J Fail. Anal. and Preven. (2023) 23:795–811
https://doi.org/10.1007/s11668-023-01616-9

TECHNICAL ARTICLE—PEER-REVIEWED

An Intelligent Fault Diagnosis Method of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network

Qi Zhang · Linfeng Deng

Submitted: 27 November 2022 / in revised form: 18 January 2023 / Accepted: 23 January 2023 / Published online: 7 February 2023
© ASM International 2023

Abstract The rolling bearing is the key component of rotating machinery, and fault diagnosis for rolling bearings can ensure the safe operation of rotating machinery. Fault diagnosis technology based on deep learning has been largely studied for bearing fault diagnosis. However, for the deep learning model based on convolutional neural network, there are some intrinsic problems of producing inconspicuous features and useful feature information loss in the process of feature extraction of the raw fault vibration signals. In this work, an intelligent fault diagnosis method of rolling bearings based on short-time Fourier transform and convolutional neural network (STFT-CNN) is proposed. The one-dimensional vibration signals are converted into time–frequency images by STFT. Then, time–frequency images are inputted into STFT-CNN model for fault feature learning and fault identification. For the STFT of the vibration signals, the window type, window width and translation overlap width of the five typical window functions are studied and optimal one is obtained. And in the STFT-CNN model, the stacked double convolutional layers are adopted to improve the nonlinear expression capability of the model. To verify the effectiveness of the proposed method, experiments are carried out on the Case Western Reserve University (CWRU) and the Machine Failure Prevention Technology (MFPT) Society bearing datasets. The results show that the proposed method outperforms other comparative methods and reaches the identification accuracy of 100% and 99.96% for CWRU and MFPT, respectively.

Q. Zhang · L. Deng (✉)
School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou, China
e-mail: denglinfeng2002@163.com

Keywords Rotating machinery · Rolling bearing · Intelligent fault diagnosis · Convolutional neural network · Short-time Fourier transform

Introduction

Rolling bearings are widely used in industrial production equipment. As the most common components of rotating machinery, the normal state of rolling bearings directly affects the efficient and safe operation of mechanical equipment [1], so it is important to carry out the fault diagnosis of rolling bearings in advance. In recent years, many fault diagnosis methods have been applied to rolling bearings, and fault diagnosis technology has developed rapidly with numerous achievements [2]. The fault diagnosis technology can be divided into four categories: model-based [3–5], signal-based [6], knowledge-based (also named as data-driven) [7], and hybrid [8]. Most of the current research that focuses in data-driven fault diagnosis are based on deep learning methods, with convolutional neural network (CNN) being the most widely used. In the last decade, CNN has been widely used in the field of machinery fault diagnosis and has achieved good results [9]. Generally, the use of CNN for fault diagnosis can be divided into four processes: data collection, model building, feature learning, and decision making [10].

Due to the high effectiveness in the vibration signal processing, one-dimensional (1D) convolutional neural networks were successfully applied to bearing fault detection and diagnosis [11, 12]. Zhang et al. [13] proposed a 1D CNN bearing fault diagnosis model acting on time domain signals. Abdeljaber et al. [14] fed raw time domain signals into a 1D CNN and applied it to real-time structural

damage detection in bleachers. Su et al. [15] proposed ResNet to directly process the raw time domain signal for fault diagnosis of a high-speed train bogie. Wang et al. [16] proposed a multi-attention one-dimensional convolutional neural network (MA1DCNN) to diagnose wheelset bearing faults. Zhao et al. [17] transformed the 1D time domain signals into frequency domain images by fast Fourier transform (FFT), which were fed into BiLSTM, LeNet, AlexNet, ResNet18 and other models for fault diagnosis. Janssens et al. [18] used the discrete Fourier transform (DFT) to change the time domain signal into a frequency domain signal, which was fed into a CNN for fault diagnosis. Although 1D CNN has been applied in the field of fault diagnosis, the following shortcomings still exist in the 1D CNN model.

(1) The advantages of CNN cannot be fully utilized when 1D signals are used as the input, and after all, CNN was originally designed to solve the learning problems of the two-dimensional (2D) images.
(2) When 1D CNN is used for processing the time domain signal directly, the useful fault feature information is lost. The 1D CNN model cannot obtain the accurate fault characteristics.

Image classification techniques using deep learning for fault diagnosis are more efficient and better applicable. Two-dimensional images often contain a wealth of fault information, and deep learning can autonomously extract features from the images that can characterize the type of faults at a deep level in bearings. The process is to convert 1D vibration signal into 2D form, followed by image classification [19] to achieve identification of bearing fault states. Bhadane et al. [20] extracted statistical features from vibration data as input to the model and further developed a 2D CNN for bearing fault classification. Hoang et al. [21] converted the original time domain signals into 2D gray-scale images according to the time series as an input to CNN for fault diagnosis. Wang et al. [22] segmented the 1D raw signals and converted them into frequency domain signals using FFT and then, converted the frequency domain signals into 2D images. Finally, the 2D images were fed into an improved LeNet-5 model, which was successfully used to rapidly evaluate the bearing reliability and predict the remaining bearing life. Wen et al. [23] proposed to convert the original time domain signals into 2D gray-scale images and input the improved LeNet-5 model for fault diagnosis.

Unlike the 2D transform above, the 1D signals can also be transformed by the short-time Fourier transform (STFT) to generate 2D time–frequency images. The time–frequency images have both time domain and frequency domain information and contain more fault information. Compared with time series signals, time–frequency images are easier to extract information in noisy environments. In the study of the fault diagnosis, it has been demonstrated that time–frequency domain inputs are better than time domain inputs [24]. STFT is widely used in the signal processing of rotating machinery fault diagnosis, and many researchers have made a lot of contributions to rolling bearings fault diagnosis based on STFT. David et al. [25] used three time–frequency analysis methods to convert time domain signals into the corresponding time–frequency images. Then, the time–frequency images were fed into the proposed CNN model for fault diagnosis, and better results were achieved with fewer learnable parameters. Zhu et al. [26] used the STFT to generate 2D images from 1D signals, followed by fault diagnosis using a capsule network. Tao et al. [27] generated time–frequency images from the original 1D vibration signals by STFT, which were input to CatGAN for fault identification.

From the above literatures, it can be found that the window width of the STFT and its effect on the diagnosis results were investigated. But the type of window function or the width of the translation overlap was hardly considered, or even both. Therefore, we analyze in depth the respective effects of the window function type, window width and translation overlap width on the fault diagnosis model. And furthermore, we construct a new network for rolling bearing fault diagnosis based on short-time Fourier transform and convolutional neural network (STFT-CNN).

The main contributions of this work are summarized as follows:

(1) Convert the 1D vibration signals into time–frequency images as the input of the STFT-CNN which can express more comprehensive fault information and illustrate more obvious fault characteristics.
(2) The effects of five different window functions in the STFT on the diagnostic performance of the proposed model are investigated, and the types of window functions, window widths and translation overlap widths suitable for fault diagnosis are determined.
(3) A new two-layer stacked CNN is proposed to improve the nonlinear expression capability of the model.

The remainder of this article is organized as follows. Section "Theoretical fundamentals" briefly introduces the CNN structure and time–frequency analysis methods. Section "The proposed method" gives a detailed description of the proposed machinery fault diagnosis method, STFT to generate time–frequency images and the STFT-CNN model structure. Section "Experimental validations" uses two cases to verify the effectiveness of the proposed method and shows a comparison analysis with recent methods for rolling bearing fault diagnosis. Finally, Sect. "Conclusion" summarizes the paper.

Theoretical Fundamentals

Convolutional Neural Network (CNN)

Traditional CNN is widely used in computer vision and is very good at extracting feature information from images. It contains many well-known networks, such as LeNet [28], AlexNet [29], VGGNet [30], ResNet [31], and MobileNet [32]. Almost all major breakthroughs in image recognition in recent years have used convolutional neural networks (CNNs). A general CNN structure is shown in Fig. 1.

A CNN is mainly composed of three parts: convolutional layer, pooling layer and fully connected layer, with the convolutional layer and the previous layer connected in a locally connected and weight sharing manner.

The convolution operation is defined as follows:

$$H_j^{l+1} = \sum_{i \in x_j} \left(H_j^l * w_{ij}^{l+1} + b_j^{l+1} \right) \quad \text{(Eq 1)}$$

where H_j^{l+1} denotes the jth feature map of the neuron at layer $l+1$, $*$ denotes the convolution operation, w_{ij}^{l+1} denotes the convolution kernel connecting the jth feature map of the neuron at layer $l+1$ and the ith feature map of the neuron at layer l, b_j^{l+1} denotes the bias, and x_j denotes the image of the input CNN.

