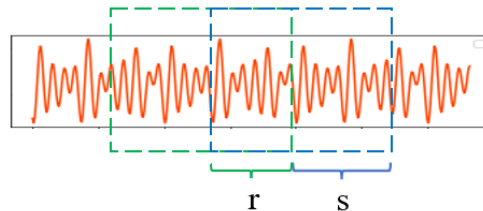
where  $a$  is the scale parameter,  $b$  is the translational value,  $\varphi(t)$  is the mother-wavelet function,  $x_t$  is the input.

### 2.4. Resampling

Deep learning requires a large number of samples for model training. Model tends to be underfitting when the training data are insufficient. Data augmentation technology could expand sample size which lays the foundation of a good model performance. Data augmentation method employed in this paper is resampling. It is performed by using a window width of  $w$  along the time axis. Resampling can be expressed as Eq (10).

$$w = r + s \quad (10)$$

where  $w$  is the window width,  $r$  is rate of repetition samples,  $s$  is the step size of sliding window. The process is shown in Figure 2.



**Figure 2.** Resampling process.

### 2.5. SNR

In signal processing, signal-to-noise ratio (SNR) refers to the ratio of signal power to noise power. It generally indicates how much the noise affects the desired signal. Its calculation formula is as follows:

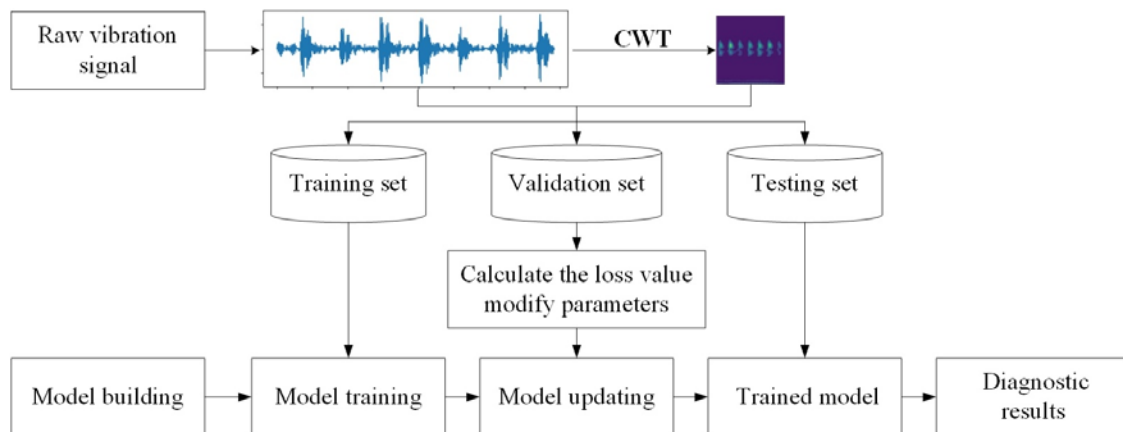
$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} = 20 \log_{10} \frac{A_{signal}}{A_{noise}} \quad (11)$$

where  $A_{signal}$  is the signal amplitude,  $A_{noise}$  is the noise amplitude,  $P_{signal}$  denotes signal power,  $P_{noise}$  denotes noise power. According to Eq (11), the smaller the SNR value, the stronger the noise compared with the signal, indicating the poorer the signal quality.

### 3. Proposed method

#### 3.1. The diagnostic procedure of the proposed model

For nonlinear vibration signals, 1D time-domain signals fail to adequately represent all fault features, and traditional deep learning methods are unable to sufficiently extract the frequency components and time-correlated features. Stacking of neural networks might lead to model overfitting and increase training time. Compared with time domain signal, time-frequency images can represent more features of non-stationary signals, and reveal the complex signal distribution of local impacts. In this paper, based on multi-domain learning, a fault diagnosis model based on parallel feature extraction with CNN and LSTM is proposed, aiming at extracting features from both time and frequency domains. Furthermore, the proposed model exhibits satisfactory accuracy and robustness with a shallow network structure. The diagnostic procedure of the model is illustrated in Figure 3.



**Figure 3.** The diagnostic procedure of the proposed model.

#### 3.2. Framework of the diagnostic model

The proposed diagnostic model can extract frequency distribution and time-correlated features in parallel. Besides, gate structure in LSTM further enhances the generalization performance of the model. The framework of the proposed diagnostic model is shown in Figure 4. It mostly contains three sub-modules: Data pre-processing, feature extraction, and fault classification. Furthermore, these three stages are further divided into five steps in detail.

**Step 1:** Data pre-processing. The dual-path feature extraction network consists of two different networks which extract time-series features and spatial features of the signal, respectively. In data preprocessing, the raw vibration signals are sliced into short samples. Then, these samples are converted to time-frequency images through CWT.

**Step 2:** Dual-path feature extraction. At this stage, a dual-path feature extraction network is constructed, and 1D vibration data and 2D time-frequency images are viewed as the input of this step: (i) 1D vibration data are input into the LSTM network for time feature extraction. (ii) time-frequency images are input into 2D-CNN for local spatial feature extraction.

**Step 3:** Feature fusion. The temporal and spatial features extracted are passed to the fusion layer

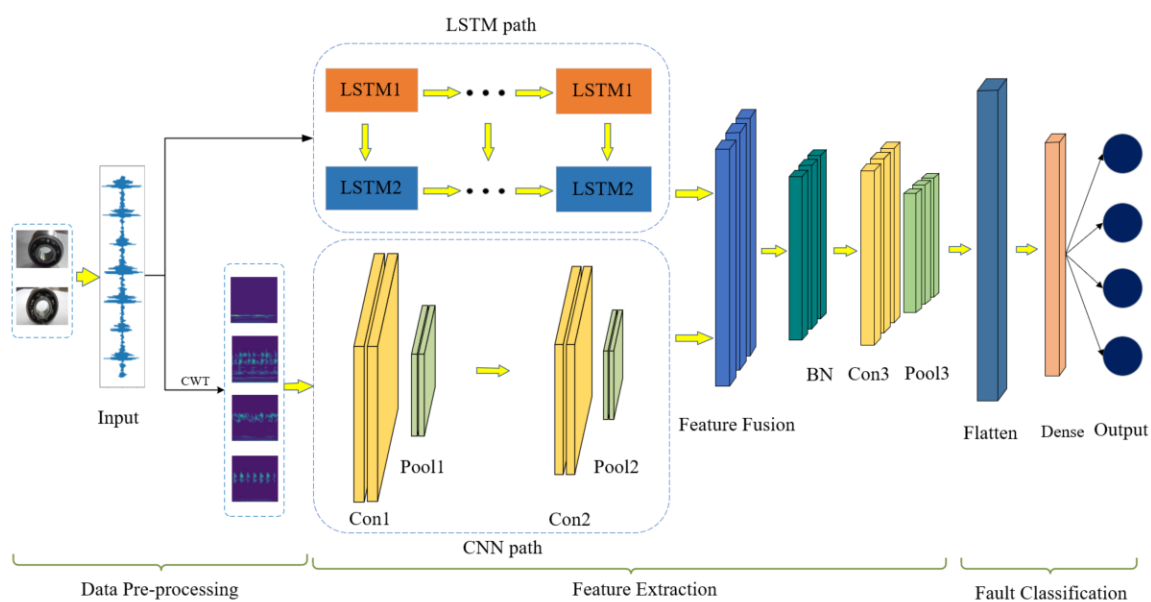
for feature fusion whose results can represent more comprehensively global features of the signal, thus improving the comprehensiveness of extracted features. The fusion process can be expressed as follows:

