**Eksploatacja i Niezawodnosc – Maintenance and Reliability**

**Volume 25 (2023), Issue 1**

journal homepage: http://www.ein.org.pl

Article citation info:
Zhang J, Kong X, Cheng L, Qi H, Yu M, Intelligent fault diagnosis of rolling bearings based on continuous wavelet transform-multiscale feature fusion and improved channel attention mechanism. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2023: 25(1) http://doi.org/10.17531/ein.2023.1.16

# Intelligent fault diagnosis of rolling bearings based on continuous wavelet transform-multiscale feature fusion and improved channel attention mechanism

Indexed by: Web of Science Group

**Jiqiang Zhang[a], Xiangwei Kong[a,b,c*], Liu Cheng[a], Haochen Qi[a], Mingzhu Yu[d]**

[a] School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China
[b] Key Laboratory of Vibration and Control of Aero-Propulsion System, Ministry of Education, Northeastern University, Shenyang 110819, China
[c] Key Laboratory of Multidisciplinary Design Optimization of Complex Equipment, Northeastern University, Shenyang 110819, China
[d] Angang Steel Company Limited, Anshan 114021, China

## Highlights

- Multiscale 2-D convolutional neural networks with different sizes are proposed.
- The CWT analysis of vibration signals and deep learning methods are combined.
- An improved channel attention mechanism is developed.
- The model can be applied to single and advanced fault-diagnosing eventualities.
- The algorithm reduces the dependence on prior knowledge and manual labor.

## Abstract

Accurate fault diagnosis is critical to operating rotating machinery safely and efficiently. Traditional fault information description methods rely on experts to extract statistical features, which inevitably leads to the problem of information loss. As a result, this paper proposes an intelligent fault diagnosis of rolling bearings based on a continuous wavelet transform(CWT)-multiscale feature fusion and an improved channel attention mechanism. Different from traditional CNNs, CWT can convert the 1-D signals into 2-D images, and extract the wavelet power spectrum, which is conducive to model recognition. In this case, the multiscale feature fusion was implemented by the parallel 2-D convolutional neural networks to accomplish deeper feature fusion. Meanwhile, the channel attention mechanism is improved by converting from compressed to extended ways in the excitation block to better obtain the evaluation score of the channel. The proposed model has been validated using two bearing datasets, and the results show that it has excellent accuracy compared to existing methods.

## Keywords

deep learning, continuous wavelet transform, improved channel attention mechanism, multi-conditions, convolutional neural network

## 1. Introduction

Bearings play the most critical role in mechanical systems, with applications in aerospace, wind power generation, rail transit, automobiles, ships, and engineering machinery. When a serious failure occurs, it can result in equipment downtime or even casualties of various degrees. Rolling bearings are responsible for almost 60% of transmission system failures, according to insufficient figures [8]. As a result, intelligent diagnostic algorithms for detecting problems and precisely identifying them should be developed as soon as possible to reduce the economic loss caused by equipment failure.

Since the 21st century, with the development of computer and sensing technologies, a large amount of data has been collected and stored, and fault diagnosis techniques based on data-driven algorithms have developed rapidly [19, 36]. Recently, an increasing number of scholars have focused on the field and obtained refreshing results. Ravikumar et al. [23] proposed a fault diagnosis algorithm for internal combustion engines combined with discrete wavelet transform (DWT) and decision tree algorithms. Dhiman et al. [5] used twin support vector machines (TWSVMs) for fault diagnosis for wind turbines. Peng et al. [21] developed a 1-D convolutional neural network (CNN) for fault diagnosis, in which the 1-D signals of bearings are preprocessed into low-frequency branches, constant mapping branches, and denoising branches, and then these branches are fed into the CNN for fault diagnosis, which can effectively detect the faults of high-speed train bearings in a

---

(*) Corresponding author.
E-mail addresses: J. Zhang (ORCID: 0000-0002-2856-0929) freedomzhangneu@163.com, X. Kong (ORCID: 0000-0001-7042-4368), shawnkongneu@gmail.com, L. Cheng (ORCID: 0000-0002-7094-3835) lchengneu@l63.com, Qi. H (ORCID: 0000-0002-9648-2141) haochen.qi@outlook.com, M. Yu (ORCID: 0000-0002-5549-9422) niunianjixiangruyi@126.com.

high-noise environment. Peng et al. [22] gave a raw vibration signal analysis algorithm based on the improved Hilbert-Huang transform (IHHT), and the experiment has proven that IHHT does show the rubbing symptoms more clearly and accurately. Although artificial intelligence algorithms based on data-driven methods in the literature have achieved excellent performance in fault diagnosis, we found that the manually extracted features can only be applied to a single working condition, which is not completely free from the limits of linearity and smoothness and has poor self-adaptability.

Deep neural networks (DNNs) and encoders with fully connected layers offer a promising answer to the matter. DNN and encoders specialize in extracting more representative features from the vibration signal, which implies that feature extraction depends on less prior knowledge and skilled expertise. Many scholars have implemented fault diagnosis using encoders and DNNs. Wu et al. [11] proposed hybrid classification encoders based on a softmax classifier to diagnose rotating machinery health conditions directly. Chen et al. [4] developed a one-dimensional self-encoder for noise reduction of high-dimensional signals while learning stratified features for fault detection and diagnosis in variable processes. Wang et al. [26]. used a DNN model for the fault diagnosis of machines, and two datasets were adopted to verify the effectiveness. Yu et al. [33] presented a deep belief network was given that combined confidence and a softmax classifier into a DNN model. Zhang et al. [37] applied a DNN model that directly recognizes raw data without signal processing. However, in both encoder and DNN models, the shallower layer neurons are fully connected to the deeper layer neurons, and the possibility of gradient explosion increases as the number of layers increases. Unlike standard fully connected deep neural networks and encoders, deep convolutional neural networks consist of multiple convolutional-pooling layers with superior performance of sparse connectivity, shared weights, and translation invariance. Scholars at home and abroad have conducted many studies on the application of CNNs. Li et al. [15] developed a fault diagnosis model based on CNN and raw infrared images. The raw infrared images of rotating machinery are acquired, and then the image is fed into the CNN for feature extraction and recognition. Zhang et al.[34] used a diagnosis model based on CNN, which can automatically perform feature extraction and diagnostic decisions. Moreover, Huang et al. [9] developed an improved CNN model to diagnose bearing faults with multiscale information. However, it is still necessary to extract the statistical features of the vibration signal as the input of the CNN, the joint distribution information in the time and frequency domains is not considered, resulting in a weak mapping relationship between the signal representation and the mechanical health state.

To solve the above problems, this manuscript proposes an intelligent fault diagnosis of rolling bearings based on a continuous wavelet transform - multiscale feature fusion and improved channel attention mechanism (CWT-CAMCNN). The raw data acquired by acceleration sensors are fed into the CWT-CAMCNN model to perceive the fault state of the bearing. The hyper-parameters are effectively tuned using the Adam optimizer to obtain an ideal model based on the CWT-CAMCNN algorithm. The main contributions of this study are listed below.

(1) Compared with the classical convolutional neural network, the CWT-CAMCNN model proposed in this manuscript introduces an improved channel attention mechanism within the convolutional neural networks and establishes the correlation between channels.

(2) Converts 1-D signals into 2-D features by utilizing a continuous wavelet transform layer that highlights fault characteristics, allowing for faster training and less proneness to overfitting.

(3) The algorithm reduces the dependence on prior knowledge and manual labor, has strong adaptability and generalization ability, and is easy to apply to different industrial equipment.

(4) This methodology will be applied to single fault diagnosing eventualities and performs well in advanced fault diagnosing eventualities. Two experimental cases demonstrate its superiority.

The remainder of the manuscript is organized as follows. In Section 2, the basic theories of the continuous wavelet transform (CWT), improved channel attention mechanism (CAM), and multiscale CNNs are briefly introduced. In Section 3, the diagnostic cases of the bearing dataset are investigated using the proposed algorithm. Finally, Section 4 gives the conclusion of the manuscript.

## 2. Theoretical background

This manuscript proposes an intelligent fault diagnosis of rolling bearings based on a continuous wavelet transform-multiscale feature fusion and an improved channel attention mechanism (CWT-CAMCNN). The proposed fault diagnosis algorithm has the excellent capability of adaptive extraction of deep-seated features, without the need for manual feature extraction to represent the state features, and can directly extract the deep-seated features from transformed 2-D images. Fig. 1 shows the procedures of the proposed CNN-based diagnosis algorithm of bearings. The proposed diagnosis algorithm mainly consists of three steps.

**Step 1**. The raw vibration data are acquired by vibration sensors and data acquisition instruments, and then the raw vibration data are divided into the training set, validation set, and testing set. In addition, data are preprocessed and status data labels are specified.

**Step 2**. The data of the training set are fed into the CWT-CAMCNN model for training, and the data of the validation set are used to optimize the hyper-parameters to meet the accuracy requirements.

**Step 3**. The preprocessed data of the testing set are fed into the CWT-CAMCNN model, and the diagnostic findings are output.