The convolution layer is a linear operation. To enhance the classification ability of the model, a nonlinear activation function is added. The commonly used activation functions are the Sigmoid function, Tanh function and ReLU function, which are defined as follows:

$$f_{\text{sigmoid}}(x) = \frac{1}{1+e^{-x}} \quad \text{(Eq 2)}$$

$$f_{\text{Tanh}}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{(Eq 3)}$$

$$f_{\text{ReLU}}(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} = \max(0, x) \quad \text{(Eq 4)}$$

The three activation function curves are shown in Fig. 2.

Pooling layers are used to reduce network parameters, and it is defined as follows:

$$H_j^{l+1} = f\left(\beta_j^{l+1} \text{down}\left(H_j^l \right) + \text{b}_j^{l+1} \right) \quad \text{(Eq 5)}$$

where $down(\cdot)$ denotes a subsampling function, β denotes the multiplicative bias.

The common pooling methods are maximum pooling and average pooling. Maximum pooling outputs the maximum value of the window, and average pooling averages the values of the window and outputs them. A diagram of the two pooling methods is shown in Fig. 3. The size of the convolution kernel is 2 × 2, and the step size is 2 in the example.

The fully connected layer is used to classify the feature data extracted earlier, and this operation is expressed as follows:

$$y^k = f\left(w^k x^{k-1} + b^k \right) \quad \text{(Eq 6)}$$

where k is the k-th layer network, x^{k-1} is the input of the (k-1)-th fully connected layer, the y^k is the output of the k-th fully connected layer, w^k is the weight coefficient, b^k is the bias, and f is the classification function.

Softmax is used as the activation function of the fully connected layer to map the output of multiple neurons between (0, 1). It is used for multi-classification tasks, and the expression is shown below:

$$q(x_i) = \frac{e^{x_i}}{\sum_{c=1}^{C} e^{x_c}} \quad \text{(Eq 7)}$$

where x is the logical input value of the Softmax layer, and $q(x_i)$ is the C-dimensional probability vector corresponding to x. The Softmax function converts the output value of the last layer node into a probability value with a sum of 1 by

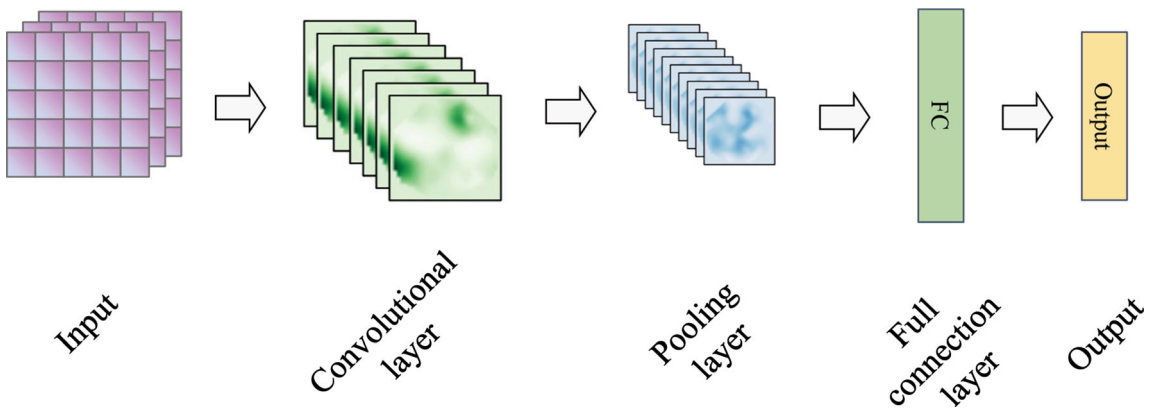

Fig. 1 General CNN structure

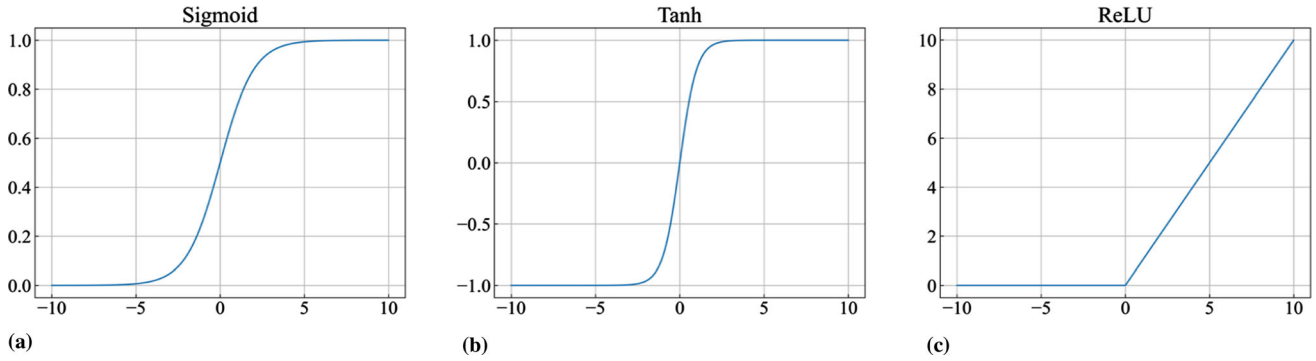

Fig. 2 Three activation function curve graphs. (**a**) Sigmoid, (**b**) Tanh, (**c**) ReLU

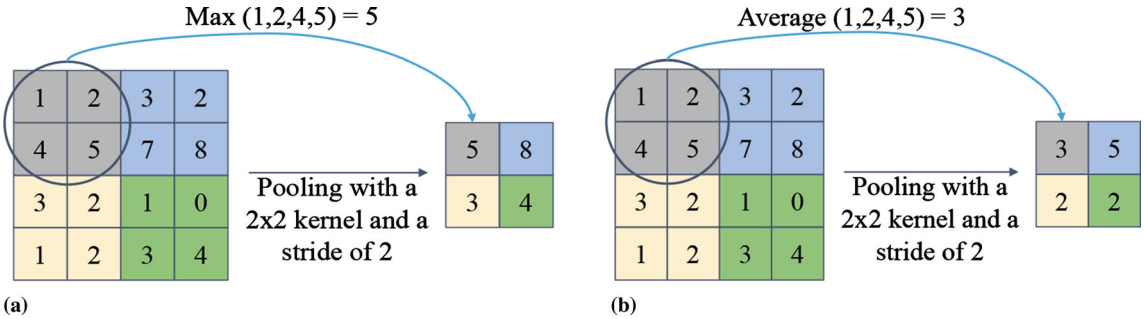

Fig. 3 Pooling methods. (**a**) Maximum pooling (**b**) Average pooling

the formula permutation. The label corresponding to the maximum probability value is the type to which the sample belongs.

For the most of classification problems, the cross-entropy loss function is often used to represent the difference between the true probability and the predicted probability distribution. The smaller the value of cross-entropy, the better the model predicts the classification effect. In addition, the cross-entropy loss function is often used with the Softmax function, and it is defined as follows:

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)) \quad \text{(Eq 8)}$$

where $p(x_i)$ denotes the true distribution of the input sample x, and $q(x_i)$ denotes the predicted distribution of the input sample x.

Time–Frequency Analysis

Time–frequency analysis technology is widely used in speech processing, signal detection, state detection and equipment fault diagnosis. The research on time–frequency analysis technology began in the 1940s. Common time–frequency analysis methods include continuous wavelet transform (CWT), S-Transform, STFT, etc. Time–frequency analysis maps 1D time domain signals to 2D time–frequency planes and reflects the joint time–frequency characteristics of the signal [33].

Suppose the wavelet mother function is $\psi(t)$, and the scale factor and translation factor are a and b, respectively. The wavelet mother function is scaled and translated to obtain the subfunction, whose equation is shown as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) a,b \in R, a > 0 \quad \text{(Eq 9)}$$

Since the scale factor and translation factor are continuously transformed, it is called the continuous wavelet function basis. The expansion of the continuous signal $s(t)$ in the wavelet basis is called the continuous wavelet transform. The CWT is defined as follows:

$$WT(a,b) = \frac{1}{\sqrt{a}}\int_R s(t)\psi^*\left(\frac{t-b}{a}\right)dt \quad \text{(Eq 10)}$$

The S-Transform is defined as follows:

$$S(\tau,f) = \int_{-\infty}^{+\infty} s(t)\frac{|f|}{\sqrt{2\pi}}e^{-\frac{(t-\tau)^2 f^2}{2}}e^{-j2\pi ft}dt \quad \text{(Eq 11)}$$

where t is the time, f is the frequency, j is an imaginary unit, and τ is the center of the Gaussian window function.