$$x = \text{concatenate}(\rho_1, \rho_2) \quad (12)$$

where  $\rho_1$ ,  $\rho_2$  denote the extracted feature vectors of LSTM and 2D-CNN,  $x$  denotes the fused feature vector,  $\text{concatenate}(\cdot)$  denotes the fusion process of 1D vector.

**Step 4:** Comprehensive feature extraction. A Batch Normalization (BN) layer is embedded which can be applied to any neural network layer to reduce the sample differences between layers by reducing the internal covariance migration [47]. The fused features are input to 1D-CNN for integrated feature extraction.

**Step 5:** Fault identification. Fault type is obtained by calculating the probability through the softmax function.



**Figure 4.** The framework of the proposed diagnostic model.

### 3.3. Model parameters

In the experiment, the feature extraction network consists of a two-layer LSTM, a two-layer 2D-CNN, and a one-layer 1D convolutional block. Fault classification is comprised of a flatten layer, two dense layers. The dropout method and mini-batch method are applied to eliminate overfitting as much as possible. The dropout rate is 0.5, the batch size is 128, and the epoch is 100. Model parameters are updated through Adam optimizer. Main parameters in the model are present in Table 1.

In fault diagnosis, accuracy, precision, recall, and F1-score can be taken as model evaluation standards:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (13)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

$$Recall = \frac{TP}{TP+FN} \quad (15)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

where TP, TN, FP and FN refer to true positives, true negatives, false positives and false negatives, respectively.

**Table 1.** Model parameters.

Layer name	Output size	Kernel size/Stride	Padding	Kernel number	Activation function	Dropout	BN
Input 1	(None, 49, 16)	—	—	—	—	—	No
Input 2	(None, 52, 52)	—	—	—	—	—	No
LSTM 1	(None, 49, 49)	49	—	—	Relu	—	No
LSTM 2	(None, 49)	49	—	—	Relu	—	No
Con 1	(None, 52, 52, 8)	$3 \times 3 / 1$	Same	8	Relu	—	No
Pool 1	(None, 26, 26, 8)	$2 \times 2 / 2$	Same	8	Relu	—	No
Con 2	(None, 26, 26, 16)	$3 \times 3 / 1$	Same	16	Relu	—	No
Pool 2	(None, 13, 13, 16)	$2 \times 2 / 2$	Same	16	Relu	—	No
Flatten1	(None, 2704)	—	—	—	—	—	No
Con 3	(None, 2753, 8)	$3 \times 1 / 1$	Same	8	Relu	—	Yes
Pool 3	(None, 1377, 8)	$2 \times 1 / 2$	Same	8	Relu	—	No
Flatten2	(None, 11016)	—	—	—	—	—	No
Dense 1	(None, 64)	64	—	—	—	0.5	No
Dense 2	(None, 10)	10	—	—	Softmax	—	No

#### 4. Experiments and result analysis

In this paper, two open access datasets, Case Western Reserve University (CWRU) and Jiangnan University (JNU) bearing datasets, are used to evaluate the effectiveness of the method.

##### 4.1. CWRU dataset experiment

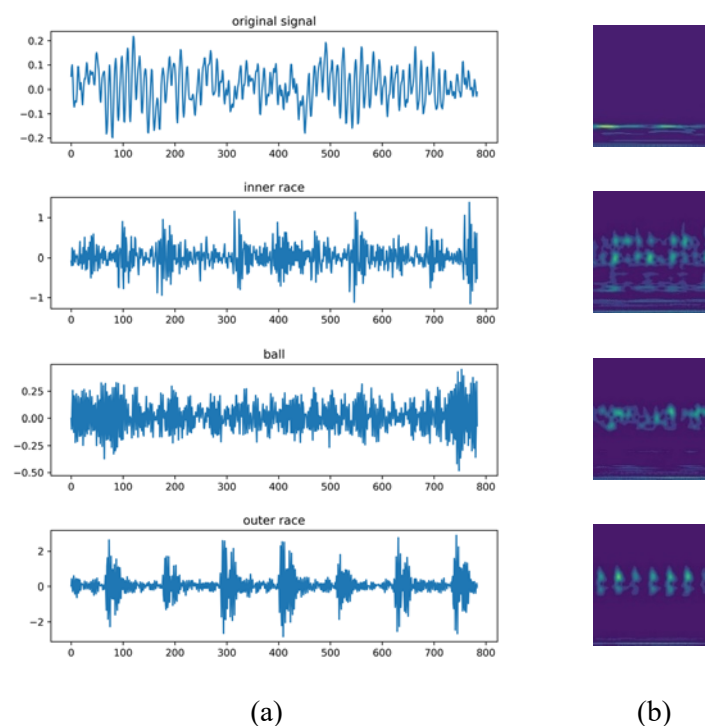
CWRU dataset, a standard bearing dataset, was collected from the Simulated Bearing Experimental Platform at Case Western Reserve University. The signals were acquired by acceleration sensors installed on the drive end housing of the motor (2 hp) and contained four loads, with a sampling frequency of 12 kHz. Single-point damage of different diameters was introduced through the EDM technique, and bearing failures were categorized as inner ring damage (ID), ball damage (BD) and outer ring damage (OD). Table 2 summarizes the features of CWRU bearing dataset. Each dataset consists of nine fault samples with different characteristics and one normal sample. The number of data samples for each type is 600 and each data sample contains 784 data points. There are a total of 6000 samples with 10 fault types, which are divided into 4200 training samples, 900 validation samples, and 900 testing samples.