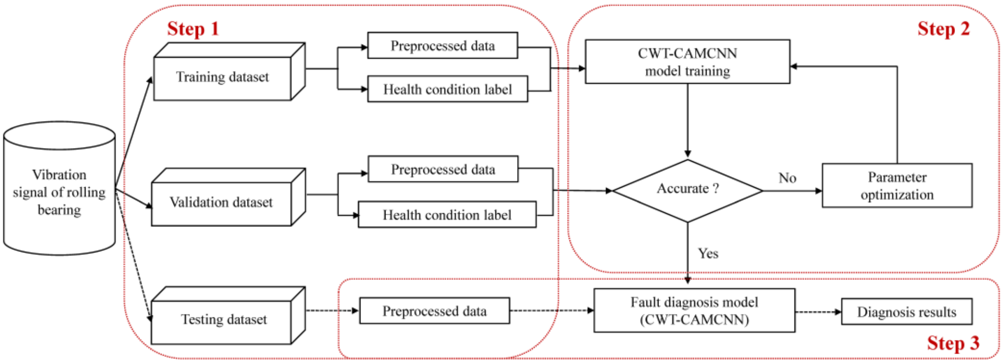


Fig. 1. The procedures of the proposed diagnosis algorithm for bearings.

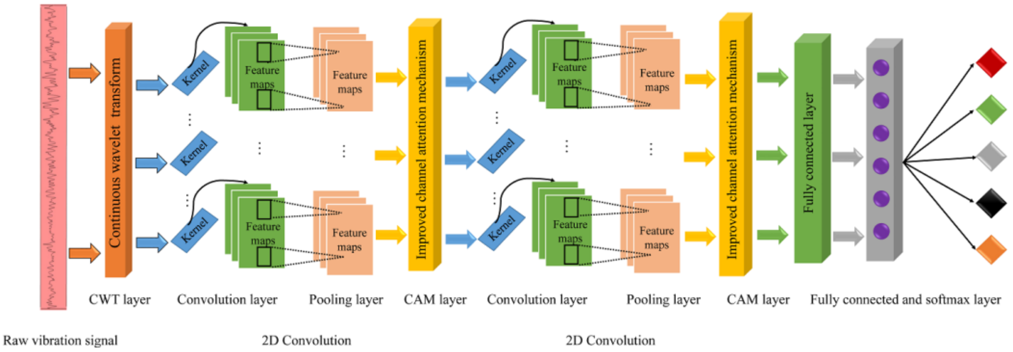


Fig. 2. The general fault diagnosis procedure of the proposed algorithm.

Fig. 2 shows the general fault diagnosis procedure of the proposed algorithm. The proposed fault diagnosis algorithm contains five essential parts: the CWT layer, convolution-pooling layers, improved CAM layers, fully connected layer, and softmax layer. Different from traditional CNNs, the raw data are fed into the CWT layer for time-frequency conversion, aiming to obtain wavelet power spectrum by combining the time domain and frequency domain, better highlighting the mapping relationship between signal representation and mechanical health state. Then, the time-frequency spectra are convolved by the multiscale convolution-pooling layers for feature extraction, and the multiscale feature fusion was implemented by the developed parallel convolutional neural networks. Meanwhile, the channel attention mechanism is improved by converting from compressed to extended ways in the excitation block to better obtain the evaluation score of the channel. In addition, the improved channel attention mechanism layer adaptively enhances or suppresses different channels to redistribute the weights for different fault levels. Fully connected layers are used for the nonlinear fitting of feature maps. Finally, the softmax layer generates the probabilities of different conditions. Model validation was conducted on the bearing datasets, and the model hyper-parameters and training parameters were kept consistent except for the difference in data length between the datasets.

## 2.1 Continuous Wavelet Transform (CWT)

CNNs specialize in handling 2-D or higher-dimensional data. In contrast, CNNs do not have ideal diagnostic performance for 1-D information [10]. Therefore, converting 1-D raw data into 2-D information forms is necessary. In the field of signal processing, many data preprocessing techniques have been developed for the time-frequency analysis of the raw signal, including short-time Fourier transform (STFT), wavelet synchro-squeezed transform (WSST), fast Fourier transform (FFT), HHT, and so on [2,21]. For nonstable signals, CWT is an expert at adaptively extracting the deep-level features adaptively that can be mapped from 1-D space to 2-D space to represent time-frequency features [7]. In addition, the Morlet function, as a standard wavelet function, has nonorthogonality. More importantly, the Morlet function is well-balanced between time-domain and frequency-domain analysis [2]. Therefore, the complex Morlet wavelet (cmor3-3) [7,31,24] is chosen as the wavelet generating function.

The CWT of the 1-D vibration signal sequence can be expressed as:

$$CWT(s,t)=\frac{1}{\sqrt{s}}\int_{-\infty}^{\infty}x(t)\psi\frac{t-\tau}{s}\mathrm{d}t\ s,\tau\in R,s>0 \quad (1)$$

where $S$ is the scale parameter. When $s>1$, it realizes the stretching of the $\psi(t)$, which is beneficial to extract the low-frequency features. In a similar way, when $0<s<1$, it realizes the compression of the $\psi(t)$, which is beneficial to extract the high-frequency features. $\tau$ is the translation parameter, which is used to specify the position of wavelet function translation along the time axis.

The CWT reflects the similarity between the mother wavelet and the signal through the inner product operation, and the mother wavelet should be chosen to be as similar as possible to the characteristics of the fault pulse, while being able to portray the fast oscillation characteristics of the signal. The selection of the mother wavelet function can be referred to in the literature [10], In this paper, the complex Morlet wavelet is used as the mother wavelet function, and its expression is.

$$\psi(t)=\pi^{-1/4}e^{-t^2/2}\cos(w_0t) \quad (2)$$

where $\boldsymbol{w}_0$ is the center frequency of the wavelet.

Fig. 3 shows the CWT time-frequency maps under different health conditions. Take the drive end bearing for the mixed working condition in the CWRU bearing dataset as an example, as shown in Table 6. As can be seen in Fig. 3, the CWT time-frequency maps for different working conditions have some similarities, it is difficult to visually observe the variations between these maps, and it is impractical to employ the information to identify faults. Hence, developing an intelligent algorithm to extract valuable features and identify fault types is critical and important.

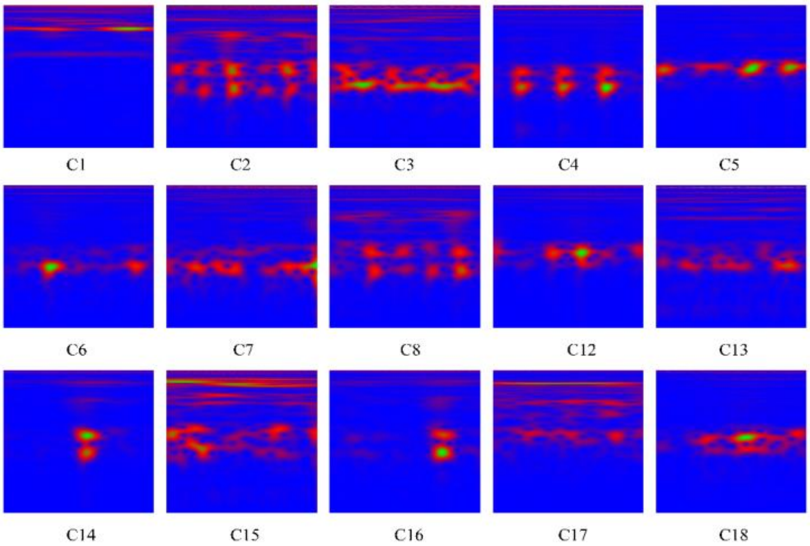


Fig. 3. The CWT maps of the raw data under different health conditions. C1: NO. C2: 0.007 IF. C3: 0.007 BF. C4: 0.007 OF1. C5: 0.007 OF2. C6: 0.007 OF3. C7: 0.014 IF. C8: 0.014 BF. C12: 0.021 IF. C13: 0.021 BF. C14: 0.021 OF1. C15: 0.021 OF2. C16: 0.021 OF3. C17: 0.028 IF. C18: 0.028 BF.