In 1946, Dennis Gabor proposed the STFT. The basic idea is to assume that the signal is smooth in a very short

time. A window function is used to divide the signal into small segments, and STFT is performed on each segment of the signal. Then, connect all the spectral analysis to form a time–frequency image, and the operation process is expressed as follows:

$$STFT(t,\omega) = \int_{-\infty}^{\infty} s(\tau)g(\tau - t)e^{-j\omega t}d\tau \quad \text{(Eq 12)}$$

Its spectrogram can be computed as follows:

$$|STFT(t,\omega)|^2 = \left|\int_{-\infty}^{\infty} s(\tau)g(\tau - t)e^{-j\omega t}d\tau\right|^2 = G(t,\omega) \quad \text{(Eq 13)}$$

where $s(t)$ denotes the signal, $g(t)$ denotes the window function, t and τ denote the moment, $g(\tau - t)$ denotes the window function whose center is located at moment t, and ω denotes the frequency.

The Proposed Method

Procedures of Proposed Method

Based on the theoretical fundamentals given above, we design a bearing fault diagnosis method. Figure 4 shows the flowchart of the proposed fault diagnosis framework based on STFT and CNN. It can be clearly seen that 1D vibration signals are sampled and subsequently conducted by the optimal STFT to form time–frequency images, and then, the images are inputted into the 2D CNN for fault classification and identification. The details of the fault diagnosis procedure are described as follows.

(1) Data preprocessing stage: Sensors collect the bearing vibration signals, and the vibration signals are divided into sample sequences in order. After that, the sample sequences are transformed into time-frequency images via the optimal STFT. To speed up the data processing, the time-frequency image data are normalized by Z-Score, and the formula is shown as Eq. 14:

$$X_i = \frac{x_i - \mu}{\sigma} \quad \text{(Eq 14)}$$

where x_i represents each data in the time–frequency image, X_i represents each data in the time–frequency image after normalization, μ and σ represent the mean and standard deviation of the data in the time–frequency image.

(2) Model training stage: The training set samples are inputted into the designed 2D-CNN model. The trained model can be obtained by continuously updating the weights iteratively to minimize and stabilize the loss function.

(3) Fault diagnosis stage: The testing set samples are inputted to the trained model to obtain fault diagnosis results.

Data Split

In the data preprocessing stage, we need to split the acquired signal. The most common method of data split in intelligent fault diagnosis field is random sampling, as shown in Fig. 5. We use this method in the experiments. To prevent information leakage, we did not perform overlapped samplings. In the generated samples, 80% of the samples are randomly selected as the training set and 20% samples as the testing set. After that, the generated samples are inputted into the STFT-CNN model.

Signal to Image Conversion

STFT contains both time domain and frequency domain information after transforming 1D vibration signals into 2D time–frequency images. STFT has two important parameters, window width and translation overlap width. The wider the width of the window function provides higher frequency domain resolution, and the narrower the width provides higher the time domain resolution. According to the Heisenberg inaccuracy principle, it is known that both cannot be obtained. So only the appropriate window width can be chosen to achieve the optimal result.

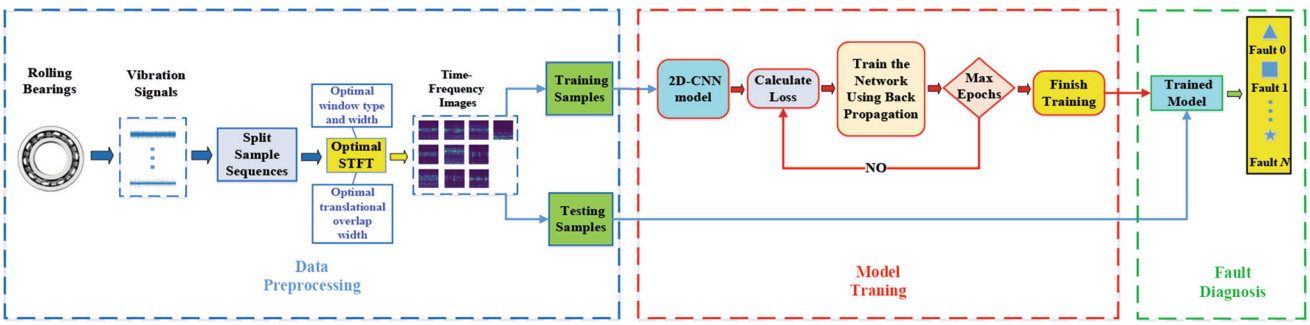

Fig. 4 Flowchart of the proposed bearing fault diagnosis method

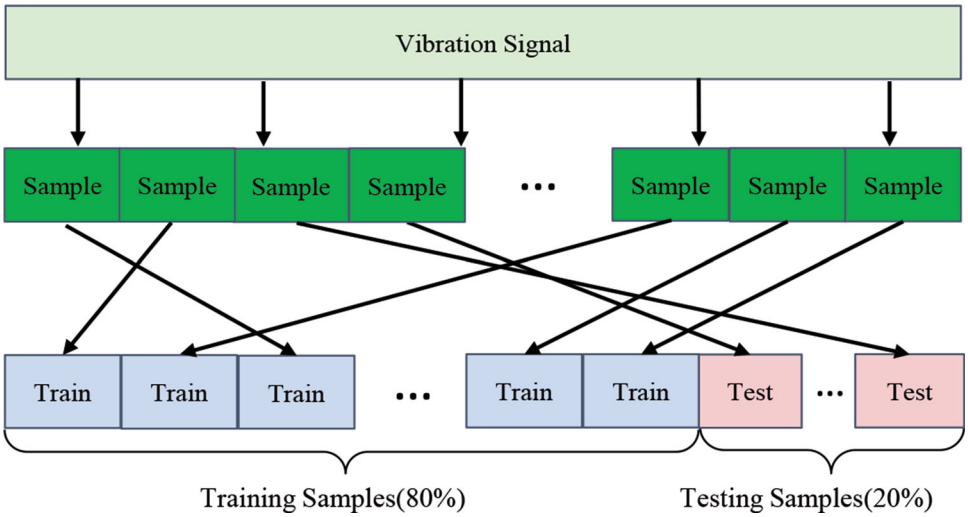

Fig. 5 Diagram of data split

Table 1 STFT window parameters setting

N_w	N_o	Size
64	34	33 × 33
128	114	65 × 65
256	250	129 × 129
512	510	257 × 257

The one-dimensional signal is passed through STFT to form a time–frequency image, and the time domain resolution is calculated as:

$$T = \left[\frac{N_x - N_o}{N_w - N_o}\right] \quad \text{(Eq 15)}$$

where $[\cdot]$ denotes rounding down, N_x is the sample length, N_o is the window function translation overlap width, and N_w represents the window function width.

The formula for the frequency domain resolution is divided into two cases where N_w is even and odd:

(1) When N_w is an even number, the frequency domain resolution equation is:

$$F = \frac{N_w}{2} + 1 \quad \text{(Eq 16)}$$

(2) When N_w is an odd number, the frequency domain resolution equation is:

$$F = \frac{N_w + 1}{2} \quad \text{(Eq 17)}$$

Reasonable time domain and frequency domain resolution can make the fault signal more obvious and reduce noise interference. Since the CNN input is preferably square matrices, four window function widths N_w and translation overlap widths N_o are set in this experiment, as shown in Table 1. The time–frequency images generated from four different window widths are input to the proposed STFT-CNN model for training and testing. Determine the window width and translation overlap width of the window function according to the accuracy index and time index.

Details of the STFT-CNN Model Structure

Figure 6 shows the proposed STFT-CNN, which consists of five convolutional layers (C), two maximum pooling layers (MP), one adaptive maximum pooling layer (AMP) and three fully connected layers (FC). FM denotes the feature map; OP denotes the output bearing state result. The original signals are converted into images, which are fed into the proposed STFT-CNN model to classify the images. In this study, the proposed STFT-CNN model is used to solve the fault diagnosis task.

