



A multiscale convolution neural network for bearing fault diagnosis based on frequency division denoising under complex noise conditions

Youming Wang¹ · Gongqing Cao¹

Received: 16 July 2022 / Accepted: 14 November 2022 / Published online: 30 December 2022
© The Author(s) 2022

Abstract

The condition of bearings has a significant impact on the healthy operation of mechanical equipment, which leads to a tremendous attention on fault diagnosis algorithms. However, due to the complex working environment and severe noise interference, training a robust bearing fault diagnosis model is considered to be a difficult task. To address this problem, a multiscale frequency division denoising network (MFDDN) model is proposed, where the frequency division denoising modules are presented to extract the detail fault features, and multiscale convolution neural network is employed to learn and enrich the overall fault features through two-scale convolution channels communication. The stacking convolution pooling layers are adopted to deepen the large-scale convolution channel and learn abundant global features. To remove the noise in the small-scale convolution channel, the frequency division denoising layers are constructed based on wavelet analysis to acquire the features of noise, where the input feature map is separated into high frequency and low-frequency features, and a sub-network based on attention mechanism is established for adaptive denoising. The superior features of MFDDN are the fusion of important fault features at each scale and custom learning of fine-grained features for the adaptive denoising, which improves the network feature extraction capability and noise robustness. This paper compares the performance of MFDDN with several common bearing fault diagnosis models on two benchmark bearing fault datasets. Extensive experiments show the state-of-the-art performance including robustness, generalization, and accuracy compared to the other methods under complex noise environment.

Keywords Convolutional neural network · Attention mechanism · Frequency division · Fault diagnosis · Denoising

Introduction

With the development of the automation and intellectualization of modern machinery, the operation monitoring and fault diagnosis of machine parts including rolling bearings, gears, rotors, etc., are vitally important for health assessment and safety management [1]. Generally, fault diagnosis methods of the rolling bearing are mainly based on signal processing and intelligent methods. The former methods such as the fast Fourier transformation (FFT) method [2, 3], spectral analysis [4, 5], or wavelet analysis [6–8] are common tools to extract useful information of operational state from noisy sensor data based on filtering, spectral estimation, statistical analysis, etc.

However, these methods rely on sufficient expert experience to choose appropriate parameters in the process for filtering, spectral estimation, statistical analysis, etc., according to the attributes of a signal, such as the amplitude, magnitude, frequency, phase, duration, shape, etc. [9] For the latter methods, including support vector machine (SVM) [10, 11], BP neural network [12, 13], artificial neural network [14], etc., they have become increasingly popular for fault detection and classification since these methods require little expertise for information extraction. Despite these methods have achieved favorable performance in fault diagnosis, it is still an arduous task to extract deep feature information from high-dimensional or complex nonlinear signals because they apply shallow structures to learn complicated features of the collected data.

Recently, deep learning (DL) method have been introduced as a powerful feature extraction and fault diagnosis tool for fault recognition. Since DL method employs multiple layers and complex deep structure to capture the information

✉ Youming Wang
xautroland@126.com

¹ School of Automation, Xi'an Key Laboratory of Advanced Control and Intelligent Process, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

with high dimensionality and nonlinearity, it is very suitable for the extraction of nonlinear and temporal features of big datasets [15]. Some typical DL methods such as the deep belief network (DBN), long-short term memory (LSTM), deep automatic encoder (DAE), etc. have achieved important results and developments in the fault diagnosis of bearings, gears, rotors, etc. [16–18]. As a representative deep learning algorithm, convolution neural network (CNN) has the capability of nonlinear mapping to extract the fault features based on its local perception and parameter sharing [19–22]. However, the noise in the collected signal usually blurs the spatial and temporal features, which causes many difficulties in feature extraction and fault recognition. Many CNN-based deep learning methods are proposed to improve the fault diagnosis performance of mechanical devices under complex noise conditions [23–28]. Although these methods can extract the fault feature of input data by network model under noise environment, it is difficult to deal with low signal-to-noise ratio (SNR) signals for fault diagnosis due to the fact that their spatial resolution is too coarse for CNNs to preserve the crucial multiscale features and remove noise from the signal, which has limited the accuracy of fault diagnosis in noisy conditions.

The emerging multiscale CNN methods employ the deep features at different resolutions and scales for the fault diagnosis under noise environment, which can extend the generalization of the features represented and fuse different feature maps to achieve higher performance. The multiscale CNN structure can focus on the temporal features of vibration signals at different scales and mine the vibration signal data features. The multiscale features usually contain different kinds of feature information, and convolutional kernels of different scale size are constructed in the multiscale structure to obtain different perceptual fields, which in turn are layered to learn features at different scales. Yu et al. [29] constructed a network model to extract multiscale features of vibration data under noise interference through the embedded multiscale attention mechanism. Zhang et al. [30] proposed a fault diagnosis method in the noisy environment based on the multiscale azimuth feature extraction model. Yu et al. [31] proposed a multiscale fusion global sparse network for extracting time-series features of vibration signals for fault diagnosis in complex environment. In the above mentioned techniques, the multiscale CNN-based fault diagnosis method is the very attractive way since the multiscale architecture can extract abundant fault features at different scales from raw input data and obtain more efficient learning capability, which is helpful for the further improvement of the performance of the clarification results.

Although multiscale CNN is considered as a practical and effective method, serious noise interference from the measuring system can reduce the distinguishing ability of fault features and undermine the feature differences between

various fault categories, which may lead to low accuracy and unreliability for the fault diagnosis. Therefore, a specific denoising operation is required for accurate feature extraction and fault detection in strong noise environments. Over the past few decades, traditional signal denoising methods have been proposed, including various time, frequency and time-to-frequency conversion methods, which can capture the desired signal by reducing the noise component [34–36]. As a classical signal denoising method, wavelet multi-resolution analysis reveals the local and salient features of the signals in the time–frequency domain. In the wavelet transform, the original input is decomposed into high-frequency and low-frequency components for denoising based on the wavelet function, where high frequencies are referred to as signal details and low frequencies are approximations of the signal. However, the selection of wavelet functions and denoising thresholds based on expert experience can be inaccurate, making the denoising effect inaccurate. To solve above problems, based on the wavelet denoising concept, we propose a two-channel multiscale convolution network based on frequency division denoising (MFDNN) for bearing fault diagnosis under complex noise conditions. Two channels including stacking convolution pooling layers and the frequency division denoising layers are established to extract global and local features of bearing vibration signals respectively, which are cascaded to fuse and capture important fault features at each scale. Considering the intrinsic multiscale feature of bearing vibration signals, the frequency division denoising layers based on wavelet analysis is presented to customize the learning of fine-grained features for the noise, where the input feature map is separated into high-frequency and low-frequency features, and a sub-network based on attention mechanism is established for adaptive denoising. The output of the multiscale model is sent to the fully connected layer for feature aggregation, and then served as input to the Softmax layer to obtain probabilistic classification results. The novelty of the proposed method lies in that the feature extraction ability of the proposed network is enhanced by stacking convolution kernels and its noise robustness is improved by a multiscale structure embedded with the frequency division denoising module.

The main contributions of our work are summarized as follows.

1. The frequency division denoising module based on feature scale and rotation invariance is proposed for the custom learning of fine-grained features for the noise and the training of the denoising thresholds for the removal of noise features.
2. Two large-scale and small-scale convolution kernel channels are constructed respectively to extract the global and local features of bearing vibration signals in complex

noisy environments, where the global and local features are fused by cascading to learn rich fault features for fault diagnosis.

3. The effectiveness of MFDDN model is verified in different datasets and compared with common models such as SVM, CNN, etc. in terms of accuracy, robustness and other parameters to verify the superiority of the proposed model.

The rest of this paper is described as follows. The relevant theoretical background of this study is presented in “Preliminaries”. In “Proposed method”, the frequency division denoising module and multiscale convolution neural network based on frequency division denoising are described. In “Experiment and analysis”, the generalization of MFDDN is verified with different bearing datasets and the superiority of the model is demonstrated through comparative experiments. The conclusions are presented in “Conclusion”.

Preliminaries

Multiscale Convolutional Neural Network

As a classical deep learning structure, a convolutional neural network, or CNN, employs a direct connection of convolutional and pooling layers for feature extraction, where the fault features at deep and abstract levels can learn and identify fault patterns. CNNs are aggregated for the invariance of multi-stage features in classification tasks, and the discriminative power of the network is enhanced by deepening the number of network layers through multi-layer stacking. However, deepening the network may lead to the possible vanishing or explosion in gradient backward conduction, which increases the fitting difficulty. To solve this problem, a kind of multiple parallel CNN structures is proposed to merge the feature extraction results of each branch by diverse convolution kernels and reduces the depth of the network. As multiscale CNN simply multiply the feature extraction of multichannel signals, it has unique ability to allocate different feature learning polices to different components of the input signals [32]. As shown in Fig. 1, the top convolution layers are adopted to learn high-level global information, and bottom layers are employed to capture the low-level detail information in the multiscale CNN. Compared to traditional single-scale CNN, multiscale CNN employs the top-layer convolutional layer to capture more specific input feature information and extract low-level abstract features by deep convolutional kernels for learning detailed information, which leads to better classification and robust performance in fault diagnosis [33].

Wavelet analysis

As a powerful and effective tool for the representation of vibration signals, the wavelet threshold denoising method is popular in signal processing, where the threshold function is employed to eliminate out-of-range noise in the wavelet coefficients [24]. Since useful signals usually appear as low frequency and noise signals appear as high-frequency signals in practice, the quality of denoising depends on the selection of threshold functions in the processing of the signals with noise in the high-frequency wavelet coefficients. So far, hard and soft threshold functions are common applied with ease of operation in the wavelet denoising. The hard threshold function sets the decomposition coefficients smaller than the threshold to zero in different scale spaces and retains the decomposition coefficients larger than the threshold, which will lead to some fluctuations in the reconstruction of the original signal due to the fixed threshold setting. The soft threshold function is to smooth the decomposition coefficients to avoid fluctuations which exceed the threshold range as as given by Eq. (1)

$$y = \begin{cases} x - \delta & x > \delta \\ 0 & -\delta \leq x \leq \delta \\ x + \delta & x < -\delta \end{cases}, \quad (1)$$

where x is the input, y is the output, δ is the threshold parameter. The derivative of the soft thresholding function is

$$\frac{\partial y}{\partial x} = \begin{cases} 1 & x > \delta \\ 0 & -\delta \leq x \leq \delta \\ 1 & x < -\delta \end{cases}. \quad (2)$$

It can be seen from Eq. (2) that the derivative of the soft threshold has a similar functions to the ReLu activation function, which can prevent gradient disappearance or gradient explosion.

Attention mechanism

To solve the limitations in CNN feature extraction operation and improve the training speed of the model, an attention mechanism is adopted to choose the important parts for processing and ignore less important parts, which can learn the weight distribution between feature maps. The basic structure of the channel attention mechanism is shown in Fig. 2 [37]. The operation of the attention module is mainly divided into two steps, including the squeeze for shrinking the input data from the previous layer and the excitation for the acquisition of the correlation between different channels. In channel attention mechanism, the global average pooling and maximum pooling are adopted to aggregate the compressed one-dimensional feature information and the fully

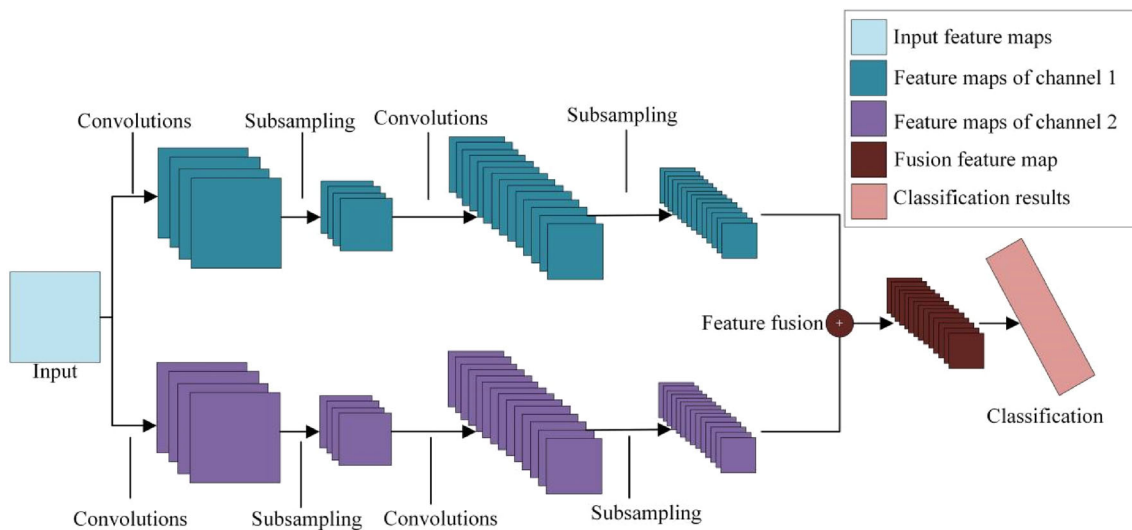


Fig. 1 Classical structure of multiscale CNN

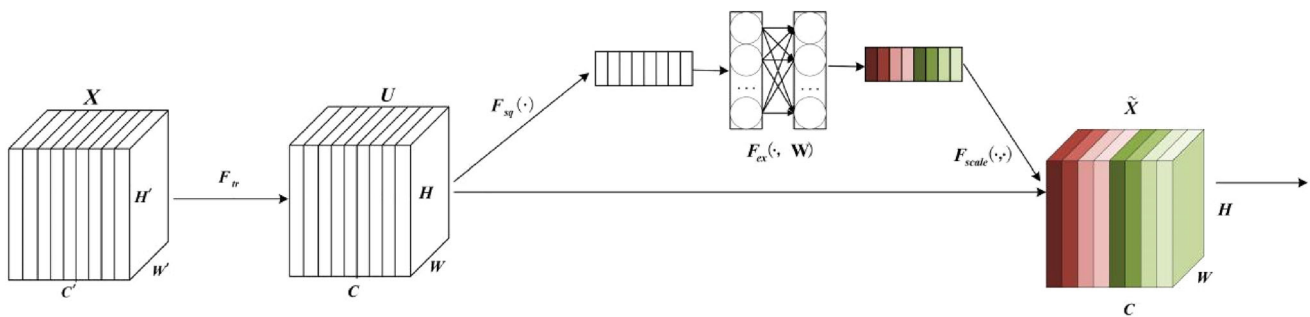


Fig. 2 Basic structure of channel attention mechanism

connected layer is employed to multiply the input by the weight matrices.