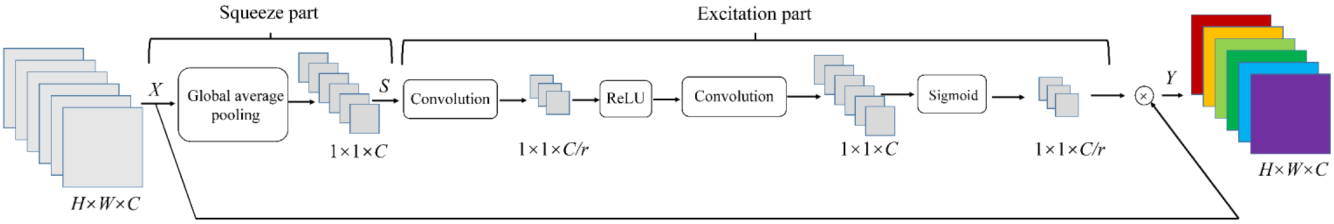


Fig. 4. The improved attention unit

### 2.2 Improved Channel Attention Mechanism

Attentional mechanisms play a critical role in human perception. When humans observe the environment, the visual system does not immediately process the whole scene but selectively acquires information. In the field of fault diagnosis, the CAM does precisely that the fault sensitivity of each channel is learned by modeling, and then the weight of different channels is adjusted to enhance or suppress. Finally, the adjusted weight is applied to the original channels to improve the performance of the model. The channel attention mechanism proposed in this manuscript is based on an improved version of the channel attention mechanism in the literature [14], where the network structure of the excitation block is converted from a compressed pattern to an extended pattern to better obtain the evaluation scores of the channels. The improved attention unit includes the squeeze and excitation blocks [16]. Fig. 4 shows the structure of the improved attention unit, which can be implemented in the following three steps. Table 1 shows the detailed parameters of the improved attention unit.

**Step 1**: After the multiscale convolution-pooling layer network operation, the size of the output feature map is $X = [x_1, x_2, x_3, \cdots, x_i]$, $W \times H$ are the width and length each feature map respectively, $C$ is the number of channels. Global average pooling (GAP) is applied to each channel to provide global information to the features after squeezing, it can be expressed as follows.

$$Z_{\mathrm{c}} = F_{\mathrm{sq}}(X_{\mathrm{c}}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{\mathrm{c}}(i,j) \cdot \tag{3}$$

where $Z_{\mathrm{c}}$ represents the compression value of the cth channel feature map. $X_c(i,j)$represents the eigenvalue of the cth channel at the position of $i$ and $j$.

**Step 2**: After the GAP operation, the size of the feature map is $C \times 1 \times 1$, and then the feature map is fed into the excitation block, which contains three cascaded layers including two convolutional layers (conv1, conv2), and one sigmoid layer. The conv1 layer uses ReLU, and the compression ratio r = 4 is used to reduce the number of channels. In this case, the size of the output feature map was $1 \times 1 \times C/r$. This can be conducted as follows.

$$\begin{cases} S_{\mathrm{e1}} = A_{\mathrm{relu}}\,(w_1 \times z + b_1) \\ A_{\mathrm{relu}}\,(x) = max(x, 0) \end{cases} \tag{4}$$

where $Z$ is the size of the feature map after the GAP operation. $w_1$ and $b_1$ are the weights and bias of the conv1, respectively. $A_{\mathrm{relu}}\,(\cdot)$ is the activation function.

The conv2 layer can recover the number of compressed channels and can adjust the weights of each channel while mapping the values to 0~1 by the sigmoid function. This process is expressed as follows.

$$\begin{cases} s = A_{\text{sigmoid}}\left(w_2 \times s_{e1} + b_2\right) \\ A_{\text{sigmoid}}(x) = \frac{1}{1+e^{-x}} \end{cases} \tag{5}$$

where $w_2$ and $b_2$ are the weights and bias of the conv2 layer, respectively. $A_{\text{sigmoid}}(\cdot)$ is the sigmoid activation function.

**Step 3**: Multiplying the weights obtained from Step 2 with the original feature map, the process is expressed as follows.

$$F_{\text{scale}}(X,s) = \left[s_1 \times X_{1,}s_2 \times X_{2,}, \cdots, s_i \times X_i\right] \tag{6}$$

Table 1 Detailed parameters of the improved attention unit

| Improved attention unit | | |
|---|---|---|
| Layer | Input size | Output size |
| GAP | H×W×C | 1×1×C |
| Conv1 | 1×1×C/r | 1×1×C |
| Conv2 | 1×1×C | 1×1×C/r |
| Sigmoid | 1×1×C/r | 1×1×C/r |
| Reshape | 1×1×C | H×W×C |
| Scale | H×W×C | H×W×C |

### 2.3 Multi-scale Convolutional Neural Network

CNNs are an essential branch of deep convolutional neural networks [6]. Based on the excellent performance of the sparse connectivity mechanism, shared weights mechanism, and translation invariance mechanism [28], CNNs have achieved outstanding success in both academic research and industry fields. At present, 1-D CNNs are widely applied in most studies. The classical convolutional neural network is designed for two-dimensional images and shows excellent ability in image classification tasks, however, it cannot be processed for one-dimensional mechanical signals, so this paper converts one-dimensional mechanical signals into two-dimensional image signals as input to the convolutional neural network through the continuous wavelet transform to realize fault identification. In this section, we introduce the basic theory of 2-D CNN models.

#### 2.3.1 Convolution-Pooling Module

The convolutional layer is primarily applied to the network by involving convolutional operations to accomplish feature extraction of the input images. A set of learnable kernel functions is an essential parameter in the feature extraction stage and directly determines the quality of the output features. While the kernel size corresponds to the perceptual range, its depth determines the number of output channels. The convolution operation is performed by multiplying the input image with a convolution kernel, adding a bias constant and a sliding window of specific step size. Therefore, the convolution process can be described as follows.

$$\begin{cases} Z^{l+1}(i,j) = [Z^l \otimes w^{l+1}](i,j) + b = \sum_{K=1}^{K_l} \sum_{x=1}^{f} \sum_{y=1}^{f} \left[Z_K^l(s_0 i + x, s_0 j + y) w_k^{l+1}(x,y)\right] + b \\ (i,j) \in \{0,1,\cdots,L_{l+1}\} L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1 \end{cases} \tag{7}$$

where $Z^l$ and $Z^{l+1}$ are the input and output of the $l+1$ layer, respectively. $b$ is the bias. $Z(i,j)$ is the image element value. $K$ is the number of channels. $f$ is the size of the convolution kernel. $S_0$ is the size of the step. $p$ is the number of padding. $L_{l+1}$ is the output size of $Z_{l+1}$.

After the convolution layer, pooling is usually carried out for downsampling operations. In pooling operations, the standard procedures are average pooling and max-pooling, which average or maximize the pixel values of the pooled window. The pooling operation is equivalent to reducing the input feature graph so that the model can extract a broader range of features. After the pooling operation, the length and width of the feature graph are reduced. Therefore, the number of parameters and the amount of calculation are reduced. Reducing the number of parameters prevents the model from overfitting, and the pooling operation has a certain robustness to the noise interference. The most common pooling layer is max-pooling or average pooling, and the max-pooling process is represented as follows.

$$\begin{cases} fea_{i,j}^{L} = f\left(fea_{i,j}^{L-1}\right) \\ i' = \left(0,1,2,\dots,\left[\frac{i-d}{s}\right]\right) \end{cases} \tag{8}$$

where $fea_{i,j}^{L-1}$ denotes the jth feature map of size i of the $L$ layer. $f$ denotes the pooling process. $d$ denotes the size of the pooling function. $S$ denotes the move step of the pooling function. $fea_{i,j}^{L}$ is the output feature map after the pooling operation.

#### 2.3.2 Fully Connected Layer

After passing through multiple convolutional layers and pooling layers, the fully connected layer is usually connected to integrate the local features extracted from the convolutional layer or pooling layer. The full connected layer is a traditional multilayer perceptron that connects each neuron in the lower layer to each neuron in the upper layer. The primary role of the fully connected layer is to extract features further and connect the output with the softmax classifier. The fully connected layer usually consists of two to three hidden layers. To improve the performance of CNN models, the activation function of neurons in the fully connected layer typically employs the ReLU function, and all neurons in the hidden layer are interconnected according to the following definition.

$$z_j^L = \text{ReLU}\left(\sum_{i=1}^{M} x_i^{L-1} w_{j,i}^L + b_j^L\right), j = 1,2,\dots,N \tag{9}$$

where $w_{j,i}^L$ is the connection weight from the ith neuron of the $L-1$ layer to the j-th neuron of the $L$ layer. $x_i^{L-1}$ denotes the input of the i-th neuron of the $L$ layer. $b_j^L$ is the bias of the i-th neuron of the $L$ layer. $M$ and $N$ are the number of neurons of the $L-1$ and $L$ layer, respectively.

#### 2.3.3 Zero-Padding Method

When applying multilayer convolution, we always miss the edge information. To address this matter, we proposed a zero-padding method. The padding method means that the zero elements were padded on both sides of the input height and width, and used for CNN models to prevent the loss of feature dimensionality. $Q$ represents the input dimension. $O$ represents the output dimension. $F$ represents the convolutional kernel size. $S$ represents the step size, and the left padding $L$ and right padding $R$ can be expressed as

$$\begin{cases} O = \text{ceil}\left(\frac{Q}{S}\right) \\ T = (O-1) \times S + F - Q \\ L = \text{floor}\left(\frac{T}{2}\right) \\ R = T - L \end{cases} \tag{10}$$

where ceil $(\cdot)$ and floor $(\cdot)$ are the ceil and floor functions, respectively.

To demonstrate the proposed model structure more clearly, Table 2 shows the network details and main parameters.