**Table 2.** CWRU bearing dataset.

Label	Fault status (inch)	Load A (HP)	Load B (HP)	Load C (HP)	Load D (HP)
0	Normal	0	1	2	3
1	0.007 ID	0	1	2	3
2	0.014 ID	0	1	2	3
3	0.021 ID	0	1	2	3
4	0.007 OD	0	1	2	3
5	0.014 OD	0	1	2	3
6	0.021 OD	0	1	2	3
7	0.007 BD	0	1	2	3
8	0.014 BD	0	1	2	3
9	0.021 BD	0	1	2	3

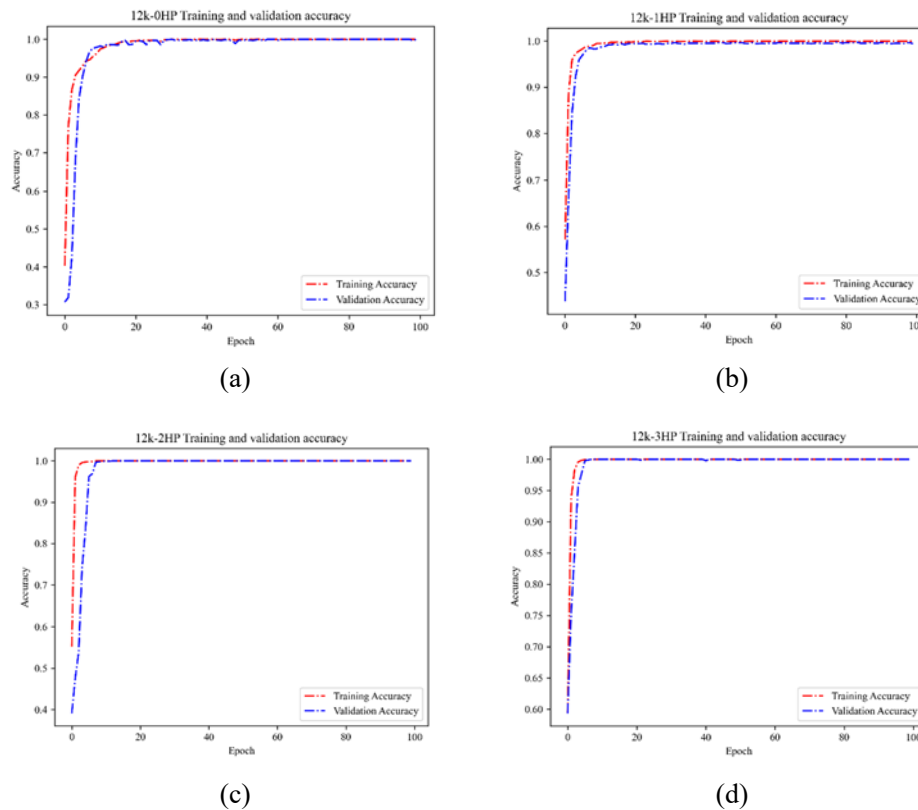
#### 4.1.1. Data pre-processing

The original signal is cut into sub-signal with a same size by the resampling method; then, time-frequency images are obtained through CWT. The time-domain and time-frequency images with the four fault types (original signal, inner race, ball, outer race) in CWRU dataset are shown in Figure 5. Figure 5(a) exhibits time-domain waveform of the signal, while Figure 5(b) exhibits their time-frequency images after CWT.

**Figure 5.** Time-domain and time-frequency images. (a) Time-domain. (b) Time-frequency.

#### 4.1.2. Performance under constant load

The convergence and robustness of the model are investigated under four constant loads (0, 1, 2 and 3 HP), and the convergence status and loss function of model training process is shown in Figures 6 and 7. According to the accuracy of the four loads in Figure 6, the proposed model can converge quickly and stably to 1 under all loads. Furthermore, the loss converges to 0 steadily during model training in Figure 7. When the epoch is 20, the loss values remain close under four loads. The above results show that the model has satisfactory robustness under constant loads.