The weight of each channel is obtained by compressing the spatial dimension of the feature map in the form

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \tag{3}$$

where H is the length of the input feature, W is the width of the input feature, u_c is the input feature, c is the channel information, z_c is the result of the squeeze function, F_{sq} is the squeeze function.

Generally, the multi-layer perceptron (MLP) network is to adjust the weights of each channel and the final weight matrix can be obtained by Sigmoid function in the form of

$$S_c = F_{ex}(Z, W) = \sigma[g(Z, W)] = \sigma[W_2 \delta(W_1 Z)], \tag{4}$$

where Z is the result of the squeeze function, W_1 and W_2 are the weights of hidden cells, δ is the activation function, σ is the excitation function, F_{ex} is the excitation function.

The weight matrix is adjusted to each channel of the input feature map by the function F_{scale} and the final weighted feature map can be obtained as

$$\hat{X}_c = F_{scale}(X_c, S_c) = X_c \otimes S_c, \tag{5}$$

where \otimes is element by element multiplication, X_c is the input feature map, S_c is output of the F_{ex} , \hat{X}_c is the output feature map combine with weight matrix.

This simple attention architecture makes it possible to assign weights to each feature channel, which allows the model to focus on the useful information and ignore the less important features to improve the learning performance of the network.

Proposed method

Inspired by the soft threshold denoising in wavelet analysis, a multiscale frequency division denoising network (MFDDN) model is proposed, where the frequency division denoising modules are integrated into the multiscale CNN model

to learn the fine-grained features of the signal and realize adaptive denoising based on the attention mechanism. The convolution kernels are applied to perform frequency division of the feature maps and allow to extract fine-grained features for signal denoising. Similarly, the sub-network based on attention mechanism are designed to allocate the adaptive threshold for each channel and remove noise-related feature maps. In addition, the cascading operations in multi-scale CNN are introduced to fuse the global and local fault features to capture rich fault features. Finally, the extracted features from the multiscale network are subsequently captured by the feature fusion layer and sent to the Softmax layer for fault classification and diagnosis.

Frequency division denoising module

To overcome the learning difficulties of traditional multiscale CNN for the fault diagnosis under strong noise conditions, the frequency division denoising method is presented to divide the high and low-frequency information in the feature map and construct to perform independent denoising, where the low-frequency component of the input feature map mainly contains the useful feature information of the signal, and the high-frequency component mainly contains the feature of the noise. According to the scale and rotational invariance of the features, the feature maps can be divided at different spatial frequencies to refine its features, which provides customized feature extraction [38]. To comprehensively remove the fault features, the attention mechanism is employed to construct a sub-network for adaptive denoising based on the concept of wavelet soft thresholding.

The structure of the frequency division denoising module is shown in Fig. 3. The frequency division denoising module is described as follows, which can be divided into two parts. To enable efficient inter-frequency communication, the convolution kernel W is divided into two components $W = [W^H, W^L]$, and two convolution kernel component can be divided into processing units of different frequencies $W = [W^{H \rightarrow H}, W^{H \rightarrow L}, W^{L \rightarrow H}, W^{L \rightarrow L}]$. High-frequency features are fused with low-frequency features after down-sampling. Similarly, the low-frequency features are fused with the high-frequency features by up-sampling, as shown in Fig. 3. The second part is employed to implement adaptive threshold denoising. The input features are compressed through a global average pooling layer to aggregate features. The scaling parameter of each channel is set by the MLP network. To ensure efficient threshold denoising of frequency division denoising module, the scaling parameter is set to a range of 0 to 1 by the Sigmoid function, and the scaling parameter is multiplied by the absolute value of the feature map.

After the two-dimensional convolution kernel operation, the input feature map can be represented as $X^{H \times W \times C}$,

where H and W are the dimensions of the input feature map, C is the number of channels of the feature map. The length and width of low-frequency features are set to half of the length and width of high-frequency features, respectively. The channels with different frequency features are partitioned by frequency division parameters α , $\alpha \in [0, 1][0, 1]$ which means that the number of channels corresponding to the low-frequency component is αC , and the number of channels corresponding to the high-frequency feature component is $(1 - \alpha)C$. The feature maps of different frequencies are

$$X_L = X^{0.5H \times 0.5W \times \alpha C}, \tag{6}$$

$$X_H = X^{H \times W \times (1-\alpha)C}, \tag{7}$$

where X_L and X_H are the low-frequency and high-frequency components in the feature map respectively.

The feature information of different frequencies at the position (p, q) of the input feature map is learned by a convolution kernel in the form of

$$\begin{aligned} Y_{p,q}^H &= Y_{p,q}^{H \rightarrow H} + Y_{p,q}^{L \rightarrow H} = \sum_{i,j \in N_k} W_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{H \rightarrow H} X_{p+i, q+j}^H \\ &\quad + \sum_{i,j \in N_k} W_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{L \rightarrow H} X_{(\lfloor \frac{p}{2} \rfloor + i)(\lfloor \frac{q}{2} \rfloor + j)}^L, \\ Y_{p,q}^L &= Y_{p,q}^{L \rightarrow L} + Y_{p,q}^{H \rightarrow L} = \sum_{i,j \in N_k} W_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{L \rightarrow L} X_{p+i, q+j}^L \\ &\quad + \sum_{i,j \in N_k} W_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{H \rightarrow L} X_{(\lfloor \frac{p}{2} \rfloor + i + 0.5)(\lfloor \frac{q}{2} \rfloor + j + 0.5)}^H, \end{aligned} \tag{8}$$

where the sampling operation is represented by a factor of 2 and the dimensionality of the feature map is consistent by moving half a step, $Y_{p,q}^H$ and $Y_{p,q}^L$ are the high-frequency output and low-frequency output of the current position, respectively.

The frequency-divided feature maps are applied as the input to the sub-network for denoising. The squeeze function performs a global average pooling of the input high-frequency and low-frequency features respectively as

$$\begin{aligned} y_c &= F_{sq}(Y_c^H) = \frac{1}{H \times W} \sum_{p=1}^H \sum_{q=1}^W Y_c^H(p, q) \\ y_{c'} &= F_{sq}(Y_{c'}^L) = \frac{1}{0.5H \times 0.5W} \sum_{p=1}^{0.5H} \sum_{q=1}^{0.5W} Y_{c'}^L(p, q), \end{aligned} \tag{9}$$

where F_{sq} is the squeeze function to carry out global average pooling on the input feature map, Y_c^H and $Y_{c'}^L$ are the outputs of global average pooling. The output of the global average

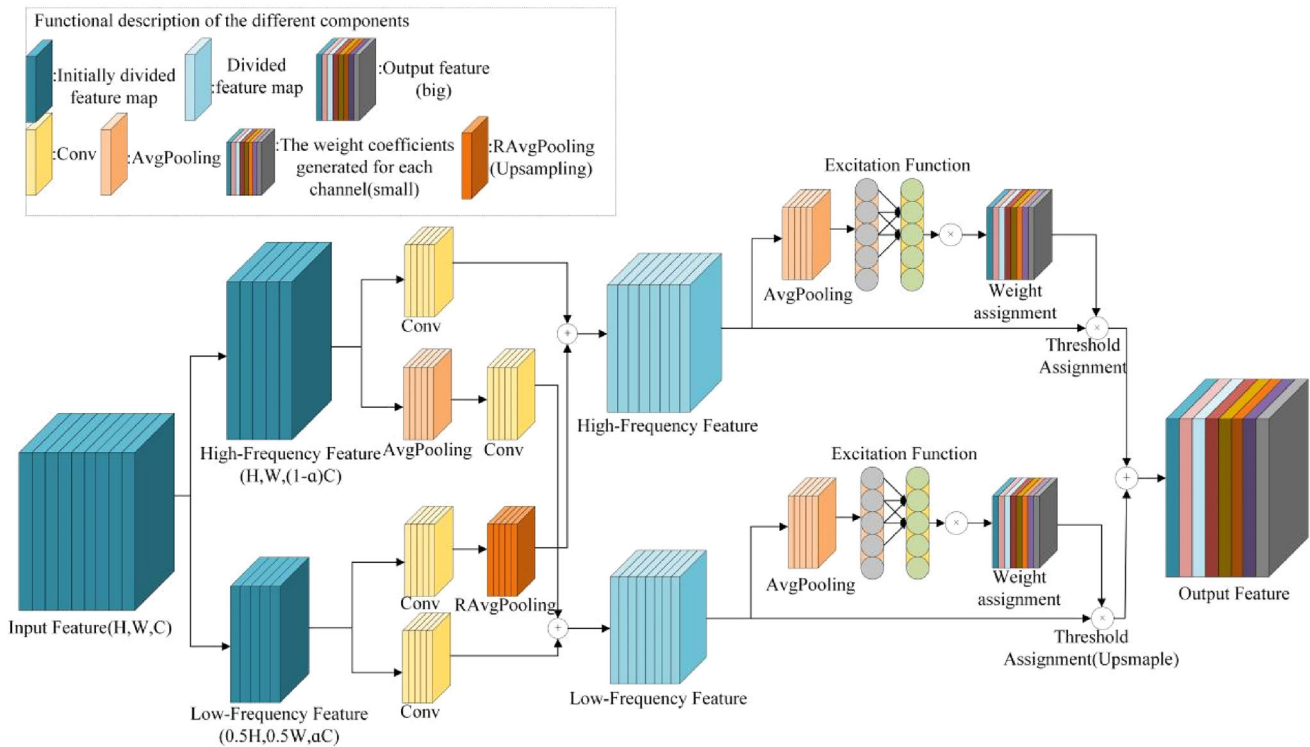


Fig. 3 Frequency division denoising module

pooling is obtained by the nonlinear transformation of a fully connected network in the form of

$$\begin{aligned}
 S &= F_{ex}(y, W) = \sigma(g(y, W)) = \sigma(W_2\delta(W_1y)) \\
 S' &= F_{ex}(y', W) = \sigma(g(y', W)) = \sigma(W_2\delta(W_1y')),
 \end{aligned}
 \tag{10}$$

where F_{ex} is the excitation function, W_1 and W_2 are the parameters of the full connection layer, δ is the ReLu activation function, σ is the Sigmoid activation function, F_{scale} is weighted with the input feature map in the form of

$$\begin{aligned}
 \hat{Y}^H &= F_{scale}(S) = SY^H \\
 \hat{Y}^L &= F_{scale}(S') = S'Y^L,
 \end{aligned}
 \tag{11}$$

where S and S' are denoising threshold matrix, Y^H and Y^L are the high and low-frequency features, \hat{Y}^H and \hat{Y}^L are high and low-frequency feature maps with denoising thresholds assigned, respectively. The output is obtained by the fusion of the high and low-frequency components in the feature map as

$$Y = \hat{Y}^H + f_{upsample}(\hat{Y}^L),
 \tag{12}$$

where $f_{upsample}$ is the up-sampling operation, and Y is the output result of the frequency division denoising module.