Table 2 Network details and main parameters

| Layers | Output size | Channel | Filter | Stride | Padding |
|---|---|---|---|---|---|
| Conv2d_1 | (None,32,115,335) | (1,32) | (3,3) | (2,2) | (32,32) |
| Attention unit_1 | (None,32,115,335) | 128 | / | / | / |
| Conv2d_2 | (None,64,52,107) | (32,64) | (2,2) | (2,2) | (48,48) |
| Conv2d_3 | (None,32,13,26) | (64,32) | (2,2) | (2,2) | (0,0) |
| Conv2d_4 | (None,16,38,45) | (32,16) | (2,2) | (2,2) | (32,32) |
| Conv2d_5 | (None,6,38,45) | (32,16) | (2,2) | (2,2) | (32,32) |
| Conv2d_6 | (None,16,38,45) | (32,16) | (2,2) | (2,2) | (32,32) |
| Concatenatio_1 (Conv2d_4_5_6) | (None,48,38,45) | 48 | / | / | / |
| Conv2d_7 | (None,16,12,15) | (48,16) | (3,3) | (3,3) | (0,0) |
| Conv2d_8 | (None,16,12,15) | (48,16) | (3,3) | (3,3) | (0,0) |
| Avg_pooling | (None,32,6,7) | / | (2,2) | (2,2) | (0,0) |
| Attention unit_2 | (None,32,6,7) | 128 | / | / | / |
| Fully connected layer | (None,512) | 512 | / | / | / |

## 3. Experimental

In this section, the effect of the proposed fault diagnosis algorithm is discussed based on the experiment, and the proposed algorithm is tested on two datasets, including the CWRU bearing data and experimental bearing data. The CWT-CAMCNN models are written in PyTorch and run on a computer with an i7-10700 CPU, RAM 16.00 GB, and GTX1660 Ti GPU. The operating system is 64-bit Win10. During CWT-CAMCNN training, the learning rate is 0.001, the batch size is 16, the epoch is 60, and the Adam optimizer is selected for optimization.

### 3.1 Case study 1: On the CWRU bearing datasets

### 3.1.1 Datasets description

All the data in the CWRU datasets are collected on the bearing testbed, which is shown in Fig. 5. The testbed primarily consists of a motor, a fan end bearing, a drive end bearing, a coupling, and a dynamometer. The testbed collects raw vibration signals from the fan end bearing (FE) and the drive end bearing (DE). The single-point fault of the bearing has four fault diameters: 0.18, 0.36, 0.54, and 0.71 mm. The testbed includes four fault forms: normal (NO), inner race faults (IF), ball faults (BF), and outer race faults (OF). The outer race failures are divided into three forms of failure according to the location of the failure(OF1/OF2/OF3). Two acceleration sensors were placed to collect the vibration signals of the two bearings with a sampling frequency of 12 kHz. For the DE fault, the sampling frequency is 48 kHz, and speed and horsepower data are collected using a torque transducer, which contains four working conditions: 0 - 1797, 0.75 - 1772, 1.5 - 1750, and 2.25 - 1730 (kW - rpm). Therefore, the frequency is approximately 30 Hz, so the time series length at one rotation cycle is 400. In the case of nonoverlapping samples, the fault signal sequence is segmented, 75% of which is used for training, and the remaining sample is used for testing the network.

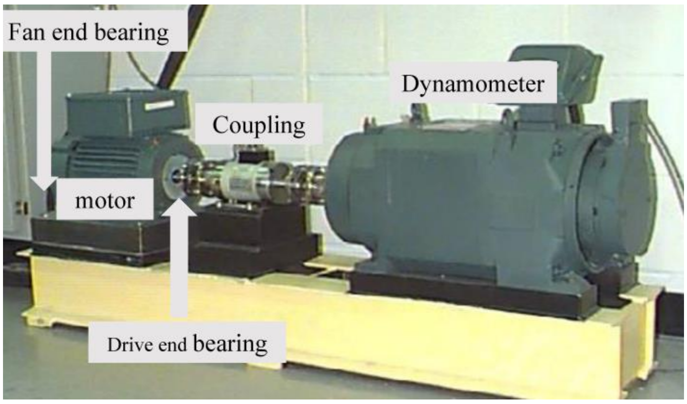


Fig. 5. Test rig used in the CWRU lab

### 3.1.2 Single bearing fault diagnosis

Tables 3 and 4 show the test accuracy of DE and FE at different fault sizes and working conditions, respectively. Table 3 shows that the fault sizes are 0.18, 0.36, 0.54, and 0.71 mm, and the working conditions are 0 - 1797, 0.75 - 1772, 1.5 - 1750, and 2.25 - 1730 (kW - rpm), respectively. The test accuracy for single working conditions can be 100%, except for a miscalculation at 1.5 kW - 1750 rpm and a fault size of 0.54 mm. As seen in Table 4, except for the working conditions with fault sizes of 0.18 and 0.36 mm, misjudgment occurred, but the accuracy exceeded 98%. In summary, the proposed diagnostic algorithm has higher diagnostic accuracy in all cases.

Table 3 Accuracy on the single DE

| Fault size (mm) | Working condition (kW - rpm) | NO | 12 k DE faults | | | | | Average accuracy |
|---|---|---|---|---|---|---|---|---|
| | | | IF | BF | OF1 | OF2 | OF3 | |
| 0.18 | 0 - 1797 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 0.75 - 1772 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 1.5 - 1750 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 2.25 - 1730 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.36 | 0 - 1797 | 1 | 1 | 1 | 1 | / | / | 1 |
| | 0.75 - 1772 | 1 | 1 | 1 | 1 | / | / | 1 |
| | 1.5 - 1750 | 1 | 1 | 1 | 1 | / | / | 1 |
| | 2.25 - 1730 | 1 | 1 | 1 | 1 | / | / | 1 |
| 0.54 | 0 - 1797 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 0.75 - 1772 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 1.5 - 1750 | 1 | 1 | 1 | 0.986 | 1 | 1 | 0.998 |
| | 2.25 - 1730 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.71 | 0 - 1797 | 1 | 1 | 1 | / | / | / | 1 |
| | 0.75 - 1772 | 1 | 1 | 1 | / | / | / | 1 |
| | 1.5 - 1750 | 1 | 1 | 1 | / | / | / | 1 |
| | 2.25 - 1730 | 1 | 1 | | / | / | / | 1 |

Table 4 Accuracy on the  single FE

| Fault size (mm) | Working condition (kW - rpm) | NO | 12 k DE faults | | | | | Average accuracy |
|---|---|---|---|---|---|---|---|---|
| | | | IF | BF | OF1 | OF2 | OF3 | |
| 0.18 | 0 - 1797 | | 1 | 1 | 1 | 0.989 | 1 | 0.998 |
| | 0.75 - 1772 | | 1 | 1 | 1 | 0.961 | 1 | 0.993 |
| | 1.5 - 1750 | | 1 | 1 | 0.985 | 0.972 | 0.936 | 0.982 |
| | 2.25 - 1730 | | 1 | 1 | 1 | 0.973 | 0.986 | 0.993 |
| 0.36 | 0 - 1797 | | 1 | 1 | / | 1 | / | 1 |
| | 0.75 - 1772 | | 0.988 | 1 | / | 0.988 | / | 0.994 |
| | 1.5 - 1750 | | 1 | 1 | / | 1 | / | 1 |
| | 2.25 - 1730 | | 1 | 1 | / | 1 | / | 1 |
| 0.54 | 0 - 1797 | | 1 | 1 | / | 1 | / | 1 |
| | 0.75 - 1772 | | 1 | 1 | / | 1 | / | 1 |
| | 1.5 - 1750 | | 1 | 1 | / | 1 | / | 1 |
| | 2.25 - 1730 | | 1 | 1 | / | 1 | / | 1 |

T-SNE is an unsupervised, nonlinear learning dimension reduction algorithm and visualization for representing high-dimensional datasets in low-dimensional space. The feature distributions of the learning results of different layers are visualized as probability distributions with t-SNE. In this manuscript, we employ t-SNE to visually interpret how the model makes decisions for fault diagnosis.

The acquired vibration signals are segmented and fed into the diagnostic model. To prove the effectiveness of the proposed algorithm, take the FE bearing(Table 4) as an example. Fig. 6 shows the diagnostic decision process of the proposed algorithm for six bearing health states under the working condition of 0 kW - 1797 rpm for NO, IF, BF, and OF with fault sizes of 0.18 mm. It can be seen that the raw data can be analyzed after CWT transformation, and the IF, BF, and OF (OF1/OF2) of the bearings can be grouped through the conv_1 and attention unit_1 layers only and have distinct boundaries. However, at this stage, there is an apparent feature overlap between the NO and OF3. After multilayer convolution for feature extraction and merging (concatenate_1 and concatenate_2), the overlapping features of the OF3 and NO of the bearing are starting to become grouped together. After attention unit_2 and the fully connected layer (FC), the OF3 and NO features are grouped, and the features of each health state form a cluster with some distance. Therefore, the proposed fault diagnosis algorithm CWT-CAMCNN can diagnose faults using CWT time-frequency images of raw signals.