The detailed structural parameters of each layer of the STFT-CNN model are shown in Table 2. The model consists of four parts. The first part consists of 32 convolutional kernels of size 5 × 5 followed by a 2 × 2 maximum pooling layers. The second part consists of a two-layer stack of 32 convolutional kernels of size 3 × 3 followed by a 2 × 2 maximum pooling layer. The third part consists of a two-layer stack of 64 convolutional kernels of size 3 × 3 followed by a 2 × 2 adaptive maximum pooling layer. The first convolutional layer is followed by maximum pooling, the remaining two convolutional layers are stacked, and then, maximum pooling is used. The fourth part is a three-layer full connection layer with input dimensions of 256, 1024, and 128, respectively. The Sigmoid activation function is initially selected for the convolutional layer, and the ReLU activation function is used for the full connection layer of the benchmark model.

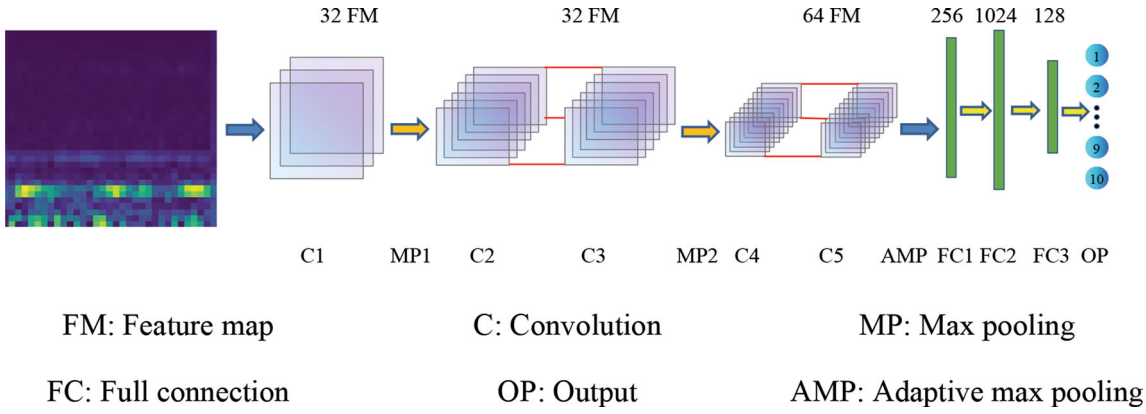

Fig. 6 Proposed STFT-CNN structure

Table 2 Structural parameters of each layer of the STFT-CNN model

Layer	Parameters
C1	Conv2d(5 × 5 × 32)
MP1	MaxPool2d(2 × 2)
C2	Conv2d(3 × 3 × 32)
C3	Conv2d(3 × 3 × 32)
MP2	MaxPool2d(2 × 2)
C4	Conv2d(3 × 3 × 64)
C5	Conv2d(3 × 3 × 64)
AMP	AdaptiveMaxPool(2 × 2)
FC1	Input dimensions = 256
FC2	Input dimensions = 1024
FC3	Input dimensions = 128

Moreover, the advantages of the STFT-CNN model are summarized as follows:

(1) The first layer of convolution adopts a 5 × 5 convolution kernel, which can extract the information from the larger neighborhood range of the time–frequency image and obtain better features. Large convolutional kernels are used in the first layer to increase the receptive field, acquire more data, and provide more information for the subsequent layers of the network, and the large convolutional kernels can better suppress high-frequency noise [34].

(2) Use two 3 × 3 convolution kernels instead of one 5 × 5 convolution kernel. Two 3 × 3 convolution layers need two activation functions, which can increase the nonlinear expression capability. Meanwhile, two stacked convolutional layers can improve the feature extraction ability, while fewer parameters can reduce the computational effort [25].

In addition to the settings of the network structural parameters, the STFT-CNN benchmark parameters are set as shown in Table 3. The loss function uses the cross-entropy loss function. The Adam optimizer is used to optimize the model parameters, the initial learning rate is 0.001, and the batch size is set to 16.

Table 3 Benchmark model parameter setting

Parameter	Value
Batch size	16
Learning rate	0.001
Optimizer	Adam
Normalization method	Z-Score
Loss function	CrossEntropy Loss

Experimental Validations

To verify the validity of the proposed model, experimental validation is conducted on two rolling bearing datasets using the deep learning framework Pytorch, and the network model is built in the Pytorch environment using Python 3.8. The computer configuration used for the experiments is CPU i7-11800H, RAM 16 GB, GPU GeForce RTX 3050Ti 4 GB.

Case Study 1: Case Western Reserve University (CWRU) Bearing Dataset

The proposed method was first validated on the CWRU bearing dataset, USA [35]. The data acquisition platform is shown in Fig. 7. A 2-hp three-phase asynchronous drive motor is used as the power source on the left side, and a torque transducer is installed in the middle to measure the speed and torque. A dynamometer is installed on the right side to generate the rated load. An acceleration sensor is installed at 12 o'clock position on the drive side to collect the vibration signal with a sampling frequency of 12,000 Hz.

The bearing type used in the experiment is the SKF 6205-2RS JEM deep groove ball bearing. The bearings were seeded with single point faults using electro-

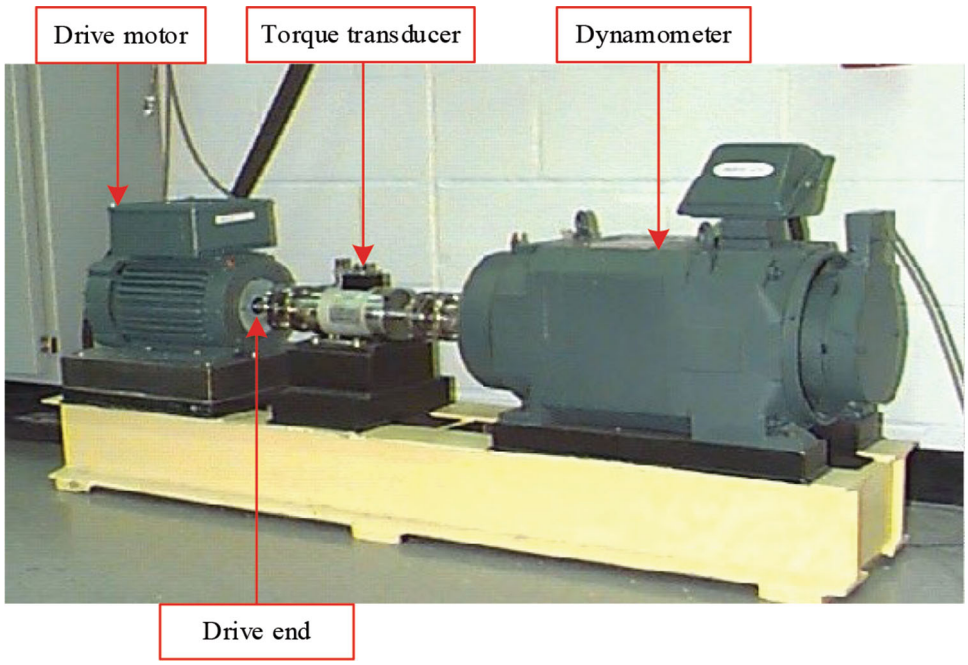

Fig. 7 Bearing data acquisition platform of CWRU

Table 4 The details of the used datasets in CWRU

Fault Location	Fault Size (mm)	Load (HP)	Label
NO	/	0,1,2,3	0
IR	0.18	0,1,2,3	1
BF	0.18	0,1,2,3	2
OF	0.18	0,1,2,3	3
IR	0.36	0,1,2,3	4
BF	0.36	0,1,2,3	5
OF	0.36	0,1,2,3	6
IR	0.54	0,1,2,3	7
BF	0.54	0,1,2,3	8
OF	0.54	0,1,2,3	9

discharge machining (EDM). And bearing faults are divided into three categories, namely inner race fault (IR), outer race fault (OR) and rolling element fault (BF). There are also three different degrees of damage for each fault type, with damage sizes of 0.18 mm, 0.36 mm, and 0.54 mm. The CWRU bearing dataset contains four load states of the bearing: 0HP, 1HP, 2HP, and 3HP. This experiment uses the drive end data at 0, 1, 2, and 3HP, and four loads to verify the performance of the proposed method. Nine fault types plus one normal state make a total of 10 health states under each load. The data of the same fault location, the same damage level, and different loads are treated as one health state. So the four loads can also be divided into 10 health states, and the detailed fault representation is shown in Table 4. To contain sufficient information for each sample, 1024 data points are used as a sample length. A total of 5940 samples can be generated for 10 health states. There are 4752 samples in the training set and 1188 samples in the test set.