Multiscale network based on frequency division denoising

Aiming at the problem of low identification ability of the model in the feature extraction, we propose a customized learning method to extract specific fine-grained features, which can improve network accuracy and enhance network robustness in complex noise environment. A multiscale network is constructed to extract the multiscale features of the bearing vibration signal, where the two convolutional kernels with different scales are constructed to extract features of different scales, as shown in Fig. 4. The large-scale convolutional kernels are employed to extract global features through large perceptual fields, and stacked convolutional layers are constructed to enhance the feature extraction capability of the network. In the small-scale convolution channel, the features are extracted by convolution kernel to focus on the local feature information and enhance the denoising capability of the network. The local feature of small-scale convolution channels is refined by stacking multi-layer frequency division denoising modules, and the number of customized modules is determined through experiments guided by model accuracy. We comprehensively discuss the number of frequency division denoising modules in the model evaluation. It follows that the classification accuracy of multiscale CNN under noisy conditions can be enhanced by embedding multiple

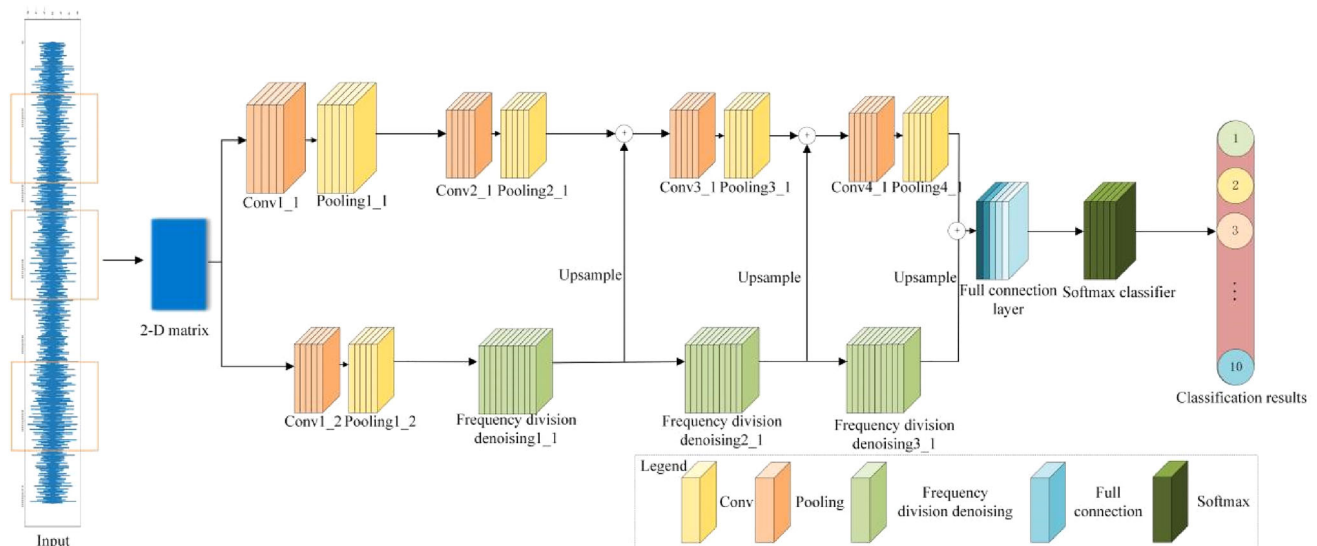


Fig. 4 Structure of MFDDN model

Table 1 Parameters of the MFDDN model

Serial number	Convolution kernel size	Output size	Zero padding
Conv1_1	$5 \times 5 \times 16$	$16 \times 16 \times 16$	Is
Pool1_1	$2 \times 2 \times 16$	$8 \times 8 \times 16$	No
Conv1_2	$3 \times 3 \times 16$	$16 \times 16 \times 16$	Is
Pool1_2	$2 \times 2 \times 16$	$8 \times 8 \times 16$	No
Conv2_1	$3 \times 3 \times 32$	$8 \times 8 \times 32$	Is
Pool2_1	$2 \times 2 \times 32$	$4 \times 4 \times 32$	No
FrequencyConv1_2	$3 \times 3 \times 32$	$4 \times 4 \times 32$ $2 \times 2 \times 32$	Is
Conv1_3	$3 \times 3 \times 64$	$4 \times 4 \times 64$	Is
Pool1_3	$2 \times 2 \times 64$	$2 \times 2 \times 64$	No
FrequencyConv2_2	$3 \times 3 \times 64$	$2 \times 2 \times 64$ $1 \times 1 \times 64$	Is
Conv1_4	$3 \times 3 \times 64$	$2 \times 2 \times 64$	Is
Pool1_4	$2 \times 2 \times 64$	$1 \times 1 \times 64$	No
FrequencyConv3_2	$3 \times 3 \times 64$	$1 \times 1 \times 64$	Is
Fcn	256×1	256×10	No
Softmax	10	10	/

frequency division denoising layers in the small-scale convolutional channel. To extract rich feature information, the feature information between different channels is fused by cascade operation, thus further improving the feature extraction capability of the model for the multiscale feature.

The MFDDN network for bearing fault diagnosis is described as follows. A multiscale network is constructed by two convolutional channels at different scales. The large-scale convolution channel contains four convolution pool layers to improve the learning capacity of the network

and the small-scale convolution channel contains three frequency division denoising modules. The scaling the output feature map of the frequency division denoising module is transmitted to the large-scale convolution channel for the feature fusion of different channels. In the feature fusion layer, the features of the two convolution channels are fused and transferred to the full connection layer. Finally, the Softmax classifier is introduced to output the fault classes and distribution probabilities.

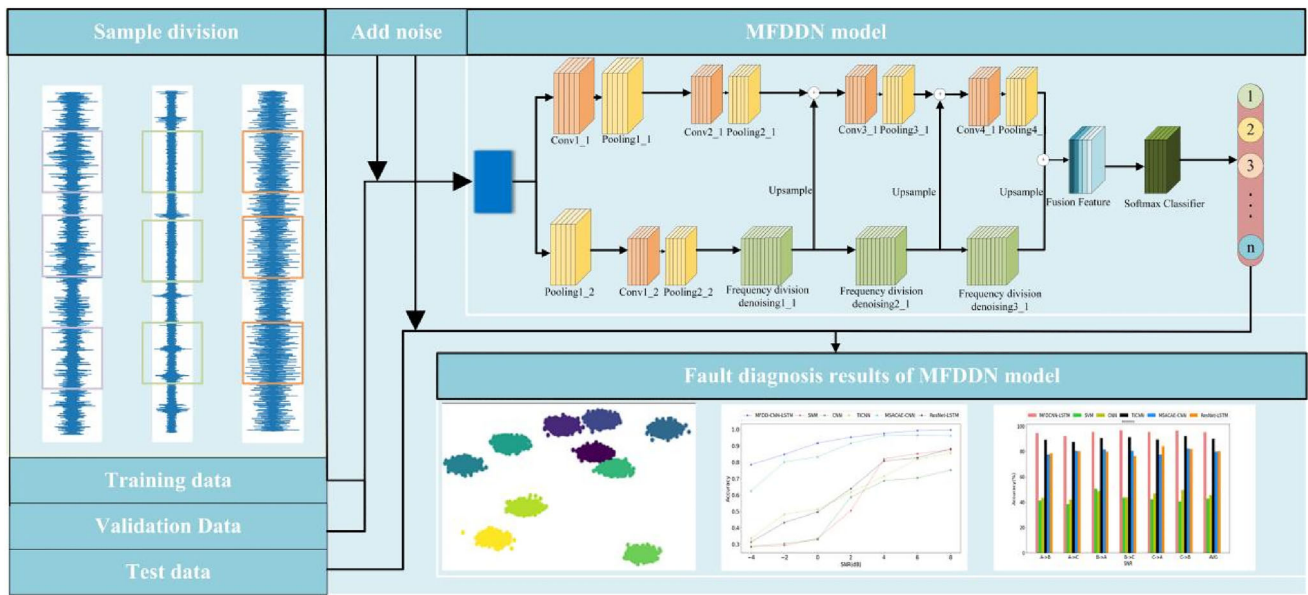


Fig. 5 MFDDN framework

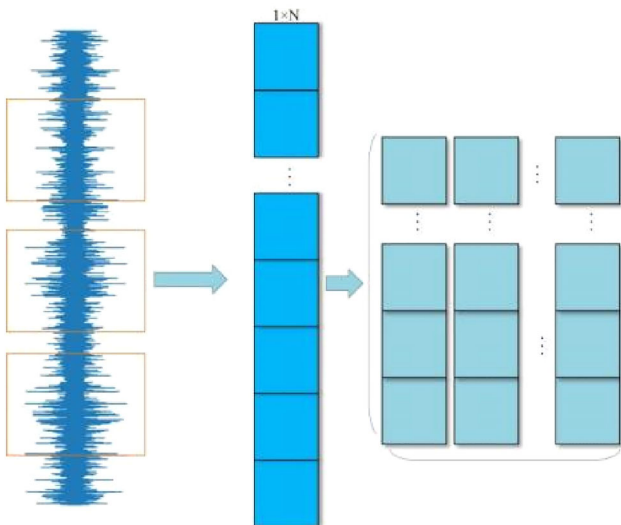


Fig. 6 Two-dimensional matrix transformation

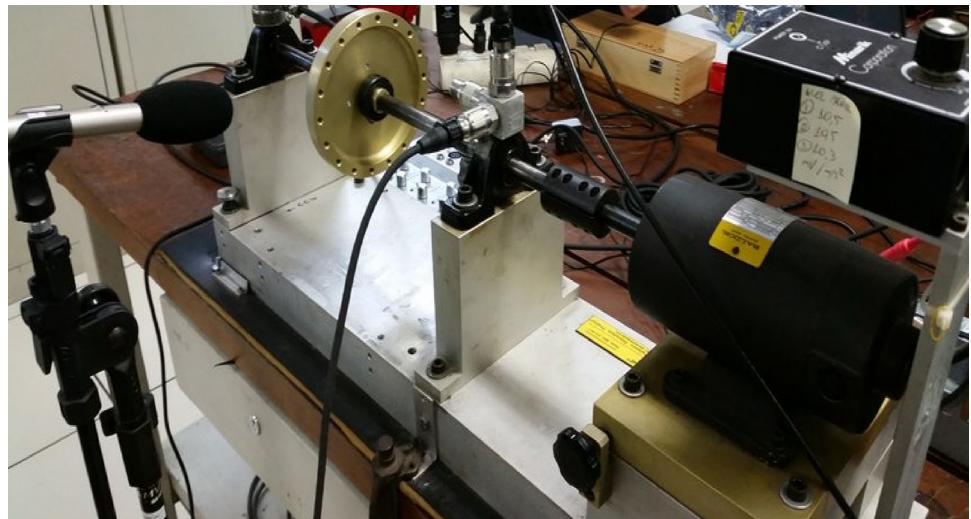
The parameters of the MFDDN model are shown in Table 1. A two-channel multiscale network is constructed to extract features through convolution kernels at different scales. In the large-scale convolution channel, a 5×5 convolution kernel is employed for feature extraction to obtain global features, and the 3×3 convolution kernels are to increase the depth to enhance feature extract ability. In the small-scale convolution channel, a 3×3 convolution kernel is introduced to improve the nonlinear fitting ability of the network, where the stacking frequency division denoising modules are employed to eliminate the noise of the feature map. The output features from

the two-channel multiscale network are sent to the full connection layer for feature fusion. The ReLu activation function is adopted to obtain nonlinear features and the Softmax function is applied for classification.

The MFDDN framework

To sum up, a multiscale convolutional neural network based on frequency division denoising is proposed for fault diagnosis in complex noise environments. As shown in Fig. 5, the architecture of the MFDDN can be divided into three sub-modules, including the data processing module, the training module, and the fault diagnosis module. In the data processing module, the vibration signals are overlapping sampled and transformed into the 2-dimensional matrix as the input, including training dataset, validation dataset, and test dataset. The MFDDN model learns the correlation between the data from the input data in the training module, where a multiscale architecture is constructed to learn the multiscale features between of the vibration data, and the frequency division denoising modules are embedded in the network to customize the learning of fine-grained features for the adaptive denoising. In the fault diagnosis module, the test set data is used to evaluate the diagnosis accuracy of the MFDDN model. Unlike traditional CNN-based fault diagnosis models, the MFDDN model introduces frequency division denoising modules to remove noisy features from the feature map and a multiscale CNN with a cascade structure was constructed to extract richer fault features.

The specific program for the MFDDN model is as follows.