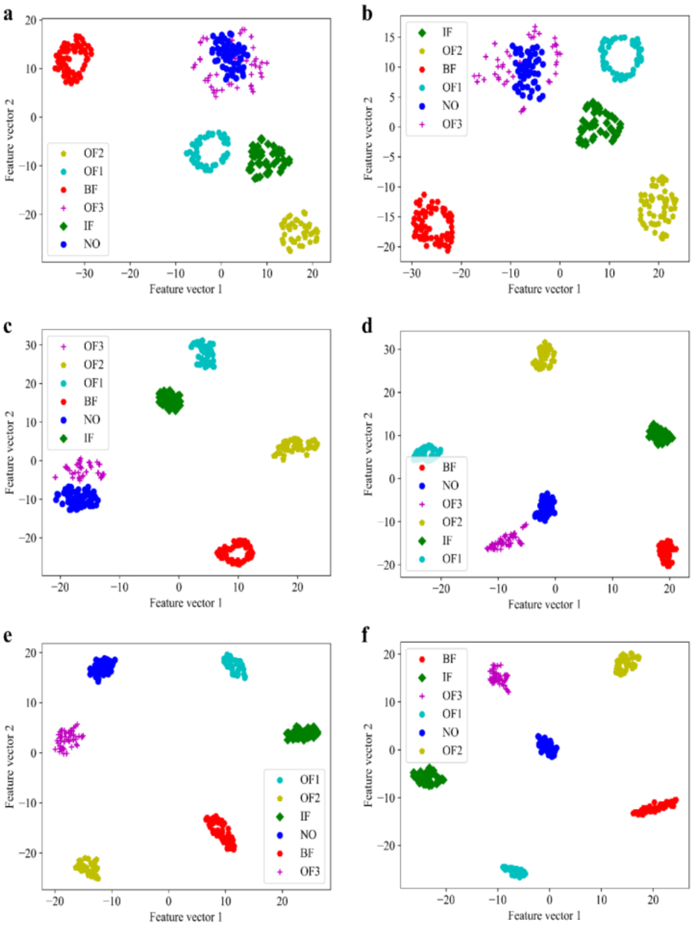


Fig. 6. Feature representations of the model via t-SNE. (a) Conv_1, (b) Attention unit_1, (c) Concatenate_1, (d) Concatenate_2, (e) Attention unit_2, (f) FC layer

### 3.1.3 Mixed Bearing Fault Diagnosis

Table 5 Accuracy under mixed fault sizes on the DE

| Fault size (mm) | Types | Label | Working condition (kW - rpm) | | | |
|---|---|---|---|---|---|---|
| | | | 0 - 1797 | 0.75 - 1772 | 1.5 - 1750 | 2.25 -1730 |
| NO | | 0 | 1 | 1 | 1 | 1 |
| 0.18 | IF1 | 1 | 1 | 1 | 1 | 1 |
| | BF1 | 2 | 1 | 1 | 1 | 1 |
| | OF1 | 3 | 1 | 1 | 1 | 1 |
| | OF2 | 4 | 1 | 1 | 1 | 1 |
| | OF3 | 5 | 1 | 1 | 1 | 1 |
| 0.36 | IF2 | 6 | 1 | 1 | 1 | 1 |
| | BF2 | 7 | 1 | 1 | 1 | 1 |
| | OF4 | 8 | 1 | 1 | 1 | 1 |
| | IF3 | 9 | 1 | 1 | 1 | 1 |
| | BF3 | 10 | 1 | 1 | 1 | 1 |
| 0.54 | OF5 | 11 | 1 | 1 | 1 | 0.988 |
| | OF6 | 12 | 1 | 1 | 1 | 1 |
| | OF7 | 13 | 0.985 | 1 | 1 | 0.988 |
| 0.71 | IF4 | 14 | 1 | 1 | 1 | 1 |
| | BF4 | 15 | 1 | 1 | 1 | 1 |
| Average | | | 0.999 | 1 | 1 | 0.999 |

Table 6 Accuracy under mixed working conditions on the DE

| Working condition (kW - rpm) | Condition number | Types | Label | Fault size (mm) | | | |
|---|---|---|---|---|---|---|---|
| | | | | 0.18 | 0.36 | 0.54 | 0.71 |
| NO | C1 | | 0 | 1 | 1 | 1 | 1 |
| 0 - 1797 | C2 | IF | 1 | 1 | 1 | 1 | 1 |
| | C3 | BF | 2 | 1 | 1 | 0.972 | 0.975 |
| | C4 | OF1 | 3 | 1 | / | 0.955 | / |
| | C5 | OF2 | 4 | 0.972 | / | 0.931 | / |
| | C6 | OF3 | 5 | 1 | 1 | 0.944 | / |
| 0.75 - 1772 | C7 | IF | 6 | 1 | 1 | 1 | 1 |
| | C8 | BF | 7 | 0.987 | 0.986 | 1 | 0.971 |
| | C9 | OF4 | 8 | 0.975 | 1 | 0.915 | / |
| | C10 | OF5 | 9 | 1 | / | 1 | / |
| | C11 | OF6 | 10 | 0.754 | / | 0.953 | / |
| 1.5 - 1750 | C12 | IF | 11 | 1 | 1 | 1 | 1 |
| | C13 | BF | 12 | 1 | 0.986 | 1 | 0.970 |
| | C14 | OF7 | 13 | 0.968 | 1 | 0.931 | / |
| | C15 | OF8 | 14 | 1 | / | 0.938 | / |
| | C16 | OF9 | 15 | 0.977 | / | 0.960 | / |
| 2.25 - 1730 | C17 | IF | 16 | 1 | 1 | 1 | 1 |
| | C18 | BF | 17 | 1 | 1 | 1 | 0.909 |
| | C19 | OF10 | 18 | 1 | 1 | 1 | / |
| | C20 | OF11 | 19 | 1 | / | 1 | / |
| | C21 | OF12 | 20 | 1 | / | 1 | / |
| Average | | | | 0.983 | 0.998 | 0.976 | 0.981 |

To further verify the effectiveness of the proposed algorithm on the multi-classification task, the algorithm is applied to fault diagnosis under mixed working conditions. As shown in Tables 5 and 6, there are 16 and 21 bearing failures corresponding to the four types of failures, respectively. As seen in table 5, the test accuracy can meet 100% for the two working conditions of 0.75 kW - 1772 rpm and 1.5 kW - 1750 rpm and 99.9% for the other two working conditions of 0 kW - 1797 rpm and 2.25 kW - 1730 rpm, the miscalculation occurs mainly in the identification of faults in outer race faults (OF5/OF7). As seen in Table 6, the test accuracy can meet more than 97% in the four mixed conditions, among which the test accuracy of the mixed condition with a fault size of 0.36 mm is as high as 99.8%. The test accuracy of the mixed conditions with fault sizes of 0.18 and 0.71 mm can meet more than 98%. The misjudgment also occurs mainly in the identification of the OF, indicating that the proposed method has a certain missing ability in the identification of defects in the case of similar fault form and fault location, but the average fault identification accuracy is still higher than 98%. Therefore, it can be shown that the proposed method in this manuscript still has higher diagnostic accuracy for mixed conditions of rolling bearings.

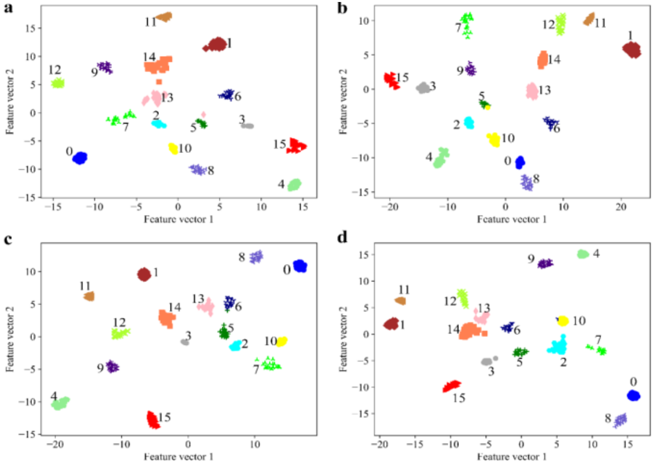


Fig. 7. Output data visualization of mixed fault size. a: 0 kW - 1797 rpm, b: 0.75 kW - 1772 rpm, c: 1.5 kW - 1750 rpm, d: 2.25 kW - 1730 rpm.

To show the classification effect of the proposed algorithm more intuitively, t-SNE is employed to visualize the output of the fully connected layer. Taking DE-bearing test accuracy for mixed fault sizes as an example(table 5), the visualization of the low-dimensional feature subset obtained after dimensionality reduction is shown in Fig.7. It can be seen from Fig. 7, the features under each healthy working condition form a family, and each family possesses obvious boundaries without feature overlap. There is a certain distance between each family. Therefore, the CWT-CAMCNN model proposed in this manuscript can be used for fault diagnosis by using the CWT time-frequency maps of vibration signals.