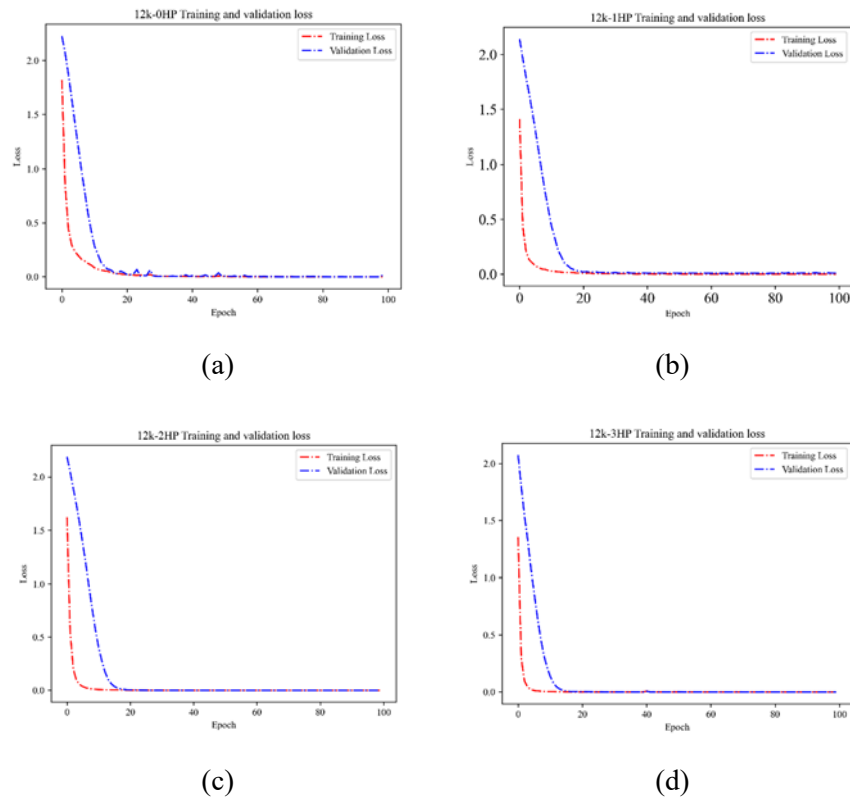


**Figure 6.** Accuracy of the training and validation. (a) 0 HP. (b) 1 HP. (c) 2 HP. (d) 3 HP.

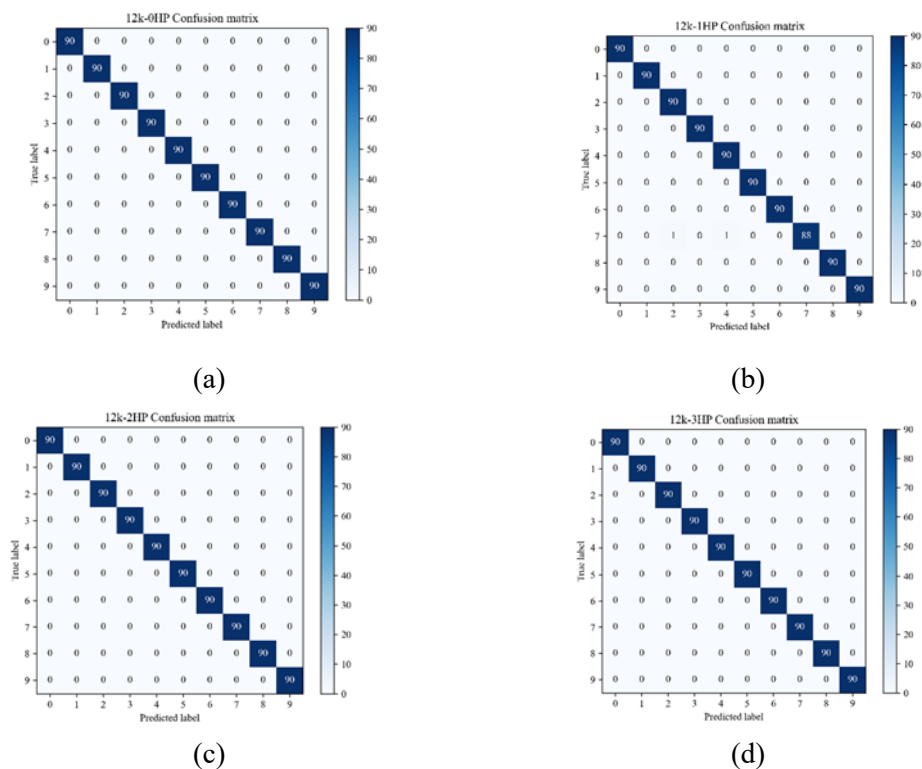
Figure 8 displays confusion matrix under four loads (0, 1, 2 and 3 HP) for the testing set, and the accuracies of the model in four different loads are shown in Table 3. According to Table 3, the model can achieve 99.78% diagnostic accuracy at least. To extensively analyze the result under 1HP load whose accuracy is the lowest, the precision, recall, and F1-score of the model under 1HP load are shown in Table 4. These specific details combining with the former results demonstrate the high diagnostic accuracy under constant load.

**Table 3.** Accuracy under four loads

Loads (HP)	0	1	2	3
Accuracy (%)	100	99.78	100	100



**Figure 7.** Loss of the training and validation. (a) 0HP, (b) 1HP, (c) 2HP, (d) 3HP.



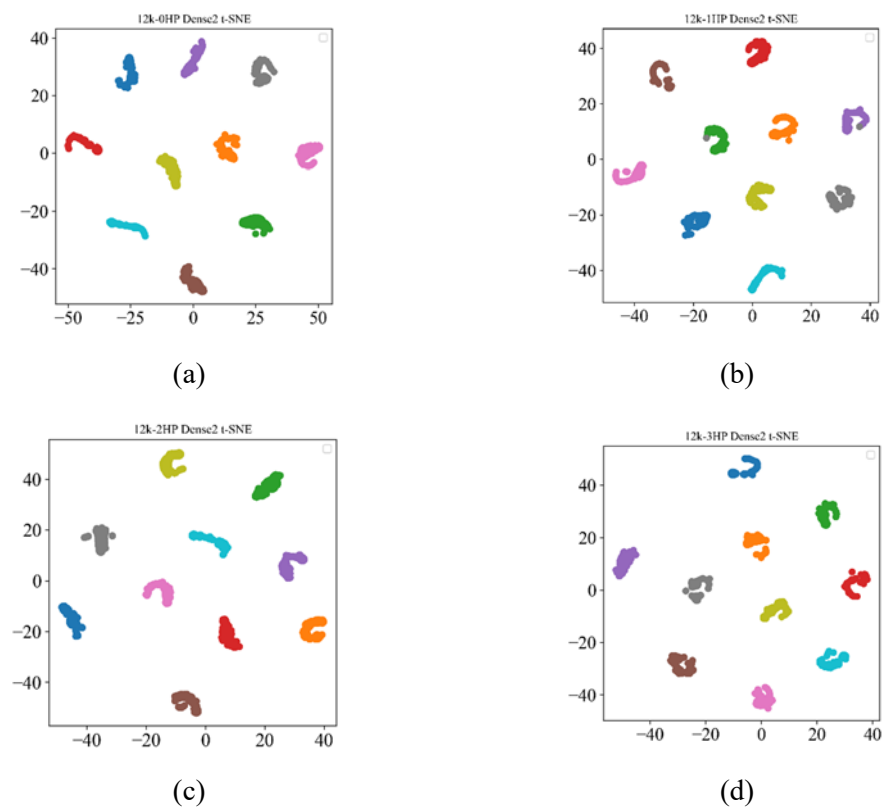
**Figure 8.** Confusion matrix of testing set. (a) 0HP, (b) 1HP, (c) 2HP, (d) 3HP.



**Table 4.** Diagnostic results under the 1HP load.

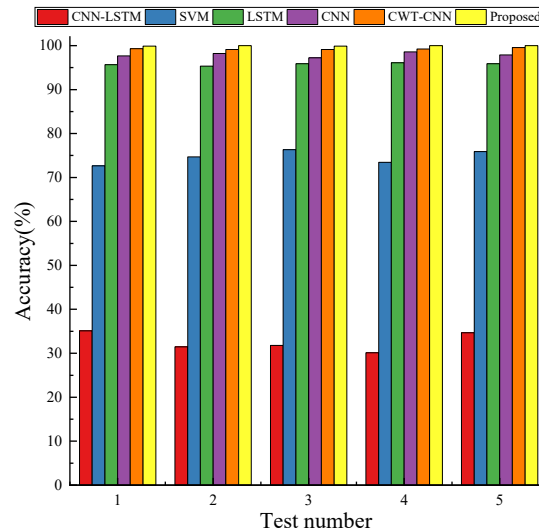
Fault label	Precision	Recall	F1-score
0	1	1	1
1	1	1	1
2	0.9890	1	0.9945
3	1	1	1
4	0.9890	1	0.9945
5	1	1	1
6	1	1	1
7	1	0.9778	0.9888
8	1	1	1
9	1	1	1

t-SNE method can effectively map features from high-dimensional to low-dimensional space. The final diagnostic results of the Dense 2 layer of testing set were visualized via t-SNE in Figure 9. Obviously, the proposed model shows distinct clusters with clear classification boundaries.