We use the proposed bearing fault diagnosis method to conduct the CWRU bearing dataset, and first, the time–frequency images can be obtained. Figure 8 shows a normal state signal and the corresponding time–frequency image. The left side is the time domain waveform of the normal signal, and the right side is the time–frequency image converted by STFT from the time domain signal. In the time–frequency image, the horizontal coordinate is time, and the vertical coordinate is frequency. The vibration components of normal state are mostly in the low- and medium-frequency band.

For STFT, different window functions will undoubtedly lead to different time–frequency spectra. So, it is much necessary to explore the effects of different window functions in STFT. Thus, five common types of window functions, namely Hamming window, Blackman window, Bartlett window, Hann window, and Rectangle window, are chosen to further investigate the effects of different window function types, window width, and translation overlap width on diagnostic results. The shapes of the five window functions in the time and frequency domains are shown in Fig. 9.

It can be observed from Fig. 9 that five different types of window functions have obviously different frequency domain characteristics although the Hamming window, Blackman window and Hann window have the same time domain shapes. Moreover, the distinction and intrinsic

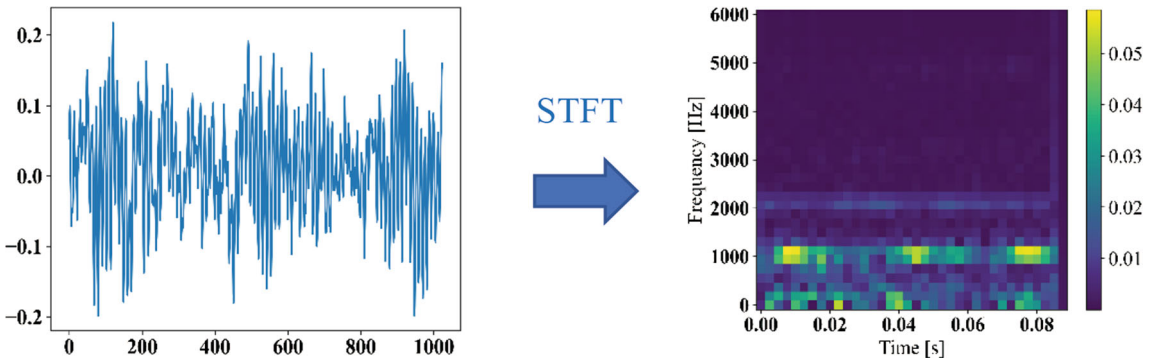

Fig. 8 1D Signal to 2D image

features of five different types of window functions can be summarized as follows:

(1) The advantage of rectangle window is that the mainlobe is more concentrated. The disadvantage is that rectangle window has a large sidelobe and a negative side lobe. It is easy to introduce high-frequency interference and leakage during the transformation process, and even negative spectrum phenomena occur. The frequency resolution is the highest, while the sidelobe leakage is the most serious, so the rectangle window is not an ideal window.

(2) Compared with the rectangle window, the Hann window has a wider mainlobe and a smaller sidelobe. From the viewpoint of reducing leakage, the Hann window is better than the rectangle window. However, the widening of the mainlobe of the Hann window corresponds to a widening of the analysis bandwidth and a decrease in frequency resolution. In this situation, the Hann window is not suitable for accurate measurements of small signals.

(3) The Hamming window and the Hann window are both cosine windows, but with different weighting factors. The weighted coefficients of the Hamming window make the sidelobe smaller, so that they decay more slowly than the Hann window and are more effective in reducing the amplitude of the sidelobe. Therefore, the Hamming window is a greatly useful window function.

(4) The Blackman window has a wide mainlobe and a small sidelobe. The amplitude resolution is the highest, but the frequency resolution is the lowest.

(5) The Bartlett window has a triangular shape in the time domain and is often used for sharpening a signal, without forming too much ripple in the frequency domain.

To obtain the optimal STFT, a comparative analysis was implemented on the five different types of window functions. In our STFT-CNN model, different window functions were utilized, and the window width and translation overlap width of each window function type are set to each value listed in Table 1. After that, 10 health state signals are used for experiment validation. The experimental results of the five window function types are shown in Fig. 10.

As can be found from Fig. 10, the bearing fault identification accuracies obtained via the proposed model with five different window functions, window widths, and translation overlap widths can reach more than 91% for 10 health states. When the window function is Hamming window, the window width is 64 and the translation overlap width is 34, the identification accuracy is the best and up to 99.94%.

Figure 11 shows the time–frequency images of 10 health states. These images are generated by using STFT with a Hamming window of width 64. The size of the generated image is 33×33, and the converted image contains 1089 pixels. From Fig. 11, it can be easily seen that any two images are obviously different, which illustrates that STFT can obtain the distinctive features of different health states.

The activation function types also affect the performance of the model. To further improve the diagnostic accuracy of the model, the activation function of the model is optimized. During a series of experiments, it is verified that the highest identification accuracy of the model was 100% when the convolutional layer used the ReLU activation function and the fully connected layer used the Tanh activation function with a training batch size of 32.

The diagnostic efficiency is also considered for the proposed model, and the time consumption of the five window functions during the verification process is recorded. The results show that the five window functions take equal time for the same window width. Thus, only the time consumption by Hamming window for the four window widths is given, as shown in Fig. 12.

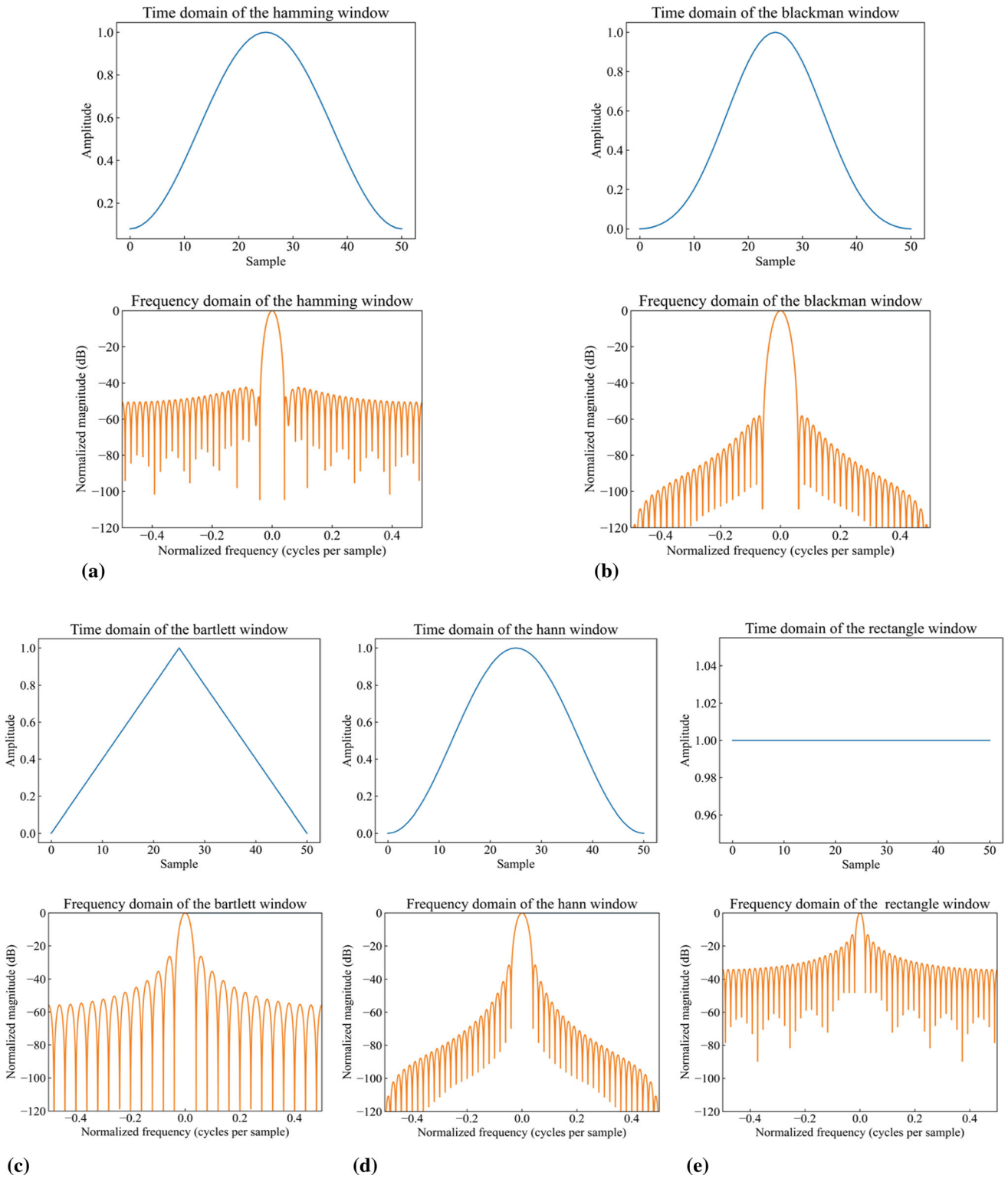

Fig. 9 The shapes of five window functions in the time and frequency domains. (**a**) Hamming window, (**b**) Blackman window, (**c**) Bartlett window, (**d**) Hann window, (**e**) Rectangle window