Take DE bearing for mixed working conditions (table 6) as an example, the accuracy and loss curves of the training set and validation were produced, as shown in Fig. 8. During training and validation, the accuracy and loss of each epoch are recorded, where five epochs are used as plot points. It can be seen that the accuracy rate fluctuates steadily after 50 epochs, and there is no obvious inflection point in the curve. In addition, to more intuitively show the superiority of the proposed in the training and validation process, the maximum accuracy and minimum loss values are marked in the unit of batch size. From the accuracy curves, we can see that the proposed model can realize over 80% diagnostic accuracy after 20 training steps, which indicates that the model has high fitting ability.

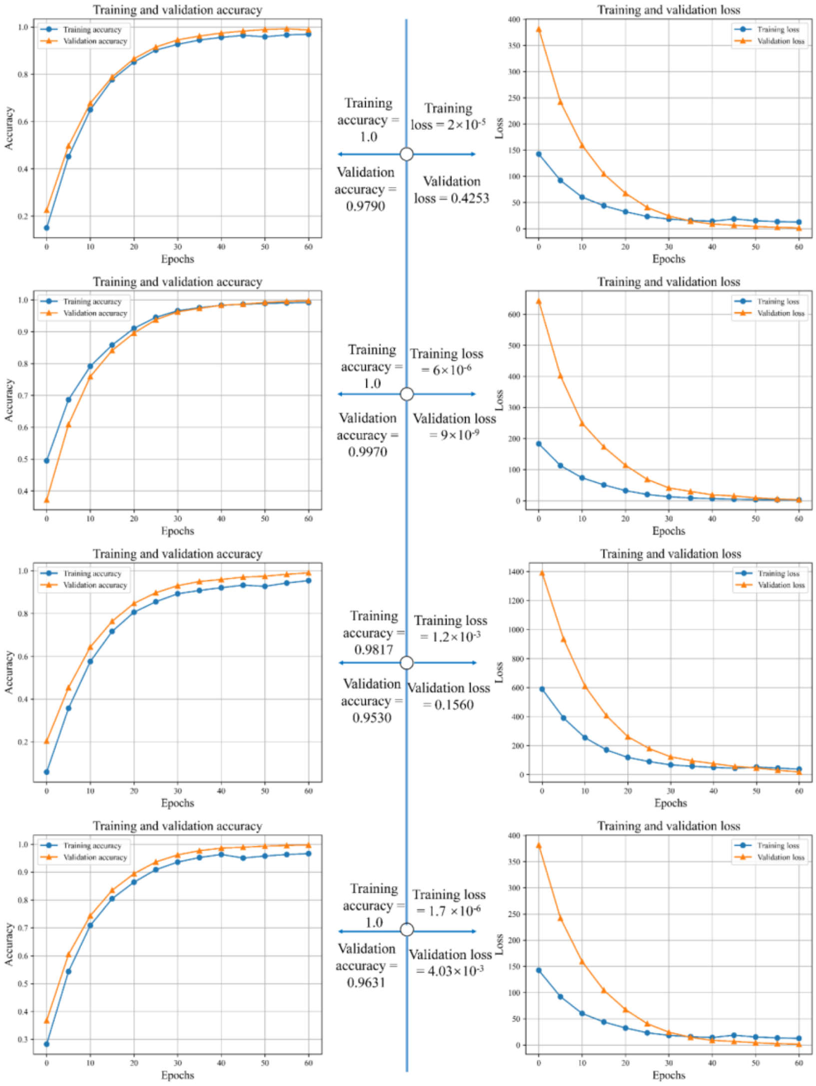


Fig. 8. The accuracy and loss curves of the proposed algorithm

### 3.1.4 Comparison with Existing DL Algorithms

To further intuitively explain the advantages of the proposed CWT-CAMCNN, other popular existing supervised and unsupervised deep learning (DL) approaches are carried out for comparison on the CWRU bearing dataset, the experimental results of different algorithms are listed in Table 7.

Table 7 Diagnosis accuracies(%) of various methods on the CWRU bearings dataset

| Reference | Feature extraction | Bearing fault types | Type number | Accuracy% |
|---|---|---|---|---|
| Proposed algorithm | CWT and CAMCNN | 1NO/4IF/4BF/7OF | 16 | 99.95 |
| Proposed method without CWT | CAMCNN | 1NO/4IF/4BF/7OF | 16 | 99.84 |
| Proposed method without CAM | CWTCNN | 1NO/4IF/4BF/7OF | 16 | 99.74 |
| Li et al.(2020) [30] | SAE and DBN | 1NO/4IF/4BF/7OF | 16 | 99.69 |
| Wang et al. (2019) [14] | SWWF | 1NO/1IF/1BF/1OF | 4 | 99.57 |
| Bai et al. (2021) [26] | MCNN and MSCF | 1NO/3IF/3BF/3OF | 10 | 99.03 |
| Xiong et al. (2016) [1] | MFDFA and ASD | 1NO/2IF/2BF/2OF | 7 | 98.6 |
| Liu et al. (2018) [31] | 10 Statistical Features | 1NO/3IF/3BF/3OF | 10 | 98.4 |
| Zhu et al. (2019) [18] | ICN | 1NO/1IF/1BF/1OF | 4 | 97.15 |
| Ying et al. (2018) [38] | EHCT | 1NO/4IF/3BF/3OF | 11 | 96.97 |
| Li et al. (2016) [33] | MDA | 1NO/4IF/3BF/3OF | 11 | 96.59 |
| Zhang et al. (2021) [13] | GRU | 1NO/3IF/3BF/3OF | 10 | 94.22 |
| Zhang et al. (2021) [13] | SVM | 1NO/3IF/3BF/3OF | 10 | 81.72 |

It can be seen that the proposed method has the highest diagnostic accuracy on the task of 16 classification. Nevertheless, the accuracy of the proposed algorithm on the CWRU-bearing dataset can reach 99.95%. Specifically, CWT-CAMCNN outperforms Li [18], Wang [27], Bai [29], and Xiong [14] with improvements of 0.26%, 0.38%, 0.92%, and 1.35%, respectively, on the CWRU bearing dataset. The effect of network structure on diagnostic accuracy is also tested, and the results show that when the proposed method uses only raw data as input, the test diagnostic accuracy can still meet 99.84%, and when the CAM module is excluded, the test diagnostic accuracy is 99.74%, which are higher than that of the existing deep learning methods. Therefore, the proposed algorithm can achieve better performance. Moreover, the experimental results also show the superiority of the deep learning-based fault diagnosis algorithm over the shallow deep learning methods of GRU and SVM algorithms.

### 3.2 Case study 2: On The Experiment Bearing Datasets

### 3.2.1 Datasets Description

Fig. 9 shows the laboratory bearing test stand, consisting of a motor frequency conversion controller, a motor, a coupling, a bearing seat, a rotor, and other parts. Fig. 10 shows the eight health conditions of the bearings designed in the experiment, including one normal condition (NO), three inner ring faults (IF1/IF2/IF3), and four outer ring faults (OF1/OF2/OF3/OF4). The performance of the model was tested under variable speed conditions of 15, 20, 25, and 30 Hz. Table 8 shows the bearing fault type.

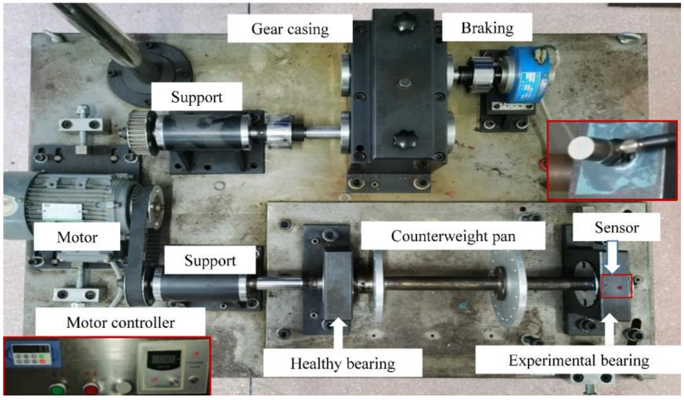


Fig. 9. The laboratory bearing test stand

Fig. 10. Different bearing healthy states

The sampling frequency is set to 19.2 kHz, and the sampling time is 110 s. The acceleration sensor is mounted on the housing of the bearing seat. To ensure that all samples of fault points can be collected within a rotation cycle, the time series length in a rotation cycle is 1280. In the case of nonoverlapping samples, the fault signal sequence is divided, 75% of which is used for training, and the remaining samples are used for testing the model.