**Figure 9.** Visualization for the Dense 2 layer under four loads. (a) 0HP, (b) 1HP, (c) 2HP, (d) 3HP.

To demonstrate the superiority of the proposed model, five comparison models are selected, and the performances of all these six models are evaluated on the same testing dataset. Moreover, 5-fold cross-validation is adopted to ensure experimental reliability. Among them, CNN-LSTM is a single-path serial fault diagnosis model. Except SVM, major parameters of the other four comparison models

are almost the same as those of the proposed model. In addition, the input of CWT-CNN is the time-frequency image, and the other four models are vibration signals. Accuracy results are depicted in Figure 10.



**Figure 10.** Accuracy of comparison models.

Based on the results in Figure 10, the accuracy of CNN-LSTM is the lowest, that of CNN reaches 97%, and that of CWT-CNN reaches 99%, indicating that the time-frequency image is competent to represent fault features compared with the time-domain signal. The parallel mechanism is adopted to extract features from local spatial information and temporal dependencies of the signal in dual-domain, and the result shows that the diagnostic accuracy of the proposed model is superior to the other five models, indicating the effectiveness of the parallel feature extraction network. For the feature extraction network with CNN and LSTM connected in series, the spatial distribution features are extracted by CNN and then the time-correlated features are extracted by LSTM. Although such model has considered the periodicity of the signal frequency distribution, CNN corrupts the temporal features of the original data, resulting in the loss of important temporal features, which would affect the quality of final extracted features. The parallel feature extraction network proposed in this paper can significantly improve the quality of the extracted features by making up for the shortcomings of the series feature extraction mechanism.

#### 4.1.3. Performance with variable loads

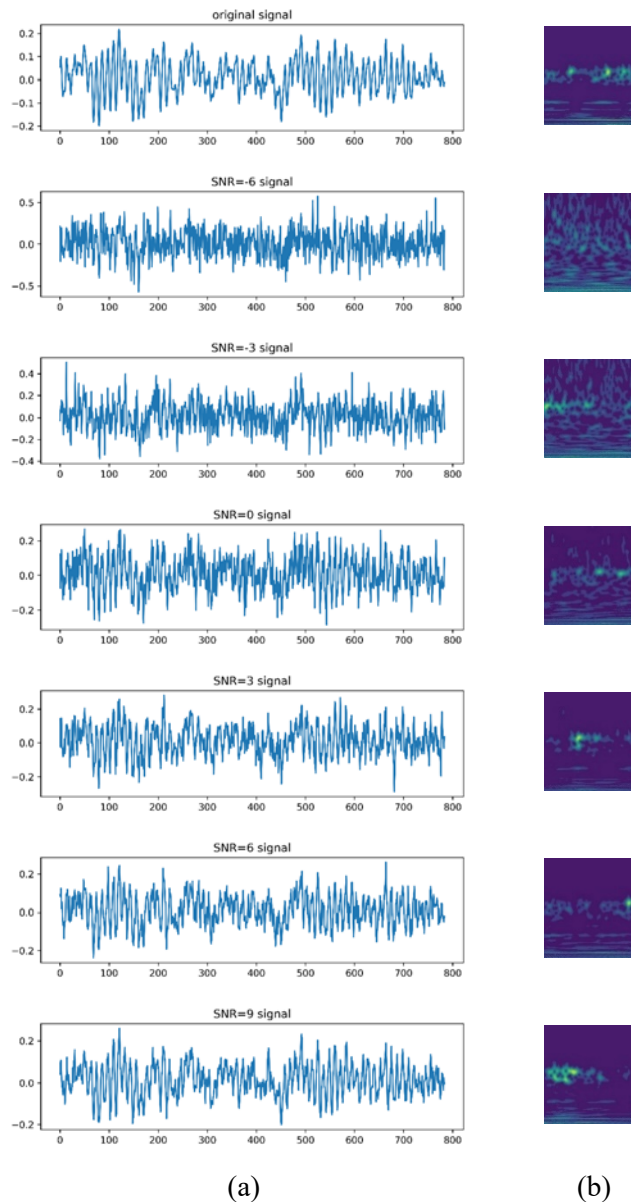
Based on CWRU dataset, composite datasets under mixed load were constructed for experiment. The dataset name implies its components (e.g., 12HP means the load contains 1 and 2HP). Composite datasets are tested on the proposed model, and their results are displayed in Table 5.

Based on Table 5, the diagnostic results of proposed model achieve satisfactory diagnostic accuracy on the composite dataset, indicating that the proposed model can be applied to variable load conditions. Therefore, the proposed model has excellent adaptability, and it is suitable for variable load conditions.

**Table 5.** Diagnostic accuracy with composite datasets.

Composite dataset (HP)	01	02	03	12	13	23	012	013	023	123	0123
Accuracy (%)	99.56	99.78	99.89	99.78	99.89	99.44	98.44	100	99.56	97.89	99.44

#### 4.1.4. Noise immunity analysis

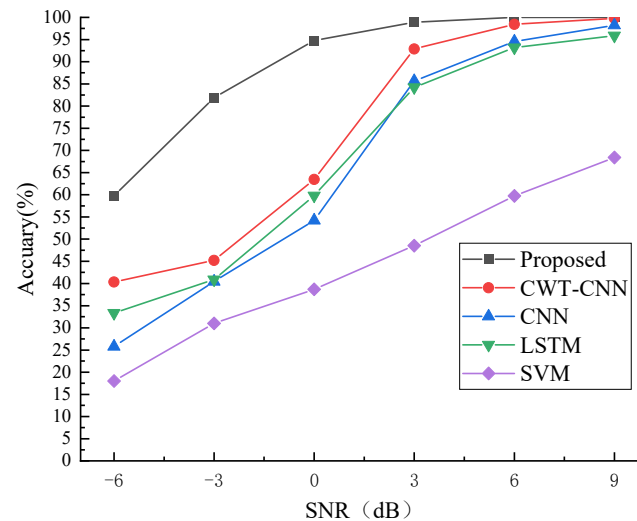


**Figure 11.** Time-domain and time-frequency images under different SNRs. (a) Time-domain. (b) Time-frequency.

Real working contexts of the bearing are usually harsh, and its vibration signal is inevitably mixed with some noises. To simulate the signals obtained in the interference of noises, Gaussian white noises with different SNRs (-6, -3, 0, 3, 6 and 9 dB) are added to the original signal, respectively. Figure 11 shows the time domain and the frequency domain images of the original signal as well as the signals with noise added at different SNRs (-6, -3, 0, 3, 6, 9). Noise is added to the bearing vibration signal,

making it challenging to distinguish certain features within the signal.