It can be easily seen from Fig. 12 that as the window width increases, the test time gradually increases in the first epoch. When the window width is 64, the least time is required, and the highest diagnostic efficiency can be achieved. Thus, the model has the highest diagnostic accuracy and diagnostic efficiency when the Hamming

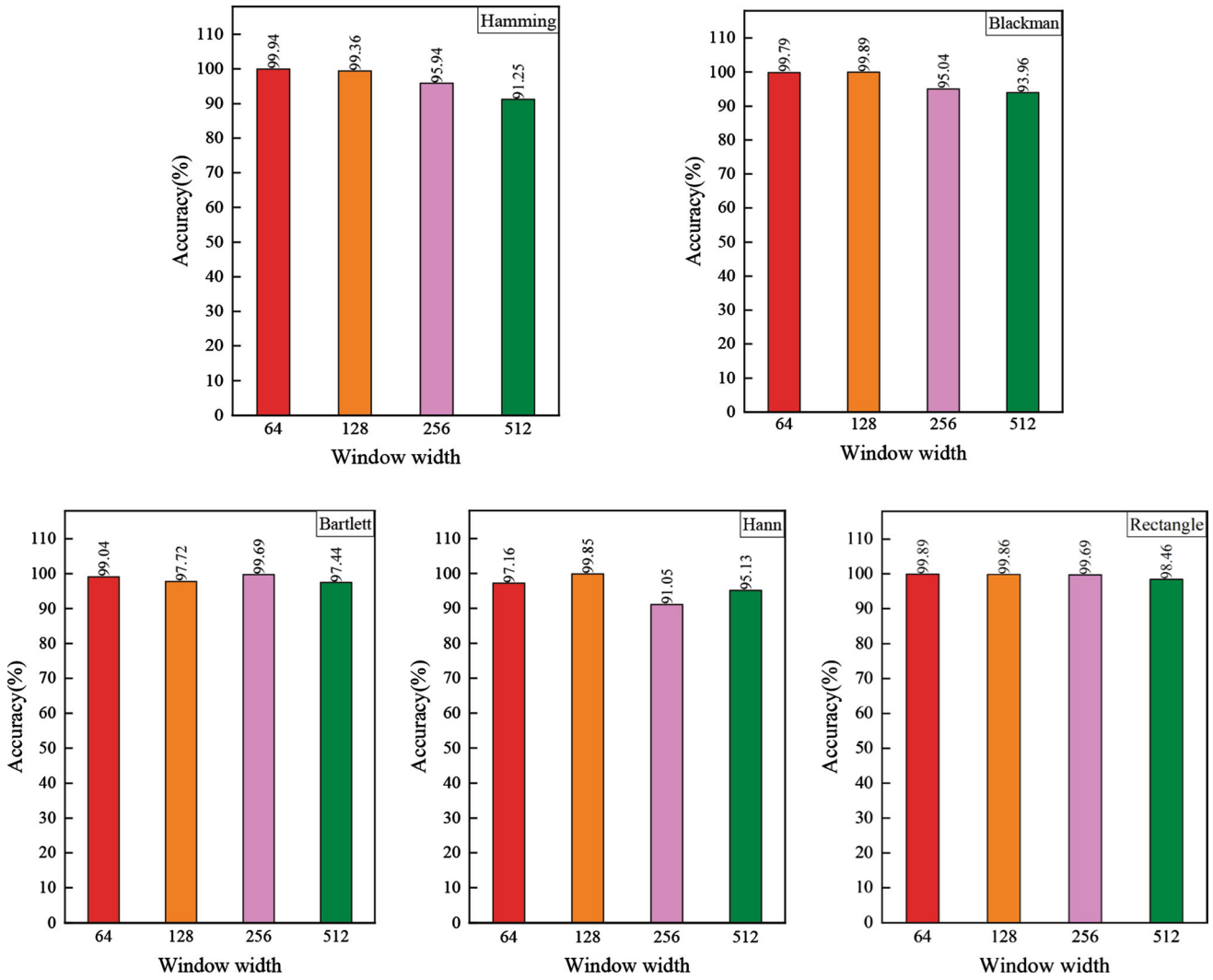

Fig. 10 The classification accuracies of five window functions on CWRU

window is selected, and the window width is 64. The above results can determine the window function type, window width, translation overlap width and the settings of the proposed model parameters.

Figure 13 shows the accuracy curves and loss curves of the training and testing sets. The curves show that the proposed model can obtain satisfactory results, reaching 100% accuracy at the 6th epoch with the loss function tending to zero and stabilizing.

To make the experimental results more intuitive, Fig. 14 shows the confusion matrix of the STFT-CNN verification experiment on the CWRU bearing dataset, with the horizontal coordinates indicating the predicted labels of each fault and the vertical coordinates indicating the true labels of the corresponding faults. The results show that the proposed method achieves 100% prediction accuracy for all fault types, implying that the proposed method can accurately identify the 10 fault status of the rolling bearing.

To further evaluate the performance of the proposed method, we compared the method with eight other methods: ADCNN [36], DBN Based HDN [37], Sparse filter [38], CNN [23], ResNet-DA [39], Self-CNN [40], gcForest [41], and DRL [42]. The classification accuracies of all the methods are shown in Table 5.

It is noted that we repeated the procedure of each method 10 times on the CWRU bearing dataset and used the average accuracy of the 10-time classification results as the performance index of each method to avoid the randomness of the experimental results. It can be easily seen that the accuracies of ADCNN, DBN Based HDN, Sparse filter, CNN, ResNet-DA, Self-CNN, gcForest, and DRL are 98.10%, 99.03%, 99.66%, 99.79%, 99.91%, 97.32%,