Fig. 7 The acquisition device of MaFaulDa**Table 2** Specification parameters of the MaFaulDa experimental setup

The specification	Specification parameters
The motor	1/4 CV DC
Frequency range	700–3600 rpm
System weight	22 kg
Shaft diameter	16 mm
Axis length	520 mm
The rotor	15.24 cm
Bearing span	390 mm
Ball number	8
The ball diameter	0.7145 cm
Basic standard frequency	1.8710 CPM/RPM
Frequency of outer ring failure	2.9980 CPM/RPM
Internal loop failure frequency	5.0020 CPM/RPM

1. The input data are transformed into a two-dimensional matrix, which can be divided into training dataset, validation dataset and test dataset.
2. The training dataset is sent into the MFDDN model for training. The early stop method or the performance of the validation dataset is introduced to be the stopping criterion of the training process.
3. The test dataset is sent to the trained network through the Softmax layer for probabilistic classification and the diagnostic results are obtained.

Experiment and analysis

In this section, the viability of the MFDDN model is demonstrated to explore the performances of MFDDN and verify the effectiveness of its diagnosis on two bearing datasets,

including bearing fault data set from Rio de Janeiro Federal University (MaFaulDa) [40] and Case Western Reserve University (CWRU) [39]. The MFDDN is implemented in Python 3.6 and TensorFlow 1.14 on Windows 64-bit operating system with the CPU Core is i5-7200 and 8G RAM.

Data preprocessing

The input data is standardized by overlapping sampling method as

$$x = \frac{x' - \mu}{\sigma}, \quad (13)$$

where the input can be expressed as $x' = [x_1, x_2, \dots, x_N]$, N is the number of sampling points, the μ is the mean value of the input, σ is the standard deviation of the input. The μ and σ in the form of

$$\mu = \frac{1}{N} \sum_{i=1}^D x'(i), \quad (14)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x'(i) - \mu)^2}, \quad (15)$$

where D is the number of data of each sample. As the two-dimensional input can contain more correlations of the original vibration information, the standardized vibration data is converted into the two-dimensional matrix, as shown in Fig. 6.

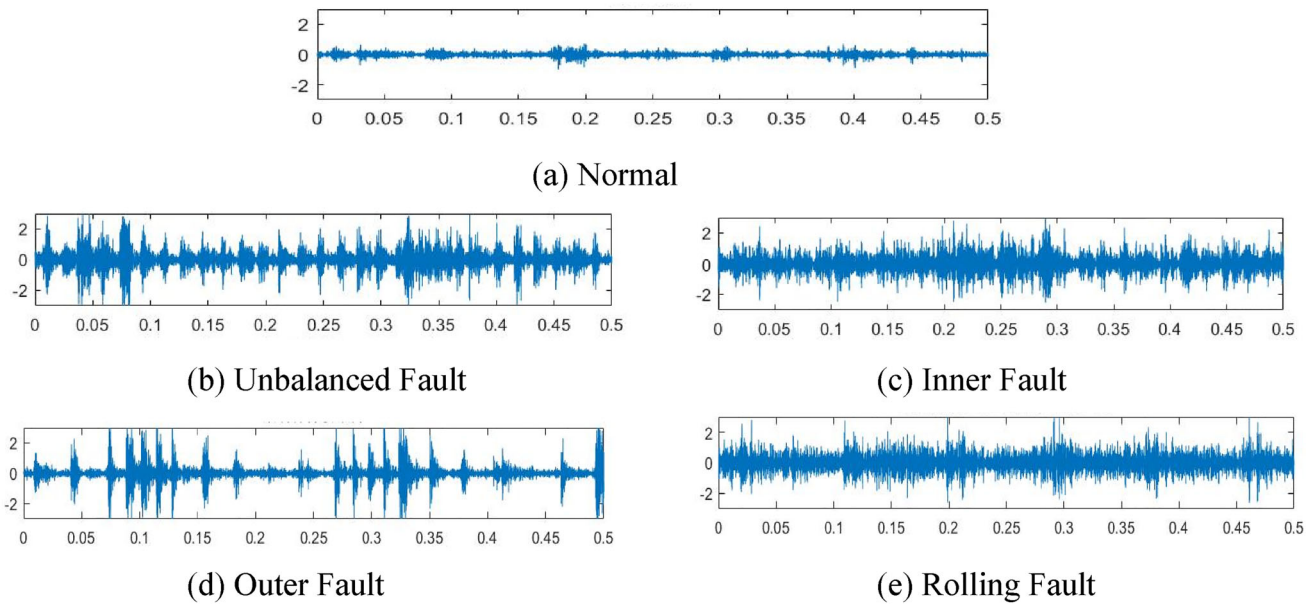


Fig. 8 Time domain waveform of MaFaulDa database under different faults

Case 1: Validation experiment with MaFaulDa

Data description and parameters setting

The MaFaulDa bearing data acquisition setup is shown in Fig. 7, including three industrial monitoring instrumentation sensors, a type 601A01 accelerometer (axial, radial and tangential), a motor, a tachometer to measure the rotation frequency of the system, and a microphone to capture the sound of the system.

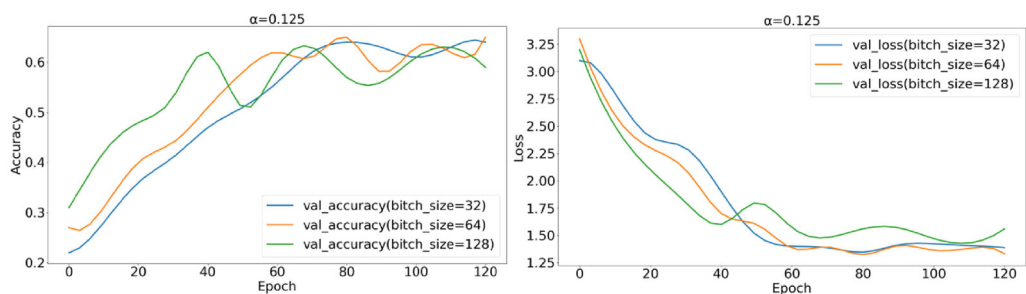
The basic specifications of the bearing failure test stand are shown in Table 2 and the time domain diagram is shown in Fig. 8. The collected bearing dataset is labeled from 0 to 4 corresponding to five working states, including normal state, unbalanced fault, outer ring fault, inner ring fault, and rolling body fault. For each label, there are 500 samples for each state and 1024 sampling points for each sample. In this experiment, the fivefold cross-validation experiment was introduced to evaluate the model. The data set was divided into five parts, four of which were used as the training set and one as the test set. To adjust the parameters in the model, the test set was divided into two parts, including validation set and test set. Therefore, the dataset was partitioned into training set, validation set, and test set with a ratio of 8:1:1. The validation set with the same distribution as the test set is constructed to estimate the training level of the model, and the test set of equal size is provided to test the performance of the model. The vibration data is transformed into a two-dimensional matrix to acquire abundant feature information for feature extraction and fault classification, where the size

is converted from 1×1024 to 32×32 as input to the model for each sample.

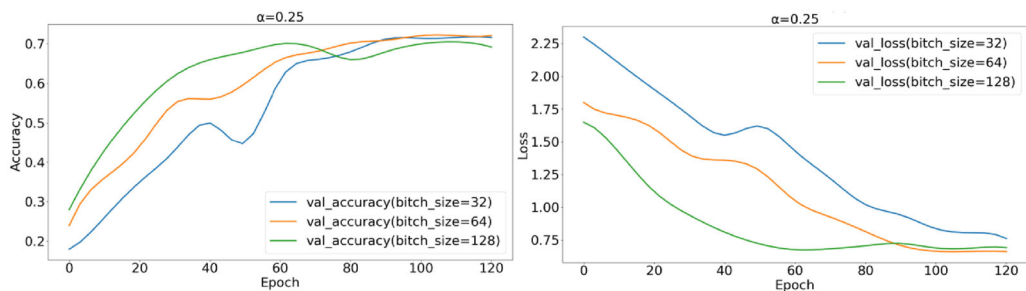
In the training process, Adam optimizer is introduced to optimize the parameters since it has strong generalization capabilities in the presence of gradient noise. To improve the network convergence and classification performance of MFDDN, the experiments on the selection of hyperparameters are conducted under the noise environment. The accuracy and loss values of the model with different epochs and batch size are introduced to verify in Fig. 9, and the maximum classification accuracy with different frequency division coefficients α are shown in Table 3.

As shown in Fig. 9, the classification accuracy for MFDDN model reaches the maximum when the frequency separation parameter α is 0.5. It can be seen that the accuracy of the model has been improved with the increase of the epochs and the loss has reached a minimum when the batch size is set to 64 and the epochs is 100.

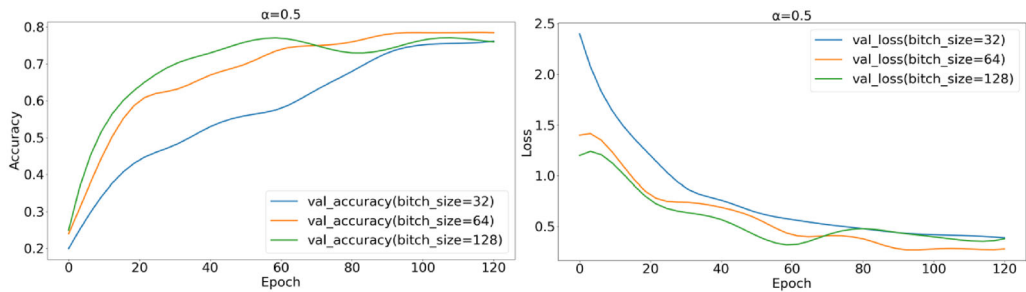
As can be seen in Table 3, the classification accuracy for the model reaches the highest when frequency separation parameter α is 0.5, due to its capability to distinguish the noise features in the fault features. When the frequency separation parameter α is smaller than 0.5, more low-frequency features are separated into the high-frequency features and many useful fault features are removed during the denoising. On the other hand, a large number of high-frequency features are mixed with low-frequency features, which results in the failure of the feature extraction for the noise. Therefore, the frequency separation parameter α of the MFDDN model is set to 0.5, the batch size is set to 64, and the epochs is set to 100.



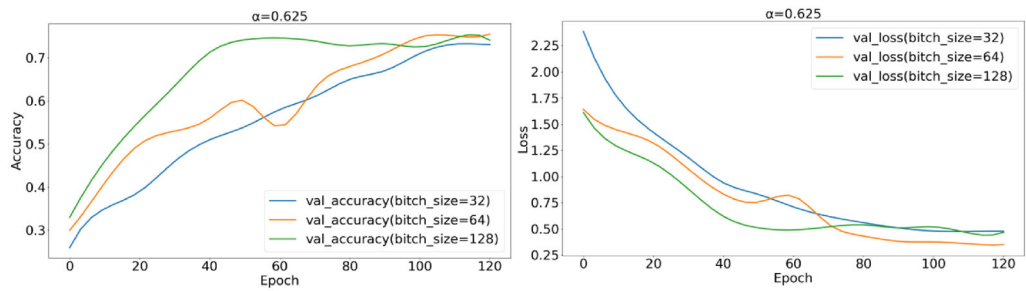
(a) Accuracy of different batch_size under epoch increment (b) Loss of different batch_size under epoch increment



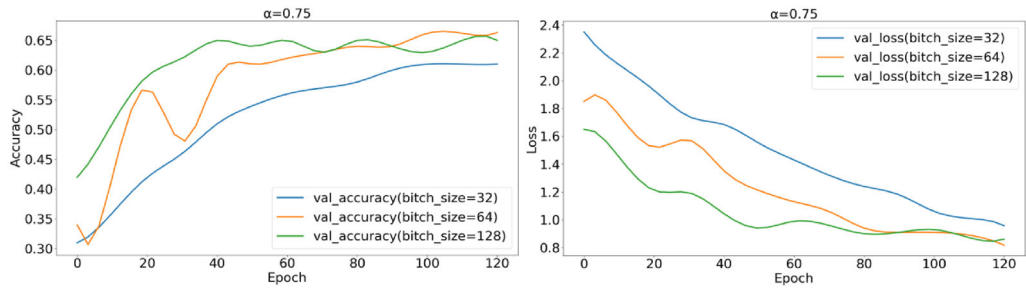
(c) Accuracy of different batch_size under epoch increment (d) Loss of different batch_size under epoch increment



(e) Accuracy of different batch_size under epoch increment (f) Loss of different batch_size under epoch increment



(g) Accuracy of different batch_size under epoch increment (h) Loss of different batch_size under epoch increment



(i) Accuracy of different batch_size under epoch increment (j) Loss of different batch_size under epoch increment

Fig. 9 Accuracy graphs and loss function graphs with different hyperparameters (SNR = - 4)

Table 3 Maximum accuracy and minimum loss of MFDDN model

Frequency division parameters	$\alpha = 0.125$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.625$	$\alpha = 0.75$
Maximum accuracy(SNR = - 4)	65.3%	71.4%	78.6%	74.1%	66.7%
Minimum loss(SNR = - 4)	1.332	0.662	0.283	0.467	0.861

Bold represents the optimal performance under different conditions

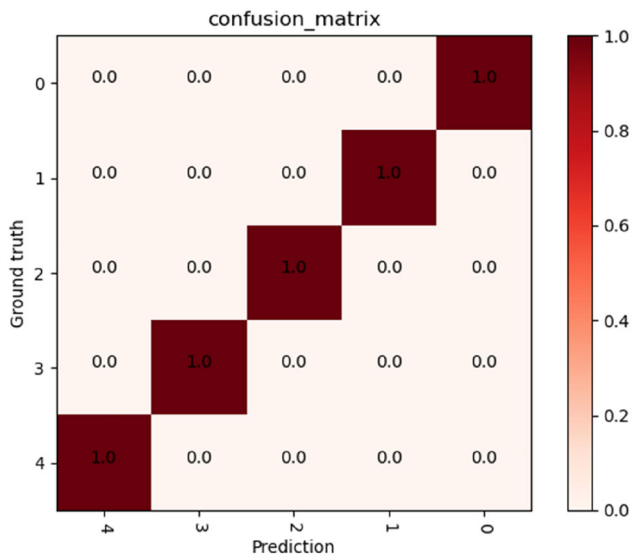


Fig. 10 Confusion matrix for the noiseless MaFaulDa dataset

Noiseless experimental analysis

In this section, the noiseless vibration signal is exploited to verify the performance of MFDDN. The diagnostic accuracy of the model is 100% without noise, as shown in Fig. 10. Without noise interference, the MFDDN model has strong feature extraction capability and achieves excellent diagnostic accuracy.