Comparison experiments of model accuracy with different SNRs were designed and their results are summarized in Figure 12. They show that noise does affect the diagnostic accuracy of the model, with the more severe the noise interference, the lower the accuracy. However, the proposed model demonstrates the highest accuracy under different SNRs compared with other models.



**Figure 12.** Diagnostic accuracy of different models under different SNRs.

#### 4.2. JNU dataset experiment

To test the generalizability of the proposed model, a testing experiment was performed on JNU bearing dataset. Data pre-processing here is the same with that on CWRU dataset.

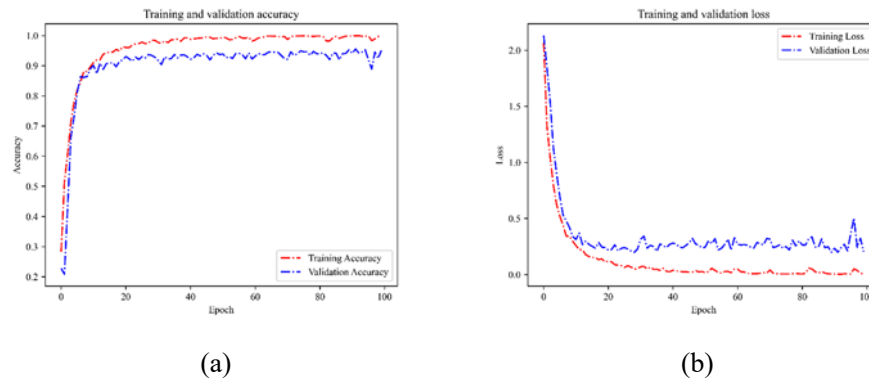
JNU bearing dataset is provided by Jiangnan University. It consists of three bearing vibration sub-datasets at different rotational speeds with the data acquisition frequency of 50 kHz. The sub-dataset contains one normal state and three fault states: ID, BD and OD. Based on differences in working conditions, 12 fault types are classified. The features of JNU bearing dataset are summarized in Table 6.

**Table 6.** JNU bearing dataset.

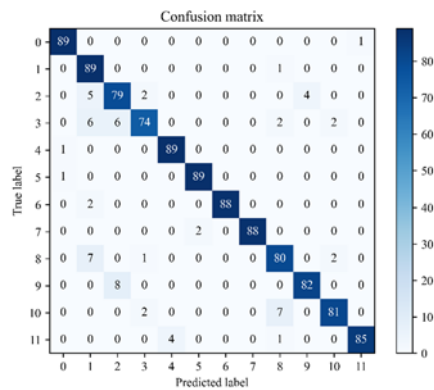
Label	Fault status	Rotating speed (rpm)	Label	Fault status	Rotating speed (rpm)
0	Normal 1	600	6	OD 1	600
1	Normal 2	800	7	OD 2	800
2	Normal 3	1000	8	OD 3	1000
3	ID 1	600	9	BD 1	600
4	ID 2	800	10	BD 2	800
5	ID 3	1000	11	BD 3	1000

Figure 13(a),(b) shows the accuracy and loss function of the model training process, where the accuracy on both the training and validation sets gradually converges and the loss function narrows down

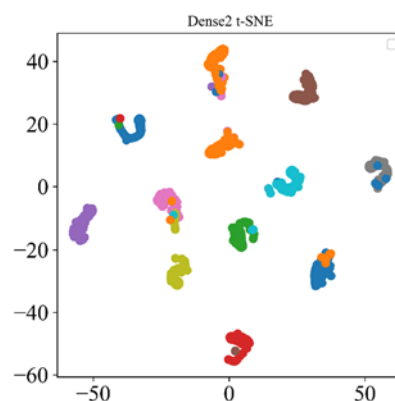
to a smaller range. Overall, the proposed model shows favorable performance on JNU bearing dataset.



**Figure 13.** Accuracy and loss on JNU dataset. (a) Accuracy. (b) Loss.



**Figure 14.** Confusion matrix of testing set.

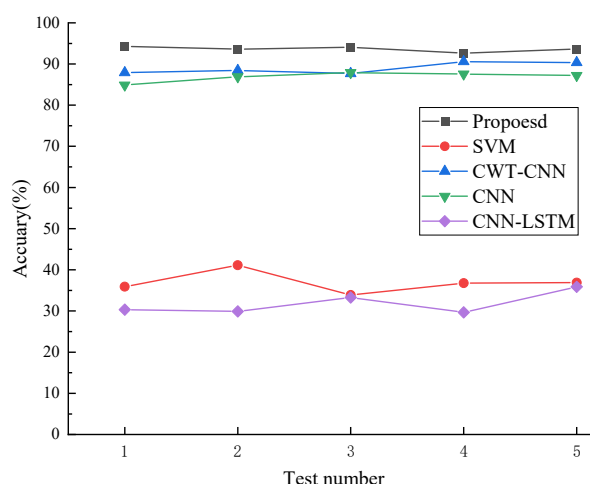


**Figure 15.** Classification results of testing set.

Confusion matrix for the testing data is presented in Figure 14. The accuracy of the proposed

model is 93.80%. Then, the classification result of testing set was visualized via t-SNE, and shown in Figure 15. Based on Figure 15, bearing failures with different types have clear classification boundaries. However, the proposed model has low accuracy for fault types with true labels of 2, 3 and 8, but has high fault identification accuracy for most fault types. As for true label of 2, 3 and 8, there are similarities between time-frequency images, which may in turn lead to misdiagnosis.

In addition, a comparison experiment is conducted, and the accuracy results is shown in Figure 16. They demonstrate that the proposed model on JNU dataset offers higher fault classification accuracy than other four models.



**Figure 16.** Accuracy of different models on JNU dataset.

In a sum, comparative analysis of the results from CWRU dataset and JNU dataset suggests that the proposed model exhibits superior generalizability on different datasets.