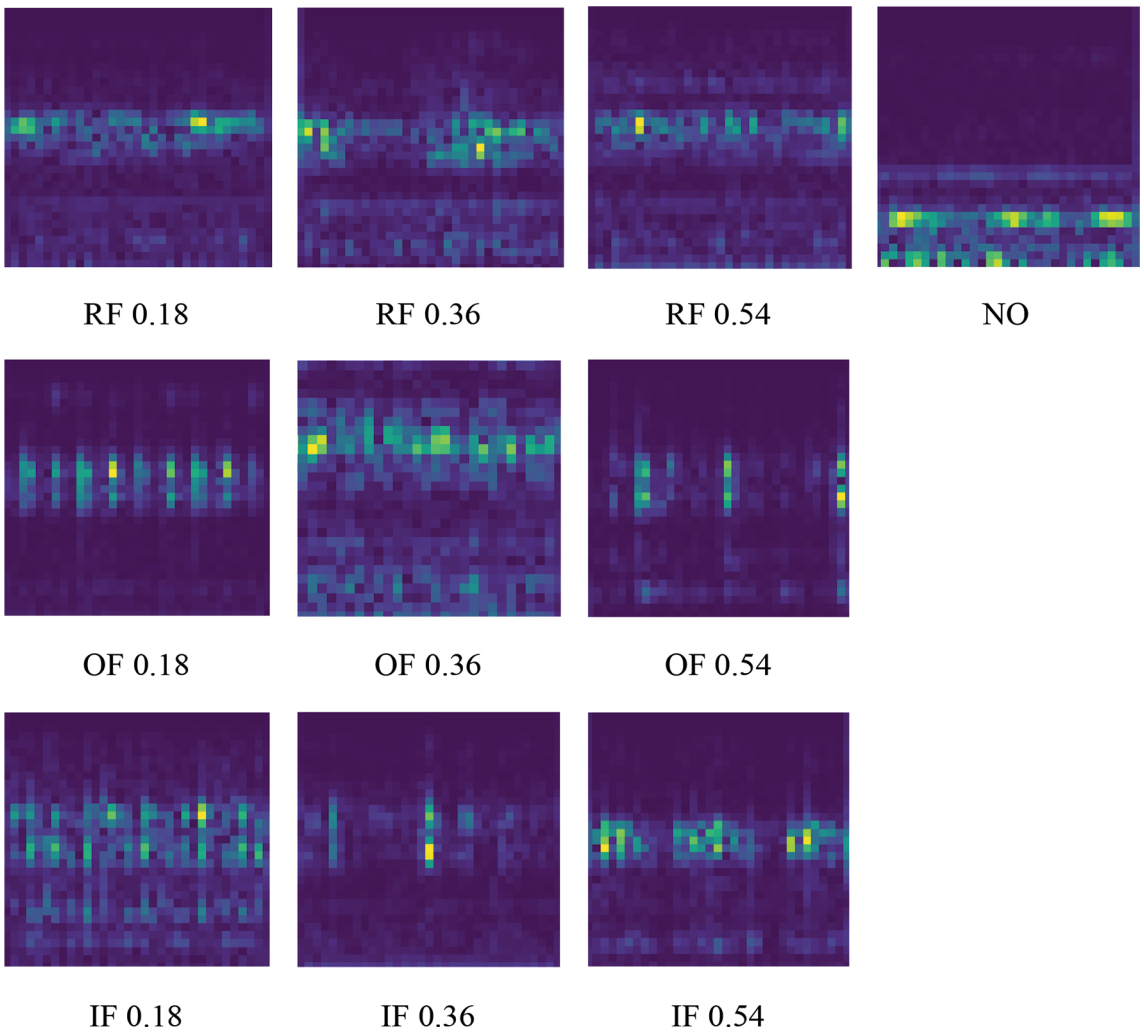

Fig. 11 Time–frequency images of 10 health states

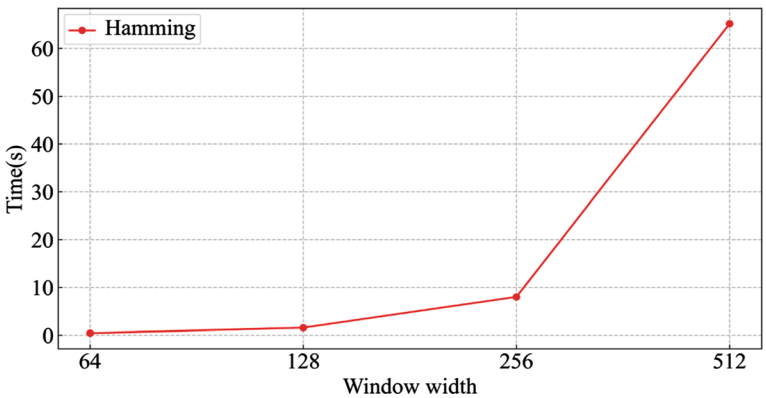

Fig. 12 Time consumption of Hamming window at four window widths

99.20%, and 99.98%, respectively. Among them, ResNet-DA and Self-CNN methods can process 1D signals directly. ADCNN, CNN, gcForest, and DRL methods need to convert 1D signals into two dimensions before processing. The accuracy of the STFT-CNN method is 100%, which is the best compared with eight other methods, indicating the effectiveness of the proposed method.

Case Study 2: Machine Failure Prevention Technology (MFPT) Bearing Dataset

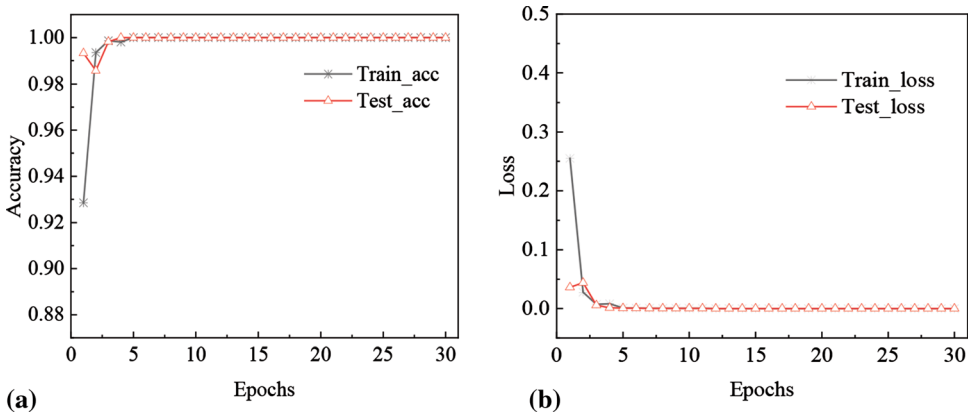

Fig. 13 The accuracy curves and loss curves on CWRU. (**a**) Accuracy curves of training and testing sets and (**b**) Loss curves of training and testing sets

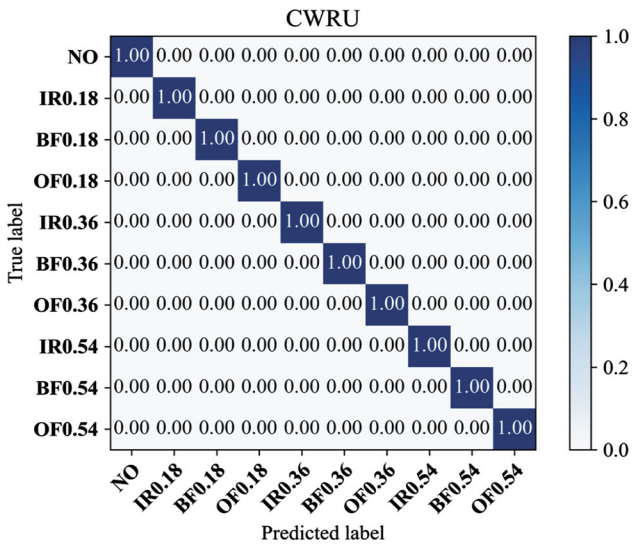

Fig. 14 Confusion matrix of bearing fault identification

Table 5 Classification results of all the methods

Methods	Year	Accuracy, %
ADCNN	2016	98.10
DBN Based HDN	2016	99.03
Sparse filter	2016	99.66
CNN	2018	99.79
ResNet-DA	2018	99.91
Self-CNN	2021	97.32
gcForest	2021	99.20
DRL	2022	99.98
STFT-CNN		100

In this experiment, bearing data were acquired from the Machine Failure Prevention Technology Society [43] public dataset, provided by Dr. Eric Bechhoefer. The installed bearings are NICE bearings, including outer race fault, inner race fault, and normal in three conditions, and the faults are shown in Fig. 15.

The bearing data are divided into four groups.

(1) Three normal states: 270 Ibs of load with a sampling frequency of 97,656 Hz and duration of 6 s.
(2) Three outer race faults: 270 Ibs of load with a sampling frequency of 97,656 Hz and duration of 6 s.
(3) Seven outer race faults: 25, 50, 100, 150, 200, 250, 300 Ibs of load with a sampling frequency of 48,828 HZ and duration of 3 s.
(4) Seven inner race faults: 0, 50, 100, 150, 200, 250, 300 Ibs of load with a sampling frequency of 48,828 Hz and duration of 3 s.

This experiment uses the three normal states, seven outer race faults, and seven inner race faults, and the detailed experimental data are given in Table 6. The number of sampling points selected in this dataset is also 1024. A total of 3717 samples can be generated for three health states. There are 2974 samples in the training set and 743 samples in the test set.

The time–frequency images generated through the Hamming window with a window width of 64 for the three health states of the MFPT dataset are shown in Fig. 16. It can be seen from Fig. 16 that the images are obviously different, and the different colors represent the size of the frequency amplitude. This allows for better classification of three fault types.

We carried out the same processing procedures as the case study 1 for the three health states of the MFPT dataset. The experimental results of the five window functions on the MFPT dataset are shown in Fig. 17.