Table 8 Bearing fault type

| Label | Bearing fault type | | |
|---|---|---|---|
| | Length/mm | Width/mm | Depth/mm |
| NO | Healthy | Healthy | Healthy |
| OF1 | 7 | 2 | 1 |
| OF2 | 3 | 2 | 1 |
| OF3 | 1 circular hole with a diameter of 1.5 mm | | |
| OF4 | 3 | 2 | 1 |
| IF1 | 3 | 3 | 1 |
| IF2 | 9 circular holes with a diameter of 2 mm | | |
| IF3 | 3 | 2 | 1 |

### 3.2.2 Experiment results and analysis

The performance of the model proposed in the manuscript is visualized layer by layer by t-SNE, taking the speed of 25 Hz as an example.

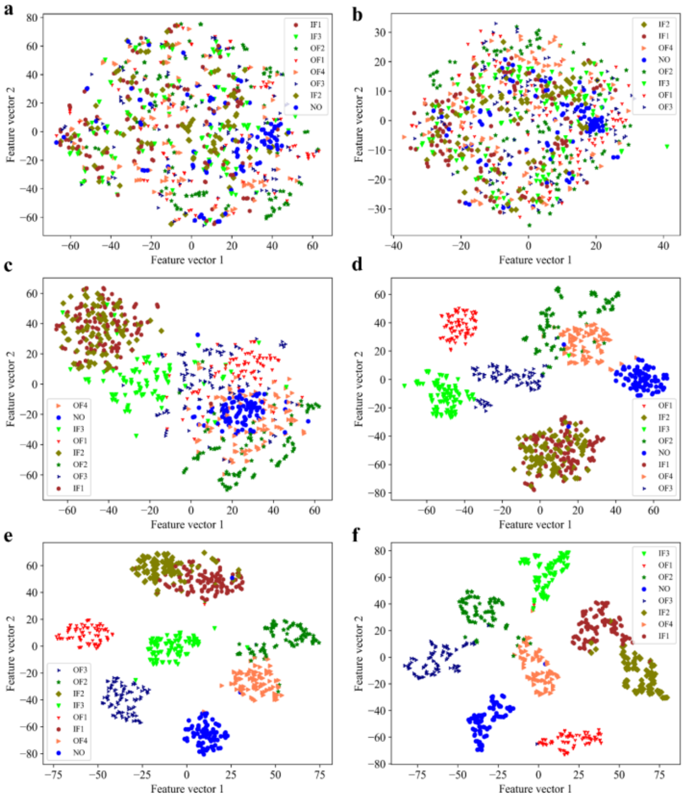


Fig. 11. Feature representations of the model via t-distributed stochastic neighbor embedding. (a) Conv_1, (b) Attention unit_1, (c) Concatenate_1, (d) Concatenate_2, (e) Attention unit_2, (f) FC layer.

Fig. 11 shows the feature representations of the model via t-SNE. After the CWT of the raw bearing data, it is not easy to distinguish the eight bearing health states through the conv_1 layer, attention unit_1 layer, and concatenate_1 layer. However, when passing through the concatenate_2 layer, NO, IF3 and OF3 can be correctly distinguished from each other and have obvious boundaries, although there are some overlapping features between OF2, OF4, IF2, and IF1. After the attention unit_2 layer, some overlapping features are distinguished; only IF2 and IF1 have little overlapping features, but the overlapping features are significantly reduced compared to the concatenate_2 layer. After passing through the FC layer. The eight bearing health states can be correctly distinguished, and there is a clear boundary between different samples. Therefore, it can be shown that the fault diagnosis of varying health states of laboratory bearings can be performed without adjusting the hyper-parameters of the algorithm proposed in this manuscript.

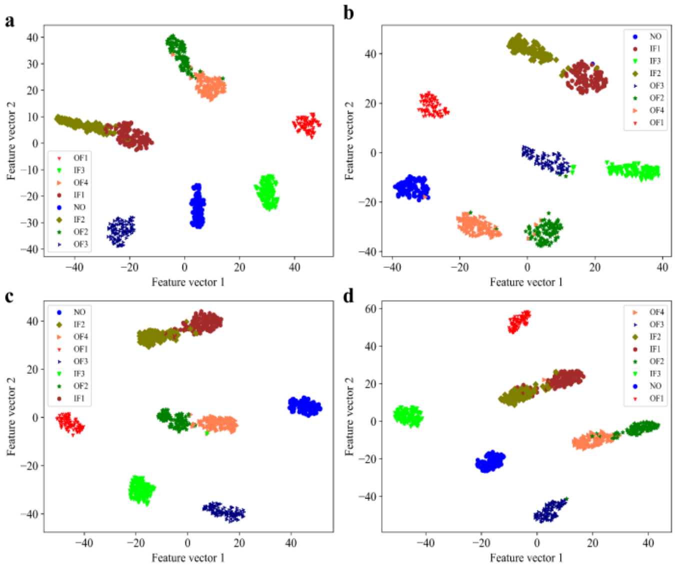


Fig. 12. The output data visualization of FC layer

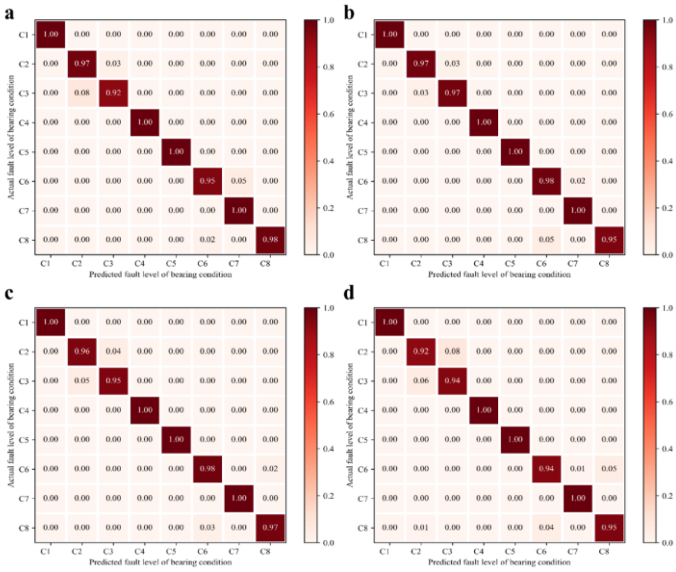
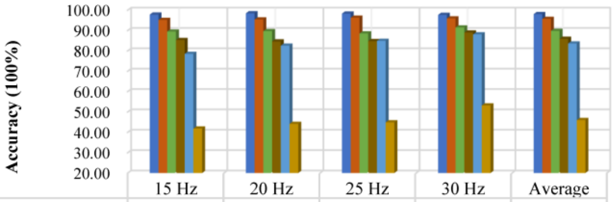


Fig. 13. Confusion matrices of variable speed conditions

To further demonstrate the performance of the proposed method for variable speed conditions, t-SNE and confusion matrices were plotted for the conditions of 15, 20, 25, and 30 Hz. Fig. 12 and Fig.13 show the output data visualization and confusion matrices of variable speed conditions. The proposed model in the manuscript can correctly distinguish the eight health states of the bearings. However, there are some overlapping characteristics of IF1, IF2, OF2, and OF3 for the four working conditions. Nevertheless, by recording the accuracy of the four working conditions, the misdiagnosis rate of the inner ring of the bearing is 0.03 and 0.08, 0.08 and 0.06, 0.04 and 0.05, and 0.03 and 0.03, the misdiagnosis rate of the outer ring of the bearing is 0.05 below. In addition to the statistics of the average accuracy under a single working condition, the accuracy reached 97.75%, 98.38%, 98.25%, and 97.63%, respectively. The average accuracy under the four working conditions is 98.00%, so the proposed algorithm CWT-CAMCNN algorithm has higher accuracy and excellent generalization ability in bearing fault diagnosis, which can provide the basis for the reliability and safety of bearings.

### 3.2.3 Comparison with existing algorithms

Fig. 14 compares the accuracies of different diagnostic algorithms under four working conditions. Under four working conditions, the proposed algorithm has the best diagnosis result among all algorithms. The proposed algorithm consists of a CWT module, an improved channel attention mechanism module, and a multiscale convolution-pooling module. To compare the effectiveness of each module in the model, the diagnostic accuracy of several different combinations of algorithms is compared. In terms of the average diagnostic accuracy, all the modules in the model play a crucial role. In addition, CWT processing has a significant impact on diagnostic accuracy.



| | 15 Hz | 20 Hz | 25 Hz | 30 Hz | Average |
|---|---|---|---|---|---|
| ■ Proposed method | 97.75 | 98.38 | 98.25 | 97.63 | 98.00 |
| ■ CWT + CNN | 95.13 | 95.37 | 96.25 | 95.87 | 95.66 |
| ■ STFT + CAM + CNN | 89.44 | 89.63 | 88.50 | 91.44 | 89.75 |
| ■ STFT + CNN | 85.25 | 84.56 | 84.75 | 88.94 | 85.88 |
| ■ Raw data + CNN | 78.44 | 82.49 | 84.84 | 88.03 | 83.55 |
| ■ Raw data + DNN | 41.79 | 44.08 | 44.85 | 53.14 | 45.97 |

Fig. 14. Diagnostic accuracy comparison of different algorithms

## 4. Conclusions

This manuscript proposes a rolling bearing fault diagnosis algorithm by integrating the CWT module, multiscale feature fusion and an improved channel attention mechanism. First, the CWT module transforms the time domain signal into the frequency domain to obtain the wavelet power spectrum. Then, feature extraction is further performed by multiscale convolutional and pooling layers. Additionally, the improved channel attention module is used in the convolutional layer to autonomously learn the sensitivity of each channel to faults. Then, the weights are redistributed by adaptively enhancing or suppressing different channels for fault sensitivity, and finally, the softmax classifier is used for fault diagnosis. Unlike most existing diagnostic models, the proposed algorithm can adaptively adjust the weights to improve the diagnosis efficiency and generalization ability. The algorithm is validated using the CWRU bearing data and experimental bearing data. It is also compared with existing deep learning algorithms. The results show that the algorithm can perform highly accurate fault diagnoses under various working conditions. However, detailed analysis of experimental results and a couple of parameter sensitivity studies should be added to evaluate the effects of the key hyper-parameters of the proposed method on the performance, this job is what we will do next.