## 5. Conclusions

Bearing vibration signals are dynamic and periodic data. Traditional shallow machine learning and single deep learning methods can hardly extract temporal and spatial features at the same time. As for those deep learning-based single-path fault diagnosis methods, feature extraction of one stage will affect that of its next stage, which would affect the final diagnostic accuracy. To tackle the insufficient feature extraction in bearing fault diagnosis, a parallel feature extraction mechanism is proposed, which can extract time-correlated information and frequency spatial distribution features in parallel; this applies CWT, combined with CNN and LSTM, to bearing fault diagnosis. First, the signal is sliced into equal-sized sub-signals via a resampling technique. Then, the CWT converts the sub-signal into a time-frequency image, which, combined with the sub-signal, are viewed as two inputs of the proposed model. Second, LSTM and 2D-CNN dual-path network are employed to extract the temporal and spatial features in parallel. After the fusion of extracted temporal and spatial features, 1D-CNN is utilized to obtain the comprehensive features. Finally, the softmax function is implemented for fault classification. In addition, a BN layer is embedded after the feature fusion layer and mini-batch method is applied to reduce the computational complexity. On CWRU dataset, the model presents satisfactory accuracy and robustness, and performs well under complex conditions under variable loads and SNRs.

Moreover, the generalization performance of the model is further validated on JNU dataset. In sum, a novel fault diagnostic model is proposed which achieves satisfactory diagnostic accuracy under constant and variable loads. However, it is unable to obtain satisfactory results under strong noise, and de-noise could be one of the future research fields.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62176150).

### Conflict of interest

The authors declare that there are no conflicts of interest.

### References

1. D. T. Hoang, H. J. Kang, A survey on deep learning based bearing fault diagnosis, *Neurocomputing*, **335** (2019), 327–335. <https://doi.org/10.1016/j.neucom.2018.06.078>
2. M. Zeng, W. Zhang, Z. Chen, Group-based K-SVD denoising for bearing fault diagnosis, *IEEE Sens. J.*, **19** (2019), 6335–6343. <https://doi.org/10.1109/JSEN.2019.2910868>
3. X. Yan, D. She, Y. Xu, Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions, *Expert Syst. Appl.*, **216** (2023), 119479. <https://doi.org/10.1016/j.eswa.2022.119479>
4. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Math. Phys. Eng. Sci.*, **454** (1998), 903–995. <https://doi.org/10.1098/rspa.1998.0193>
5. K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.*, **62** (2013), 531–544. <https://doi.org/10.1109/TSP.2013.2288675>
6. G. Xin, Z. Li, L. Jia, Q. Zhong, H. Dong, N. Hamzaoui, et al., Fault diagnosis of wheelset bearings in high-speed trains using logarithmic short-time Fourier transform and modified self-calibrated residual network, *IEEE Trans. Ind. Inf.*, **18** (2021), 7285–7295. <https://doi.org/10.1109/TIM.2021.3139706>
7. I. Daubechies, *Ten Lectures on Wavelets*, Society for industrial and applied mathematics, 1992.
8. M. Q. Tran, M. K. Liu, Q. V. Tran, T. K. Nguyen, Effective fault diagnosis based on wavelet and convolutional attention neural network for induction motors, *IEEE Trans. Instrum. Meas.*, **71** (2021), 1–13. <https://doi.org/10.1109/TIM.2021.3139706>
9. L. Yuan, D. Lian, X. Kang, Y. Chen, K. Zhai, Rolling bearing fault diagnosis based on convolutional neural network and support vector machine, *IEEE Access*, **8** (2020), 137395–137406. <https://doi.org/10.1109/ACCESS.2020.3012053>

10. T. Jin, Q. Cheng, H. Chen, S. Wang, J. Guo, C. Chen, Fault diagnosis of rotating machines based on EEMD-MPE and GA-BP, *Int. J. Adv. Manuf. Technol.*, (2021), 1-12. <https://doi.org/10.1007/s00170-021-08159-z>
11. P. Chen, X. Zhao, Q. Zhu, A novel classification method based on ICGOA-KELM for fault diagnosis of rolling bearing, *Appl. Intell.*, **50** (2020), 2833–2847.
12. H. Tao, P. Wang, Y. Chen, V. Stojanovic, H. Yang, An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks, *J. Franklin Inst.*, **357** (2020), 7286–7307. <https://doi.org/10.1007/s10489-020-01684-6>
13. J. Gu, Y. Peng, H. Lu, X. Chang, G. Chen, A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN, *Measurement*, **200** (2022), 111635. <https://doi.org/10.1016/j.measurement.2022.111635>
14. N. Saravanan, V. K. Siddabattuni, K. Ramachandran, A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box, *Expert Syst. Appl.*, **35** (2008), 1351–1366. <https://doi.org/10.1016/j.eswa.2007.08.026>
15. V. Muralidharan, V. Sugumaran, A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis, *Appl. Soft Comput.*, **12** (2012), 2023–2029. <https://doi.org/10.1016/j.asoc.2012.03.021>
16. J. Li, X. Yao, X. Wang, Q. Yu, Y. Zhang, Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis, *Measurement*, **153** (2020), 107419. <https://doi.org/10.1016/j.measurement.2019.107419>
17. H. Li, T. Liu, X. Wu, Q. Chen, Application of EEMD and improved frequency band entropy in bearing fault feature extraction, *ISA Trans.*, **88** (2019), 170–185. <https://doi.org/10.1016/j.isatra.2018.12.002>
18. H. S. Najim, J. K. Alsalaet, Cross-domain diagnosis of roller bearing faults based on the envelope analysis adaptive features and artificial neural networks, *J. Vib. Control*, **2023** (2023), 10775463231191684. <https://doi.org/10.1177/10775463231191684>
19. J. Yang, J. Liu, J. Xie, C. Wang, T. Ding, Conditional GAN and 2-D CNN for bearing fault diagnosis with small samples, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–12. <https://doi.org/10.1109/tim.2021.3119135>
20. X. Song, Y. Cong, Y. Song, Y. Chen, P. Liang, A bearing fault diagnosis model based on CNN with wide convolution kernels, *J. Ambient Intell. Hum. Comput.*, **13** (2022), 4041–4056. <https://doi.org/10.1007/s12652-021-03177-x>
21. J. Zhu, T. Hu, B. Jiang, X. Yang, Intelligent bearing fault diagnosis using PCA-DBN framework, *Neural Comput. Appl.*, **32** (2020), 10773–10781. <https://doi.org/10.1007/s00521-019-04612-z>
22. P. Zou, B. Hou, J. Lei, Z. Zhang, Bearing fault diagnosis method based on EEMD and LSTM, *Int. J. Comput. Commun.*, **15** (2020). <http://doi.org/10.15837/ijccc.2020.1.3780>
23. X. Yan, D. She, Y. Xu, M. Jia, Deep regularized variational autoencoder for intelligent fault diagnosis of rotor-bearing system within entire life-cycle process, *Knowl.-Based Syst.*, **226** (2021), 107142.
24. K. Tong, Y. Wu, F. Zhou, Recent advances in small object detection based on deep learning: A review, *Image Vis. Comput.*, **97** (2020), 103910. <https://doi.org/10.1016/j.imavis.2020.103910>
25. G. Algan, I. Ulusoy, Image classification with deep learning in the presence of noisy labels: A survey, *Knowl.-Based Syst.*, **215** (2021), 106771. <https://doi.org/10.1016/j.knosys.2021.106771>