Performance verification experiment in noise environment

Figure 11 and Table 4 compare the fault diagnosis accuracy of MFDDN, SVM, Random Forest, KNN, CNN, a CNN model based on training Interference(TICNN) [26], a multiscale CNN model based on denoising auto-encoders (MSACAE-CNN) [18], a hybrid model of LSTM and ResNet (ResNet-LSTM) [41], and an improved multiscale CNN model combining feature attention mechanism (IMS-FACNN) [42] for the input signal with - 4 to 8 dB Gaussian white noise.

It can be seen from Fig. 11 and Table 4 that the diagnostic accuracy of the MFDDN method is superior to other models under different SNR from - 4 to 8. Specifically, the MFDDN has satisfactory improvement compared to the other models at SNR of - 4. With the increase of SNR,

the model accuracy has been ahead of other models, which indicates that MFDDN has shown excellent denoising ability and feature extraction ability through customized design. The above experimental results show that multiscale features are extracted for learning and the MFDDN network performance is satisfactory by customized fine-grained features.

The confusion matrix is introduced to explore the diagnosis results of different fault types of the MFDDN under the SNR of - 4, 0, 4, and 8 respectively in Fig. 12. As can be seen from Fig. 12, MFDDN has higher classification accuracy for the data with fault labels of 1 in low SNR. With the improvement of SNR, the classification accuracy of MFDDN for each fault has been significantly improved. The results show that the MFDDN model refines the features through the frequency division denoising module to learn the effective feature information in a customized learning way to improve the model accuracy.

Case 2: Validation experiment with CWRU

Data description

The rolling bearing data acquisition center is shown in Fig. 13. The bearing data acquisition device consists of a 1.5 KW (2 HP) motor, a torque sensor (decoder), and a power meter which was installed on the drive end bearing seat. The acceleration sensor is employed to collect the vibration acceleration signal of the bearing fault. In this experiment, the speed of the test bearing is set to 1730–1797 rpm. There are three types of faults with diameters of 0.007 in., 0.014 in., and 0.021 in. The bearing operating conditions for complex work can be obtained by adding different loads to the motor, including 0HP, 1HP, 2HP, and 3HP. The signals in time domain of 10 bearing working states are shown in Fig. 14.

Noiseless experimental results analysis

The each bearing state consists of 1000 samples, where each sample consists of 1024 sequential data points of the original vibration signal. For each sample, 1×1024 one-dimensional data were converted into 32×32 two-dimensional feature matrices to retain the correlation between raw input data.

Fig. 11 Diagnostic accuracy of different models under different noise conditions

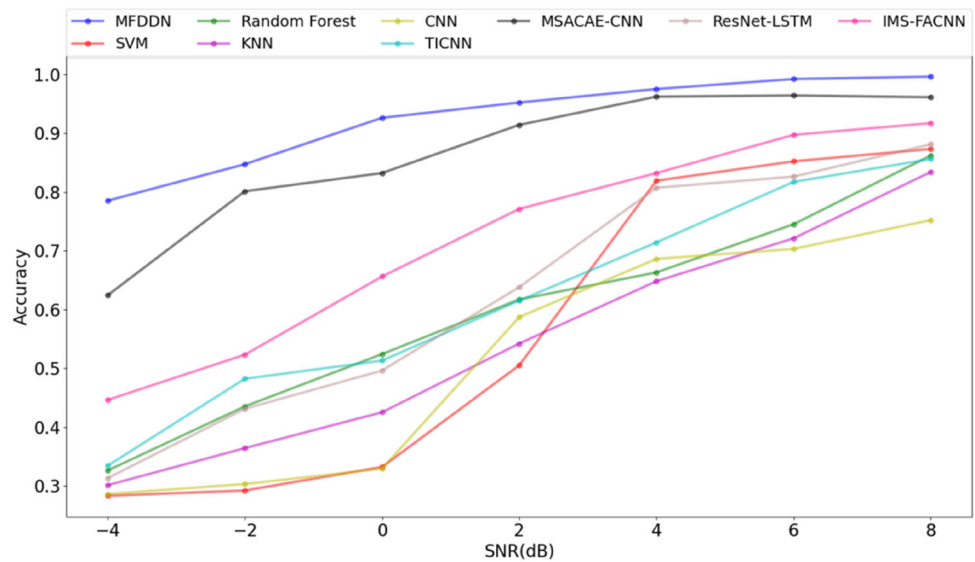


Table 4 Results of different network model under different SNR

Fault diagnosis model	SNR = - 4 (%)	SNR = - 2 (%)	SNR = 0 (%)	SNR = 2 (%)	SNR = 4 (%)	SNR = 6 (%)	SNR = 8 (%)	Avg (%)
MFDDN	78.6	84.7	92.4	95.2	97.5	98.2	99.2	92.2
SVM	28.3	29.2	33.2	50.5	81.9	85.2	87.3	56.5
Random Forest	32.6	43.5	52.4	61.7	66.3	74.5	86.2	59.6
KNN	30.1	36.4	42.5	54.2	64.8	72.1	83.4	54.7
CNN	28.6	30.3	32.9	58.7	68.6	70.3	75.2	52.1
TICNN	33.4	48.2	51.3	61.5	71.4	81.7	85.6	61.8
MSACAE-CNN	62.4	80.1	83.2	91.4	96.2	96.4	96.1	86.5
ResNet-LSTM	31.3	43.1	49.6	63.8	80.7	82.6	88.1	62.7
IMS-FACNN	44.6	52.3	65.6	77.1	83.2	89.7	91.7	72.0

Bold represents the optimal performance under different conditions

Without noise interference, the average recognition accuracy of the MFDDN achieves 100%. The diagnostic confusion matrix results in the noiseless state in Fig. 15 indicate that the dataset are correctly classified.

Performance verification experiment in noise environment

The diagnostic accuracy of MFDDN model at the different SNR are shown in Fig. 16 and Table 5. Table 5 depicts the comparison of the diagnostic accuracy of MFDDN with other state-of-the-art models at different SNR.

The MFDDN performances excellent performance in different noise environments, as shown in Table 5. The average diagnostic accuracy of the proposed model is superior to SVM, KNN, Random Forest, CNN, MSACEAE-CNN, TICNN, ResNet-LSTM, IMS-FACNN. The model accuracy reaches 100% at 2 dB SNR level, which is far better than other models. Experimental results show that the MFDDN

possesses higher accuracy for different datasets due to the customized structural design of fine-grained feature information being learned by the model. Figure 16 visualizes the accuracy trend as the SNR increases. The accuracy variation of MFDDN is compared to other models, which indicates that the fine-grained feature learning capability enables the network to have stronger robustness.

To further analyze the learning ability of the network for the features of original bearing vibration signals and explore the learning rate of the network, t-random adjacent embedding (T-SNE) algorithm is introduced for two-dimensional visualization. Figure 17 shows the feature distribution of MFDDN training samples when the SNR is - 4. With the increase of epochs, MFDDN can remove the noisy features and retain the effective fault features of the data, which shows the effectiveness of the frequency division denoising module.

The confusion matrix was introduced to evaluate the model and explore the bearing fault diagnosis performance

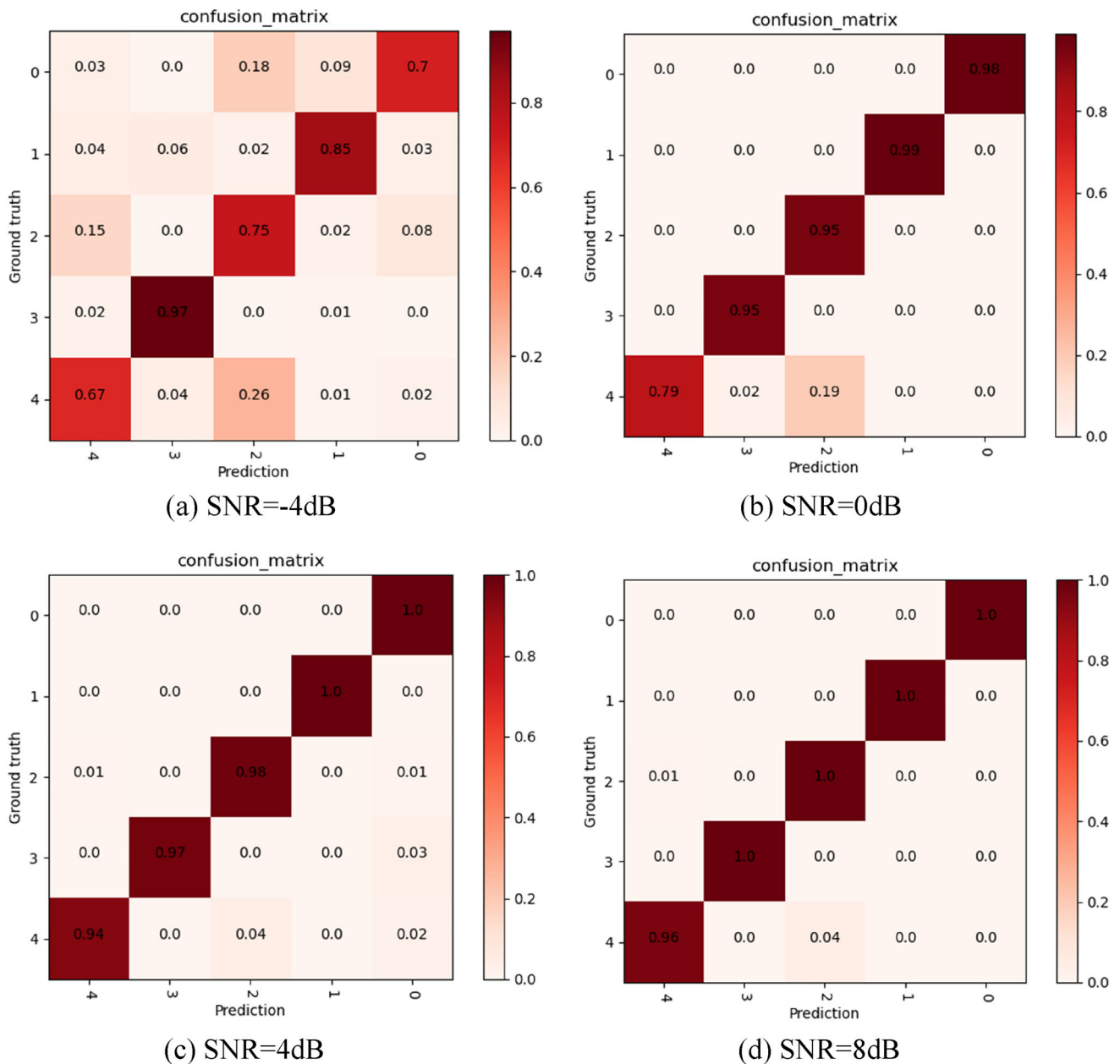


Fig. 12 Confusion matrix of fault diagnosis model

of MFDDN under four SNR conditions. As can be seen from Fig. 18, MFDDN has a low diagnostic accuracy for faults 1 and 2. With the increase of SNR, the precise control of fine-grained features enables the network to have a stronger ability to identify all types of faults.