## References

1. Bai R, Xu Q, Meng Z et al. Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation. Measurement 2021; 184, 109885, https://doi.org/10.1016/j.measurement.2021.109885.
2. Basaran M, Fidan M. Induction motor fault classification via entropy and column correlation features of 2D represented vibration data. Eksploatacja i Niezawodnosc –Maintenance and Reliability 2021; 23 (1): 132–142, http://dx.doi.org/10.17531/ein.2021.1.14.
3. Buzzoni M, Mucchi E, Dalpiaz G. A CWT-based methodology for piston slap experimental characterization. Mechanical Systems and Signal Processing 2017; 86 16-28, https://doi.org/10.1016/j.ymssp.2016.10.005.
4. Chen S, Yu J, Wang S. One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes. Journal of Process Control 2020; 87 54-67, https://doi.org/10.1016/j.jprocont.2020.01.004.
5. Dhiman H S, Deb D, Muyeen S M et al. Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. IEEE Transactions on Energy Conversion 2021; 36 3462-3469, https://10.1109/TEC.2021.3075897.
6. Guo Y, Hu T, Zhou Y et al. Multi-channel data fusion and intelligent fault diagnosis based on deep learning. Measurement Science and Technology 2022; 34 015115, https:// 10.1088/1361-6501/ac8a64.
7. Hu A, Xiang L, Zhang Y. Experimental study on the intrawave frequency modulation characteristic of rotor rub and crack fault. Mechanical Systems and Signal Processing 2019; 118 209-225, https://doi.org/10.1016/j.ymssp.2018.08.051.
8. Hu Z X, Wang Y, Ge M F et al. Data-driven fault diagnosis method based on compressed sensing and improved multiscale network. IEEE Transactions on Industrial Electronics 2019;67 3216-3225, http://ieeexplore.ieee.org.doi: 10.1109/TIE.2019.2912763.
9. Huang W, Cheng J, Yang Y et al. An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. Neurocomputing 2019; 359 77-92, https://doi.org/10.1016/j.neucom.2019.05.052.
10. Jiang Y, Tang B, Qin Y et al. Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. Renewable energy 2011; 36 2146-2153, https://doi.org/10.1016/j.renene.2011.01.009.
11. Kong X, Li X, Zhou Q et al. Attention recurrent autoencoder hybrid model for early fault diagnosis of rotating machinery. IEEE Transactions on Instrumentation and Measurement 2021; 70 1-10, https://10.1109/TIM.2021.3051948.
12. Li J, Cao Y, Ying Y et al. A rolling element bearing fault diagnosis approach based on multifractal theory and gray relation theory. PLoS One 2016; 11 0167587, https://doi.org/10.1371/journal.pone.0167587.
13. Li J, Li X, He D et al. Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor. Journal of Intelligent Manufacturing 2020; 31 1899-1916, https://doi.org/10.1007/s10845-020-01543-8.
14. Li J, Li X, He D et al. Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor. Journal of Intelligent Manufacturing 2020; 31 1899-1916, https://doi.org/10.1007/s10845-020-01543-8.
15. Li X, Li J, Zhao C et al. Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection. Mechanical Systems and Signal Processing 2020; 142 106740, https://doi.org/10.1016/j.ymssp.2020.106740.
16. Li Y, Wang K. Modified convolutional neural network with global average pooling for intelligent fault diagnosis of industrial gearbox. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (1): 63–72, https:// 10.17531/ein.2020.1.8.
17. Liu H, Zhou J, Xu Y et al. Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. Neurocomputing 2018; 315 412-424, https://doi.org/10.1016/j.neucom.2018.07.034.
18. Liu R, Meng G, Yang B et al. Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. IEEE Transactions on Industrial Informatics 2016; 13 1310-1320, https:// 10.1109/TII.2016.2645238.

19. Lu C, Wang Z Y, Qin W L et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. Signal Processing 2017; 130 377-388, https://doi.org/10.1016/j.sigpro.2016.07.028.
20. Nowakowski T, Komorski P. Diagnostics of the drive shaft bearing based on vibrations in the high-frequency range as a part of the vehicle's self-diagnostic system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2022; 24 (1): 70–79, http://doi.org/10.17531/ein.2022.1.9.
21. Peng D, Wang H, Liu Z et al. Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition. IEEE Transactions on Industrial Informatics 2020; 16 4949-4960, https:// 10.1109/TII.2020.2967557.
22. Peng Z K, Peter W T, Chu F L. An improved Hilbert–Huang transform and its application in vibration signal analysis. Journal of sound and vibration 2005; 286 187-205, https://10.1016/j.jsv.2004.10.005.
23. Ravikumar K N, Madhusudana C K, Kumar et al. Classification of gear faults in internal combustion (IC) engine gearbox using discrete wavelet transform features and K star algorithm. Engineering Science and Technology an International Journal 2022; 30 101048, https://doi.org/10.1016/j.jestch.2021.08.005.
24. Su W, Wang F, Zhu H et al. Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. Mechanical systems and signal processing 2010; 24 1458-1472, https://doi.org/10.1016/j.ymssp.2009.11.011.
25. Wang C, Gan M, Zhu C. A supervised sparsity-based wavelet feature for bearing fault diagnosis. Journal of Intelligent Manufacturing 2019; 30 229-239, https://doi.org/10.1007/s10845-016-1243-9.
26. Wang J, Li S, An Z et al. Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines. Neurocomputing 2019; 329 53-65, https://doi.org/10.1016/j.neucom.2018.10.049.
27. Wang Q, Zhao B, Ma H et al. A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion. Journal of Mechanical Science and Technology 2019; 33 2561-2571, https:// 10.1007/s12206-019-0504-x.
28. Woo S, Park J, Lee J Y et al. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) 2018; 3-19, https://link.springer.com/conference/eccv. https://doi.org/10.1007/978-3-030-01234-2_1
29. Wu X, Zhang Y, Cheng C et al. A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. Mechanical Systems and Signal Processing 2021; 149 107327, https://doi.org/10.1016/j.ymssp.2020.107327.
30. Xiong Q, Zhang W, Lu T et al. A fault diagnosis method for rolling bearings based on feature fusion of multifractal detrended fluctuation analysis and alpha stable distribution. Shock and Vibration 2016; https://doi.org/10.1155/2016/1232893.
31. Yang Y, Peng Z, Zhang W et al. Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances. Mechanical Systems and Signal Processing 2019; 119 182-221, https://doi.org/10.1016/j.ymssp.2018.07.039.
32. Ying Y, Li J, Chen Z et al.Study on rolling bearing online reliability analysis based on vibration information processing. Computers & Electrical Engineering 2018; 69 842-851, https://doi.org/10.1016/j.compeleceng.2017.11.029.
33. Yu J, Liu G. Knowledge extraction and insertion to deep belief network for gearbox fault diagnosis. Knowledge-Based Systems 2020; 197 105883, https://doi.org/10.1016/j.knosys.2020.105883.
34. Zhang J, Kong X, Li X et al. Fault diagnosis of bearings based on deep separable convolutional neural network and spatial dropout. Chinese Journal of Aeronautics 2022; 35(10):301-312, https://doi.org/10.1016/j.cja.2022.03.007.
35. Zhang J, Yi S, Liang G et al. A new bearing fault diagnosis method based on modified convolutional neural networks. Chinese Journal of Aeronautics 2020; 33 439-447, https://doi.org/10.1016/j.cja.2019.07.011.
36. Zhang K, Tang B, Qin Y et al. Fault diagnosis of planetary gearbox using a novel semi-supervised method of multiple association layers networks. Mechanical Systems and Signal Processing 2019; 31 243-260, https://doi.org/10.1016/j.ymssp.2019.05.049.
37. Zhang R, Peng Z, Wu L et al. Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence. Sensors 2017; 17 549, https://doi.org/10.3390/s17030549.
38. Zhu Z, Peng G, Chen Y et al. A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis. Neurocomputing 2019; 323 62-75, https://doi.org/10.1016/j.neucom.2018.09.050.