26. M. Morchid, Parsimonious memory unit for recurrent neural networks with application to natural language processing, *Neurocomputing*, **314** (2018), 48–64. <https://doi.org/10.1016/j.neucom.2018.05.081>
27. Y. Ma, X. Jia, H. Bai, G. Liu, G. Wang, C. Guo, et al., A new fault diagnosis method based on convolutional neural network and compressive sensing, *J. Mech. Sci. Technol.*, **33** (2019), 5177–5188. <https://doi.org/10.1007/s12206-019-1007-5>
28. C. Lu, Z. Wang, B. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification, *Adv. Eng. Inf.*, **32** (2017), 139–151. <https://doi.org/10.1016/j.aei.2017.02.005>
29. H. Liu, D. Yao, J. Yang, X. Li, Lightweight convolutional neural network and its application in rolling bearing fault diagnosis under variable working conditions, *Sensors*, **19** (2019), 4827. <https://doi.org/10.3390/s19224827>
30. Y. Jin, C. Chen, S. Zhao, Multisource data fusion diagnosis method of rolling bearings based on improved multiscale CNN, *J. Sensors*, **2021** (2021), 1–17. <https://doi.org/10.1155/2021/2251530>
31. Z. Xing, R. Zhao, Y. Wu, T. He, Intelligent fault diagnosis of rolling bearing based on novel CNN model considering data imbalance, *Appl. Intell.*, **52** (2022), 16281–16293. <https://doi.org/10.1007/s10489-022-03196-x>
32. C. Zhang, J. Feng, C. Hu, Z. Liu, L. Cheng, Y. Zhou, An intelligent fault diagnosis method of rolling bearing under variable working loads using 1-D stacked dilated convolutional neural network, *IEEE Access*, **8** (2020), 63027–63042. <https://doi.org/10.1109/ACCESS.2020.2981289>
33. M. Ye, X. Yan, N. Chen, M. Jia, Intelligent fault diagnosis of rolling bearing using variational mode extraction and improved one-dimensional convolutional neural network, *Appl. Acoust.*, **202** (2023), 109143.
34. D. Gao, Y. Zhu, Z. Ren, K. Yan, W. Kang, A novel weak fault diagnosis method for rolling bearings based on LSTM considering quasi-periodicity, *Knowl.-Based Syst.*, **231** (2021), 107413. <https://doi.org/10.1016/j.apacoust.2022.109143>
35. H. Liu, J. Zhou, Y. Zheng, W. Jiang, Y. Zhang, Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders, *ISA Trans.*, **77** (2018), 167–178. <https://doi.org/10.1016/j.isatra.2018.04.005>
36. B. An, Z. Zhao, S. Wang, S. Chen, X. Chen, Sparsity-assisted bearing fault diagnosis using multiscale period group lasso, *ISA Trans.*, **98** (2020), 338–348. <https://doi.org/10.1016/j.isatra.2019.08.042>
37. M. Qiao, S. Yan, X. Tang, C. Xu, Deep convolutional and LSTM recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads, *IEEE Access*, **8** (2020), 66257–66269. <https://doi.org/10.1109/ACCESS.2020.2985617>
38. G. Fu, Q. Wei, Y. Yang, C. Li, Bearing fault diagnosis based on CNN-BiLSTM and residual module, *Meas. Sci. Technol.*, **34** (2023), 125050. <https://doi.org/10.1088/1361-6501/acf598>
39. H. Chen, W. Meng, Y. Li, Q. Xiong, An anti-noise fault diagnosis approach for rolling bearings based on multiscale CNN-LSTM and a deep residual learning model, *Meas. Sci. Technol.*, **34** (2023), 045013. <https://doi.org/10.1088/1361-6501/acf074>
40. S. Ning, Y. Wang, W. Cai, Z. Zhang, Y. Wu, Y. Ren, et al., Research on intelligent fault diagnosis of rolling bearing based on improved shufflenetV2-LSTM, *J. Sensors*, **2022** (2022). <https://doi.org/10.1088/1361-6501/acf598>

41. Y. Guo, J. Mao, M. Zhao, Rolling bearing fault diagnosis method based on attention CNN and BiLSTM network, *Neural Process. Lett.*, **55** (2023), 3377–3410. <https://doi.org/10.1007/s11063-022-11013-2>
42. M. Ye, X. Yan, N. Chen, Y. Liu, A robust multi-scale learning network with quasi-hyperbolic momentum-based Adam optimizer for bearing intelligent fault diagnosis under sample imbalance scenarios and strong noise environment, *Struct. Health Monit.*, (2023), 14759217231192363. <https://doi.org/10.1177/14759217231192363>
43. H. Wang, Z. Liu, D. Peng, Y. Qin, Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis, *IEEE Trans. Ind. Inf.*, **16** (2019), 5735–5745. <https://doi.org/10.1109/TII.2019.2955540>
44. S. Shao, S. Mcaleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, *IEEE Trans. Ind. Inf.*, **15** (2018), 2446–2455. <https://doi.org/10.1109/TII.2018.2864759>
45. J. Zhang, Y. Sun, L. Guo, H. Gao, X. Hong, H. Song, A new bearing fault diagnosis method based on modified convolutional neural networks, *Chin. J. Aeronaut.*, **33** (2020), 439–447. <https://doi.org/10.1016/j.cja.2019.07.011>
46. L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, *IEEE Trans. Ind. Electron.*, **65** (2017), 5990–5998. <https://doi.org/10.1109/TIE.2017.2774777>
47. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *International Conference on Machine Learning*, (2015), 448–456. <https://doi.org/10.48550/arXiv.1502.03167>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)