Diagnostic experiments of noise test data under different loads

The diagnostic accuracy of MFDDN is tested under different load domains at -2 dB SNR level. The diagnostic accuracy with SVM, KNN, Random Forest, CNN, TICNN,

MSACAE-CNN, ResNet-LSTM, and IMS-FACNN models, the experimental results are shown in Fig. 19. Three data sets under different loads (1HP, 2HP, 3HP) were introduced as experimental data, which were respectively represented as data set A, data set B, and data set C.

The horizontal axis of Fig. 19 represents the load variation. For example, $A \rightarrow B$ represents data set A as the training data and data set B as the test data. The vertical axis represents the experimental accuracy. Table 6 shows the diagnostic accuracy of the bearings in the noisy environment under varying loads.

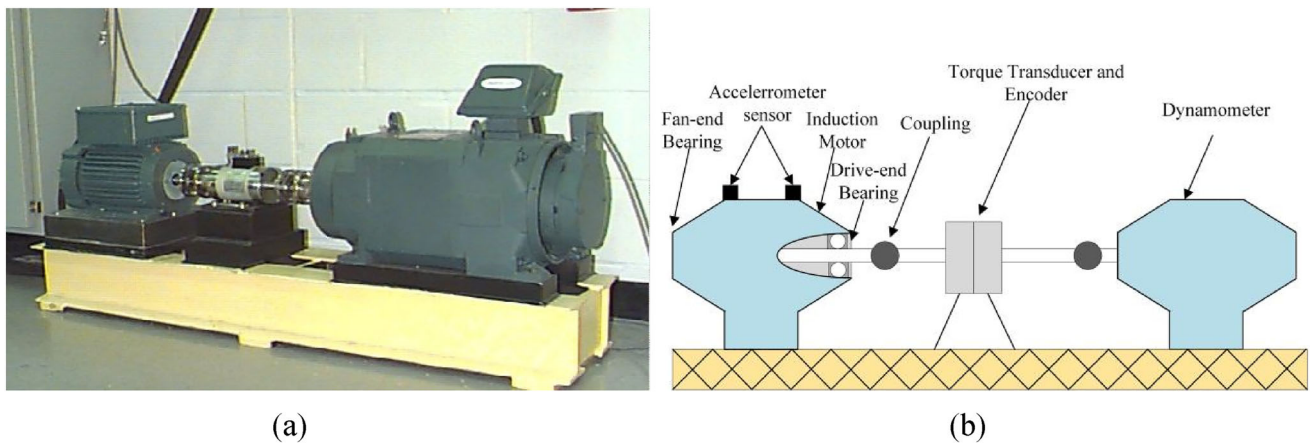


Fig. 13 **a** The bearing data acquisition device of CWRU and **b** sketch of the device

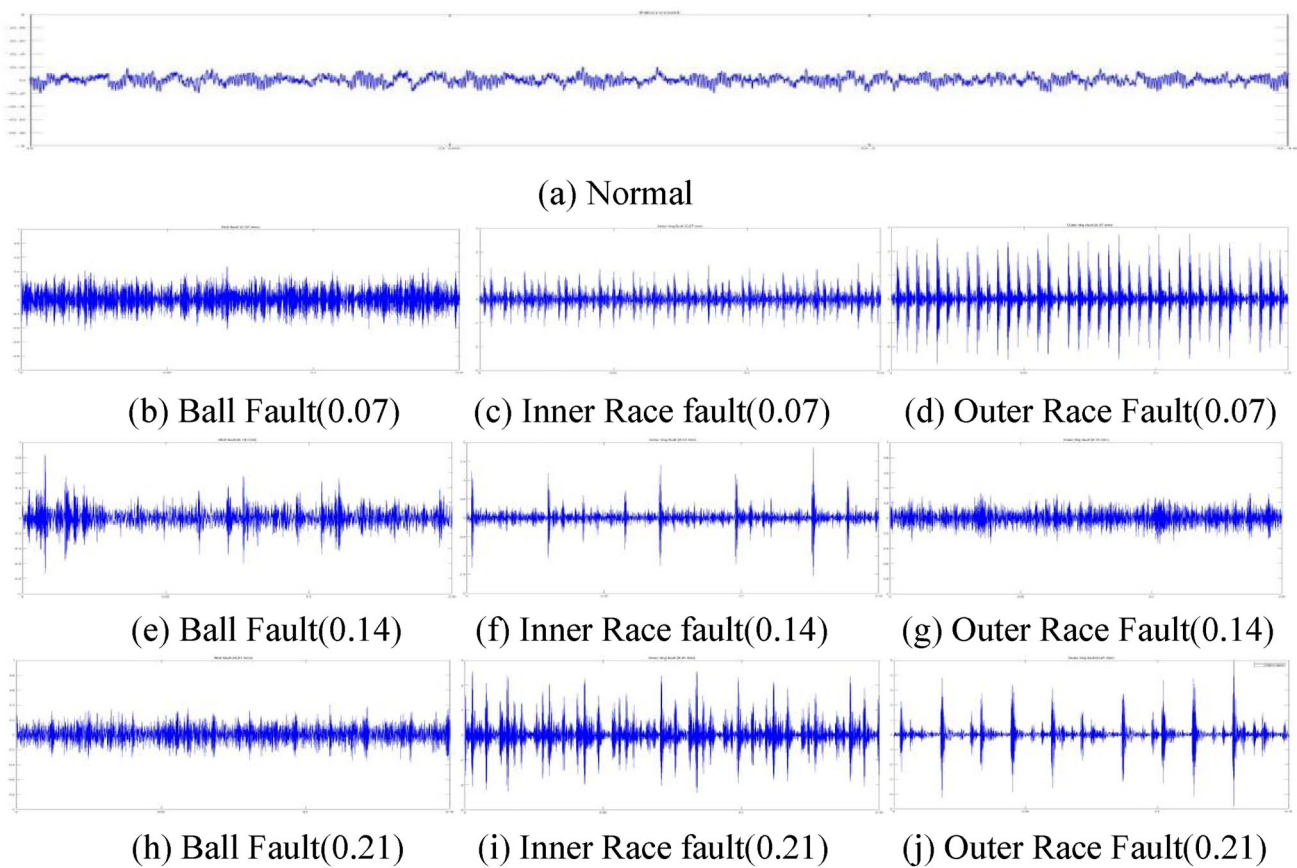


Fig. 14 Vibration signals of CWRU under 2 horsepower (hp)

The fitness of the MFDDN is verified in this section by training in one dataset and testing in the other, as shown in Table 6. It can be observed that MFDDN has a higher diagnostic accuracy of 95.1%, indicating that it can extract feature information with domain invariance from the original signal. The MFDDN model possesses stronger robustness in complex environments, which proves that the features learned by

MFDDN from the original signal are more domain invariant than other fault diagnosis model.

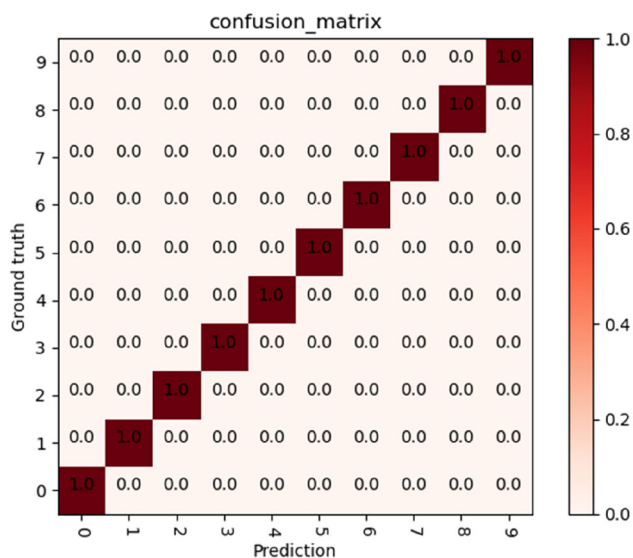


Fig. 15 Confusion matrix of CWRU dataset in noiseless condition

Evaluation model

Ablation experiment

In this section, we validate the functionality of each module to verify the effectiveness of key components under two baseline datasets, and the SNR of the signal is set to -4 . The customized learning provides the MFDDN with effective fine-grained features to denoise, where diagnostic accuracy is employed as a guide to determine the number of frequency division denoising modules. Table 7 shows the diagnostic accuracy with different numbers of frequency division denoising modules.

Table 7 indicates that the model reaches maximum accuracy with the number of modules set to 3, which means that the network has extracted specific fine-grained features at this point. As the number of custom modules increases, the effective features can be removed by the frequency division denoising module leads to degradation of model accuracy.

To further explore the influence of each key component on the accuracy of the model, we propose the following four different ablation models. A MFDDN model without multiscale structure (MFDDN-FDD-WM) was constructed to validate the effect of the multiscale architecture on model accuracy, which was embedded with three frequency division denoising module. The MFDDN-FDD-WM model does not include the multiscale architecture, which was different from MFDDN. A MFDDN model without frequency division denoising module (MFDDN-WFDD) was established to evaluate the impact of the frequency division denoising module on the model accuracy. A MFDDN model without frequency division operation (MFDDN-WFD) was employed to analyze the impact of frequency division on the

model accuracy based on the denoising sub-network, which was embedded in the denoising sub-network compared with MFDDN. A MFDDN model without sub-network (MFDDN-WN) was built to verify the effect of denoising sub-networks on model accuracy. The two datasets with $\text{SNR} = -4$ were evaluated for module functionality of the MFDDN. The accuracy comparison results are shown in Table 8.

In this study, four ablation experimental studies were carried out to verify the impact of key components in MFDDN on model accuracy. As shown in Table 8, the multiscale structure can effectively improve the ability of the model to extract multiscale features, which in turn improves the diagnostic accuracy. The comparison of the results between MFDDN and MFDDN-WFDDN demonstrates the effectiveness of the frequency division denoising module by fine-grained feature denoising. The comparison of the MFDDN-WFD shows that frequency division can provide the network with more fine-grained features to learn. The accuracy comparison with MFDDN-WN shows that the denoising sub-network improves the denoising ability of the frequency division denoising module by adaptively setting the denoising threshold. The MFDDN model achieves 78.6% and 96.5% accuracy on the two datasets by extracting fine-grained features using the sub-network to set the denoising threshold, respectively.

Evaluation indicators

We evaluate the performance of the MFDDN model by parameters, floating point operations (FLOGs) and running time. The bearing data of CWRU with a SNR of -2 were used to evaluate the complexity of the model. As shown in Table 9, the parameters, FLOGs, and running time of MFDDN are larger than most models, which leads to a higher complexity of MFDDN compared to other models. This makes the current stage of the model difficult when it comes to practical industrial applications. At the same time, the small amount of fault data in industrial sites leads to the model without sufficient samples for training. The transfer learning as an effective method can significantly improve the classification recognition accuracy for tasks with inadequate samples. Therefore, our subsequent goal is to reduce the complexity of the model based on transfer learning and maintain the accuracy for practical industrial applications.

Conclusion

A MFDDN model is presented for the intelligent fault diagnosis, which can mine multiscale features and time-series features of vibration signals under complex noise conditions. A multiscale structure was constructed to extract features of vibration signals and embeds the frequency division denoising modules for customized fine-grained feature learning for

adaptive denoising in the feature map. With the proposed frequency division denoising module, the feature map is refined to customize the removal of noise from the feature map, and a sub-network is constructed with adaptive denoising. We choose the hyperparameters of the model with accuracy-oriented, where epoch is set to 100 and batch size is set to 64. The data set is divided into training set, validation set, and test set in the ratio of 8:1:1 based on fivefold cross-validation. To verify the robustness and adaptability of the model, we validate the MFDDN model on the MaFaulDa and CWRU datasets and compare it with other classical models. The proposed MFDDN achieves an average accuracy of 100% in

a noiseless environment and 92.2% in different noisy environments, respectively. Such performance outperforms other models on the MaFaulDa dataset. For the CWRU dataset, the proposed MFDDN also achieved an average accuracy of 100% in the noiseless environment and reached an average accuracy of 99.1% in different noisy environments, which is higher than other models. In the variable load noise environment, MFDDN achieves an accuracy of 95.1%, which is higher than other models. The ablation experiments were designed to evaluate the impact of each key component on the accuracy of the model. The above results show that MFDDN achieves advanced bearing fault diagnosis accuracy, and the

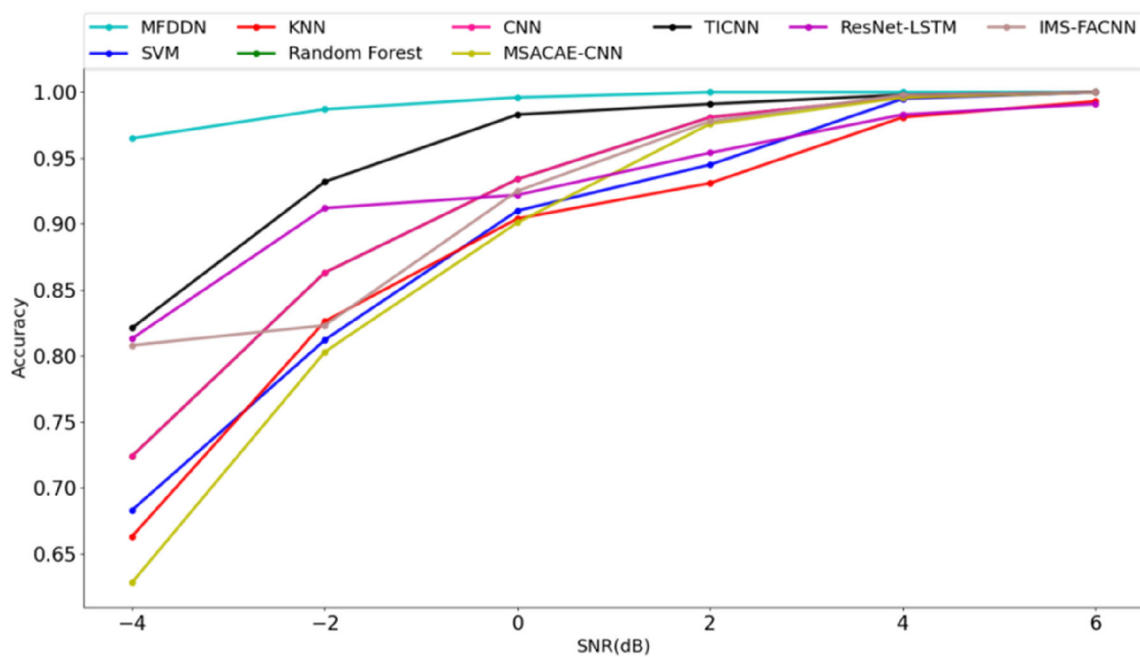


Fig. 16 Broken line graph of diagnosis results of each model under different SNR

Table 5 Comparison table of model diagnosis results under noise environment

Diagnosis model	SNR = - 4 (%)	SNR = - 2 (%)	SNR = 0 (%)	SNR = 2 (%)	SNR = 4 (%)	SNR = 6 (%)	Avg (%)
MFDDN	96.5	98.7	99.6	100	100	100	99.1
SVM	68.3	81.2	91.2	94.5	98.9	100	89.0
KNN	66.3	82.6	90.4	93.1	98.1	99.3	88.3
Random Forest	72.4	86.3	93.4	98.1	99.6	100	91.6
CNN	62.8	80.3	90.1	97.6	99.6	100	88.4
MSACAE-CNN	82.1	93.2	98.3	99.1	99.8	100	95.4
TICNN	81.3	91.2	92.2	95.4	98.3	99.1	92.9
ResNet-LSTM	80.8	82.3	92.5	97.8	99.8	100	92.2
IMS-FACNN	83.6	88.2	93.1	94.3	98.0	99.4	92.1

Bold represents the optimal performance under different conditions

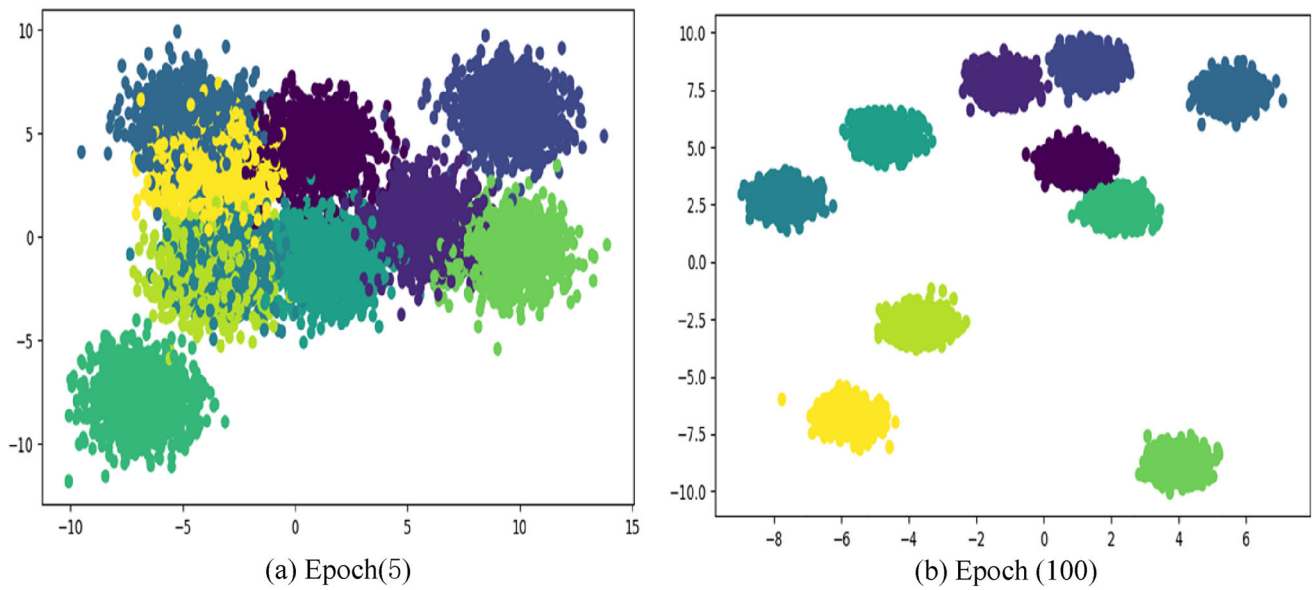


Fig. 17 Feature distribution visualization of MFDDN

customizable fine-grained feature processing method provided by the frequency division denoising module enables the model to reach high diagnostic accuracy and excellent robustness under complex conditions. However, the number of parameters, computational effort and running time are introduced to evaluate the model complexity, which is higher than that other common models. At the same time, due to the small sample of fault data in industrial sites, the accuracy

of the MFDDN under small sample is more demanding. In the subsequent study, we will focus on reducing the complexity of the model by adding 1×1 convolutional layers or separating convolutions and applying transfer learning to the MFDDN optimization.

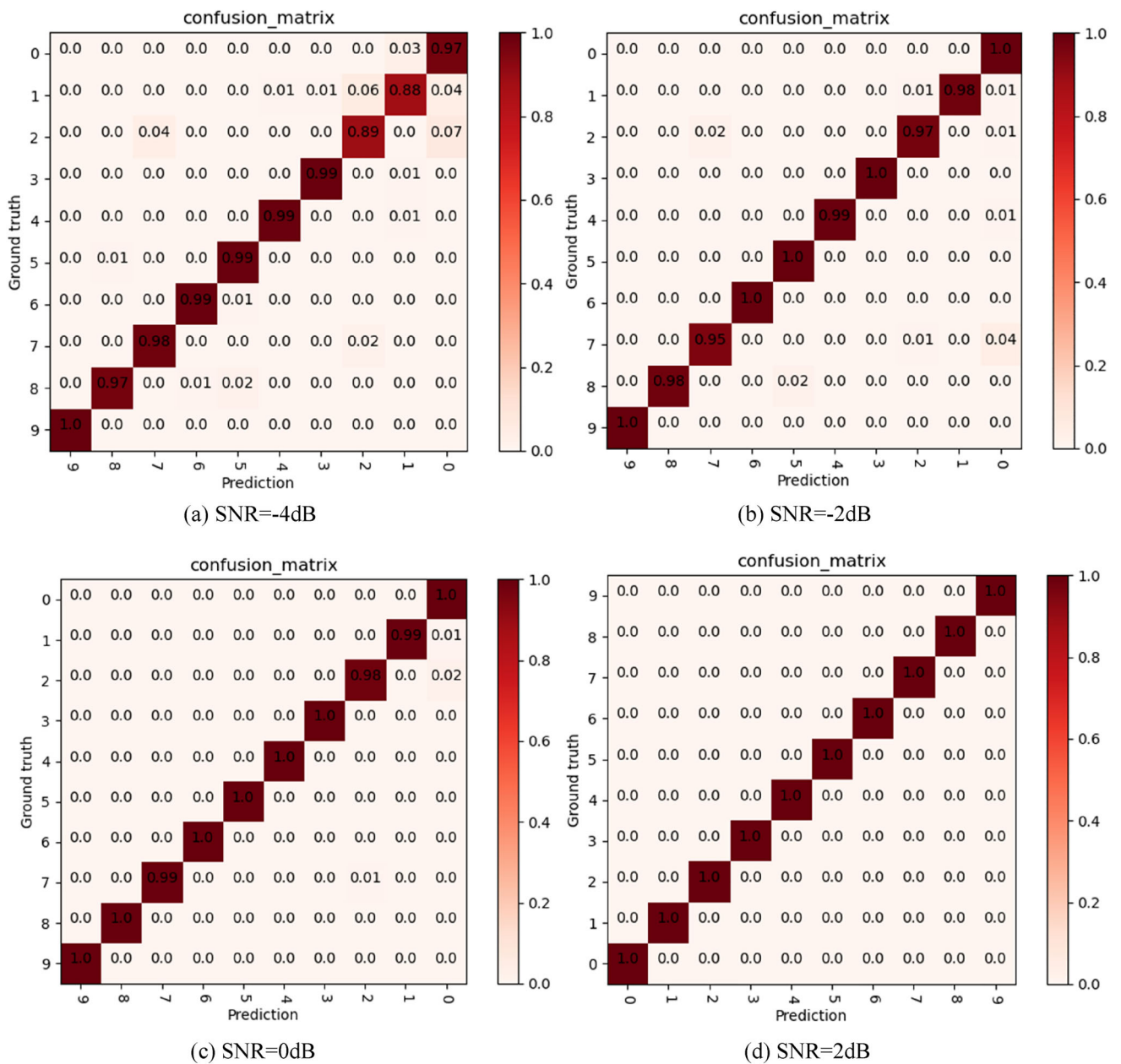


Fig. 18 Confusion matrix of fault diagnosis results under different SNR

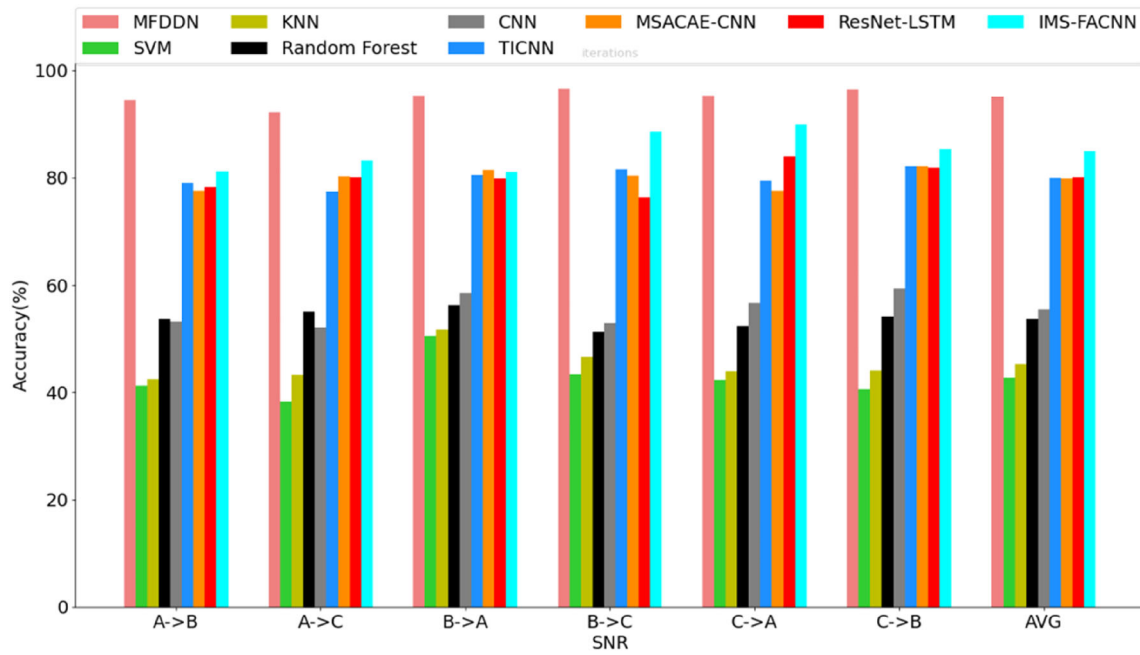


Fig. 19 Diagnosis accuracy under different loads in the noise environment

Table 6 Comparison of test result of load change under SNR of -2

Load variation	A → B (%)	A → C (%)	B → A (%)	B → C (%)	C → A (%)	C → B (%)	Average (%)
MFDDN	94.5	92.2	95.2	96.6	95.2	96.4	95.1
SVM	41.2	38.3	50.5	43.4	42.3	40.6	42.7
KNN	42.4	43.2	51.7	46.6	43.9	44.1	45.3
Random Forest	53.7	55.0	56.2	51.3	52.4	54.1	53.7
CNN	53.2	52.1	58.6	52.9	56.7	59.3	55.4
TICNN	89.1	87.4	90.5	91.6	89.4	92.1	89.9
MSACAE-CNN	77.6	80.3	81.5	80.4	77.6	82.1	79.9
ResNet-LSTM	78.3	80.1	79.9	76.4	84.0	81.8	80.1
IMS-FACNN	81.2	83.2	81.1	88.6	89.9	85.4	84.9

Bold represents the optimal performance under different conditions

Table 7 Maximum classification accuracy with different customization modules

Data set	Number of frequency division denoising module	Accuracy (%)
MaFaulDa	$n = 1$	57.2
	$n = 2$	67.1
	$n = 3$	78.6
	$n = 4$	78.4
CWRU	$n = 1$	82.5
	$n = 2$	92.8
	$n = 3$	96.5
	$n = 4$	95.3

Bold represents the optimal performance under different conditions

Table 8 Model accuracy of ablation experiments

Data set	Fault diagnosis model	Accuracy (%)
MaFaulDa	MFDDN	78.6
	MFDDN-FDD-WM	71.2
	MFDDN-WFDD	41.3
	MFDDN-WFD	52.1
	MFDDN-WN	66.7
CWRU	MFDDN	96.5
	MFDDN-FDD-WM	87.8
	MFDDN-WFDD	74.6
	MFDDN-WFD	82.8
	MFDDN-WN	86.4

Bold represents the optimal performance under different conditions

Table 9 Time complexity of MFDDN model and other models

Diagnosis model	Accuracy (%)	Parameters	FLOPs	Running time (ms)
MFDDN	98.7	268,326	59.54 M	168.98
SVM	81.2	/	/	54.32
Random Forest	86.3	/	/	78.71
KNN	82.6	/	/	67.43
TICNN	80.3	84,598	27.06 M	92.27
MSACAE-CNN	93.2	683,461	127.63 M	243.84
ResNet-LSTM	91.2	221,792	42.34 M	142.07
IMS-FACNN	82.3	117,436	32.83 M	98.50

Bold represents the optimal performance under different conditions

Acknowledgements This paper is supported by National Natural Science Foundation of China (Grant no. 51875457), the Key Research and Development Program of Shaanxi Province of China (2022SF-259), Xi'an science and technology plan project (22GXFW0128).

Data availability The datasets utilized in this work are publicly available in the following links (<https://engineering.case.edu/bearingdatacenter/>, http://www02.smt.ufrj.br/~offshore/mfs/page_01.html).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Lau ECC, Ngan HW (2010) Detection of motor bearing outer race-way defect by wavelet packet transformed motor current signature analysis. *IEEE Trans Instrum Meas* 59(10):2683–2690. <https://doi.org/10.1109/TIM.2010.2045927>
- Rai VK, Mohanty AR (2007) Bearing fault diagnosis using fft of intrinsic mode functions in hilbert huang transform. *Mech Syst Signal Process* 21(6):2607–2615. <https://doi.org/10.1016/j.ymsp.2006.12.004>
- Feng Z, Zuo MJ (2013) Fault diagnosis of planetary gearboxes via torsional vibration signal analysis. *Mech Syst Signal Process* 36(2):401–421. <https://doi.org/10.1016/j.ymsp.2012.11.004>
- Picot A, Obeid Z, Regnier J, Poignant S, Darnis O, Maussion P (2014) Statistic-based spectral indicator for bearing fault detection in permanent-magnet synchronous machines using the stator current. *Mech Syst Signal Process* 46(2):424–441. <https://doi.org/10.1016/j.ymsp.2014.01.006>
- Liu H, Jaspreet SD (2014) A time domain approach to diagnose gearbox fault based on measured vibration signals. *J Sound Vib* 333(7):2164–2180. <https://doi.org/10.1016/j.jsv.2013.11.033>
- Wang D, Sun S, Tse PW (2015) A general sequential monte carlo method based optimal wavelet filter: a bayesian approach for extracting bearing fault features. *Mech Syst Signal Process* 52–53:293–308. <https://doi.org/10.1016/j.ymsp.2014.07.005>
- Li C, Ming L, Yi Z, Hou S (2012) Multi-scale autocorrelation via morphological wavelet slices for rolling element bearing fault diagnosis. *Mech Syst Signal Process* 31:428–446. <https://doi.org/10.1016/j.ymsp.2012.03.012>

8. Wang D, Tsui KL, Zhou Q (2016) Novel gauss-hermite integration based bayesian inference on optimal wavelet parameters for bearing fault diagnosis. *Mech Syst Signal Process* 72–73:80–91. <https://doi.org/10.1016/j.ymssp.2015.11.018>
9. Zhao B, Zhang X, Hai L, Yang Z (2020) Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. *Knowl-Based Syst* 199:105971. <https://doi.org/10.1016/j.knosys.2020.105971>
10. Jegadeeshwaran R, Sugumaran V (2015) Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. *Mech Syst Signal Process* 52–53:436–446. <https://doi.org/10.1016/j.ymssp.2014.08.007>
11. Gu YK, Zhou XQ, Yu DP, Shen YJ (2018) Fault diagnosis method of rolling bearing using principal component analysis and support vector machine. *J Mech Sci Technol* 32(11):5079–5088. <https://doi.org/10.1007/s12206-018-1004-0>
12. Chow MY, Mangum PM, Yee SO (2002) A neural network approach to real-time condition monitoring of induction motors. *IEEE Trans Industr Electron* 38(6):448–453. <https://doi.org/10.1109/41.107100>
13. Ma D, Liang Y, Zhao X, Guan R, Shi X (2013) Multi-bp expert system for fault diagnosis of powersystem. *Eng Appl Artif Intell* 26(3):937–944. <https://doi.org/10.1016/j.engappai.2012.03.017>
14. Shao M, Zhu XJ, Cao HF, Shen HF (2014) An artificial neural network ensemble method for fault diagnosis of proton exchange membrane fuel cell system. *Energy* 67(1):268–275. <https://doi.org/10.1016/j.energy.2014.01.079>
15. Zhao Z, Li T, Wu J, Sun C, Chen X (2020) Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study. *ISA Trans* 107(10):224–255. <https://doi.org/10.1016/j.isatra.2020.08.010>
16. Wu Z, Jiang H, Lu T, Zhao K (2020) A deep transfer maximum classifier discrepancy method for rolling bearing fault diagnosis under few labeled data. *Knowl-Based Syst* 196:105814. <https://doi.org/10.1016/j.knosys.2020.105814>
17. Zhong T, Qu J, Fang X, Li H, Wang Z (2021) The intermittent fault diagnosis of analog circuits based on EEMD-DBN. *Neurocomputing* 436:74–91. <https://doi.org/10.1016/j.neucom.2021.01.001>
18. Kong X, Mao G, Wang Q, Ma H, Yang W (2020) A multi-ensemble method based on deep auto-encoders for fault diagnosis of rolling bearings. *Measurement* 151:107132. <https://doi.org/10.1016/j.measurement.2019.107132>
19. Wang S, Xiang J, Zhong Y, Zhou Y (2018) Convolutional neural network-based hidden Markov models for rolling element bearing fault identification. *Knowl Based Syst* 144(15):65–76. <https://doi.org/10.1016/j.knosys.2017.12.027>
20. Grezma J, Zhang J, Wang P, Gao RX (2020) Multi-stream convolutional neural network-based fault diagnosis for variable frequency drives in sustainable manufacturing systems. *Procedia Manuf* 43:511–518. <https://doi.org/10.1016/j.promfg.2020.02.181>
21. Zhu ZA, Peng GA, Chen YA, Gao H (2019) A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis. *Neurocomputing* 323:62–75. <https://doi.org/10.1016/j.neucom.2018.09.050>
22. Pandhare V, Singh J, Lee J (2019) Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features. In: 2019 Prognostics and system health management conference, pp 320–326. <https://doi.org/10.1109/PHM-Paris.2019.00061>
23. Wang H, Liu Z, Peng D, Cheng Z (2021) Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. *ISA Trans* 128:470–484. <https://doi.org/10.1016/j.isatra.2021.11.028>
24. Zhao M, Zhong S, Fu X, Tang B, Pecht M (2020) Deep residual shrinkage networks for fault diagnosis. *IEEE Trans Industr Inf* 16(7):4681–4690. <https://doi.org/10.1109/TII.2019.2943898>
25. Su K, Liu J, Xiong H (2021) Hierarchical diagnosis of bearing faults using branch convolutional neural network considering noise interference and variable working conditions. *Knowl-Based Syst* 230:107386. <https://doi.org/10.1016/j.knosys.2021.107386>
26. Zhang W, Li C, Peng G, Chen Y, Zhang Z (2017) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech Syst Signal Process* 100:439–453. <https://doi.org/10.1016/j.ymssp.2017.06.022>
27. Long Z, Xiong G, Liu H, Zou H, Guo W (2010) Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Syst Appl* 37(8):6077–6085. <https://doi.org/10.1016/j.eswa.2010.02.118>
28. Xu Y, Yan X, Sun B, Liu Z (2020) Global contextual residual convolutional neural networks for motor fault diagnosis under variable-speed conditions. *Reliab Eng Syst Saf* 225:108618. <https://doi.org/10.1016/j.res.2022.108618>
29. Yu W, Pi D, Xie L, Luo Y (2021) Multiscale attentional residual neural network framework for remaining useful life prediction of bearings. *Measurement* 177:109310. <https://doi.org/10.1016/j.measurement.2021.109310>
30. Zhang K, Chen J, Zhang T, Zhou Z (2020) A compact convolutional neural network augmented with multiscale feature extraction of acquired monitoring data for mechanical intelligent fault diagnosis. *J Manuf Syst* 55:273–284. <https://doi.org/10.1016/j.jmsy.2020.04.016>
31. Yu J, Zhou X, Lu L, Zhao Z (2021) Multiscale dynamic fusion global sparse network for gearbox fault diagnosis. *IEEE Trans Instrum Meas* 70:1–11. <https://doi.org/10.1109/TIM.2021.3076855>
32. Zhang H, Xiao Z, Wang J, Li F, Szczerbicki E (2020) A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention. *IEEE Internet Things J* 7(2):1072–1080. <https://doi.org/10.1109/JIOT.2019.2949715>
33. Xiao Z, Xu X, Zhang H, Szczerbicki E (2021) A new multi-process collaborative architecture for time series classification. *Knowl-Based Syst* 220:106934. <https://doi.org/10.1016/j.knosys.2021.106934>
34. Kong Y, Wang T, Chu F (2019) Meshing frequency modulation assisted empirical wavelet transform for fault diagnosis of wind turbine planetary ring gear. *Renew Energy* 132:1373–1388. <https://doi.org/10.1016/j.renene.2018.09.027>
35. Wang D, Zhao Y, Yi C, Tsui KL, Lin J (2018) Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings. *Mech Syst Signal Process* 101:292–308. <https://doi.org/10.1016/j.ymssp.2017.08.038>
36. Kedadouche M, Liu Z, Vu VH (2016) A new approach based on oma-empirical wavelet transforms for bearing fault diagnosis. *Measurement* 90:292–308. <https://doi.org/10.1016/j.measurement.2016.04.069>
37. Hu J, Shen L, Albanie S, Sun G, Wu E (2017) Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 42(8):2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
38. Chen Y, Fan H, Xu B, Yan Z, Kalantidis Y, Rohrbach M, Shuicheng Y, Feng J (2019) Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution. In: 2019 IEEE/CVF international conference on computer vision (ICCV) 2019, pp 3434–3443. <https://doi.org/10.1109/ICCV.2019.00353>
39. Loparo KA (2012) Case western reserve university bearing data center. *Bear. Vib. Data Sets, Case West. Reserv. Univ.* <https://engineering.case.edu/bearingdatacenter/>, pp 22–28
40. Marins MA, Ribeiro F, Netto SL, Silva E (2018) Improved similarity-based modeling for the classification of rotating-machine failures. *J Franklin Inst* 355(4):1913–1930. <https://doi.org/10.1016/j.jfranklin.2017.07.038>

41. Wang Y, Cheng L (2020) A combination of residual and long-short-term memory network for bearing fault diagnosis based on time-series model analysis. *Meas Sci Technol*. <https://doi.org/10.1088/1361-6501/abaa1e>
42. Xu Z, Li C, Yang Y (2020) Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism. *ISA Trans* 110:379–393. <https://doi.org/10.1016/j.isatra.2020.10.054>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.