The experimental results show that the proposed model can achieve identification accuracies of more than 97% for different window functions, window widths, and translation overlap widths on the MFPT bearing dataset. When the window function is Bartlett window, the window width is 128, and the translation overlap width is 114, the identification result is the best, and the identification accuracy reaches 99.98%. The window function determined in the

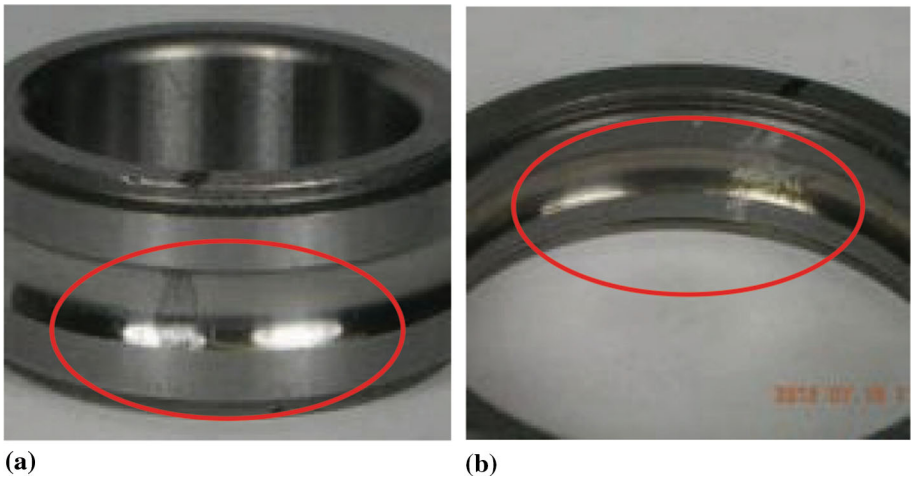

Fig. 15 Types of bearing fault. (**a**) Inner race fault and (**b**) Outer race fault

Table 6 The details of the used datasets in MFPT

Fault type	Sample rate	Load (Ibs)	Label
Three normal states	97,656	270	0
Seven outer race faults	48,828	25, 50, 100, 150, 200, 250, 300	1
Seven inner race faults	48,828	0, 50, 100, 150, 200, 250, 300	2

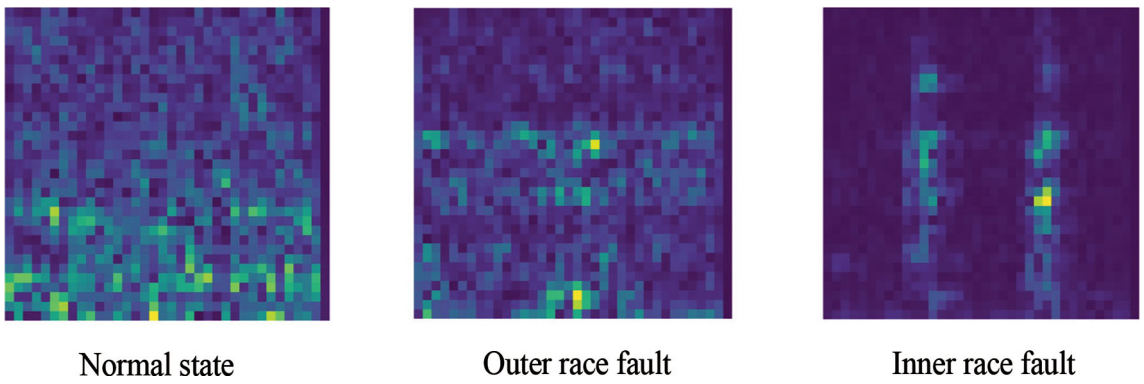

Fig. 16 Time–frequency images of three health states

CWRU bearing dataset is the Hamming window with a window width of 64, and identification accuracy is 99.96%, which is second only to the Bartlett window and can meet the needs of diagnosis. However, in terms of diagnostic efficiency, since the Bartlett window takes 1.3 s in every epoch during the test and the Hamming window only takes 0.21 s, the determined window function is more efficient for diagnosis.

When using the Hamming window, the accuracy curves, and loss curves of the training and testing sets are shown in Fig. 18. The curves illustrate that the network has been trained sufficiently, so the identification results are convincing.

To further verify the superiority of the proposed method, we compare it with four well-known deep learning methods. The processing procedure for the MFPT bearing dataset is absolutely the same as that for the CWRU bearing dataset. And thus, the average diagnostic accuracy and testing time of each method are shown in Table 7.

The accuracies of LeNet, ResNet18, ResNet34, and BiLSTM [17] are 97.65%, 99.25%, 99.79%, and 96.15%, respectively. And meantime, the STFT-CNN method can achieve the accuracy of 99.96%, which is higher than the

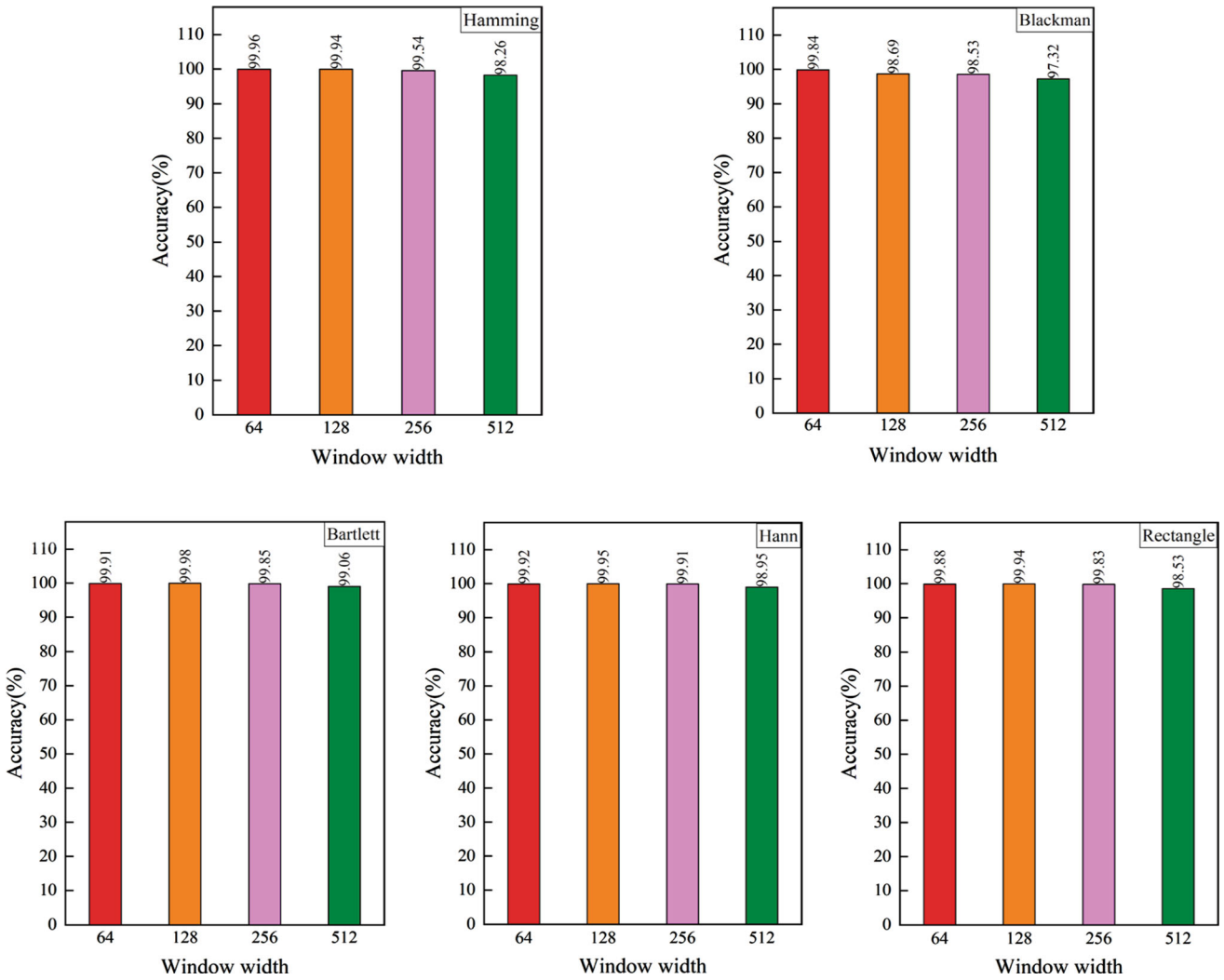

Fig. 17 The classification accuracies of five window functions on MFPT

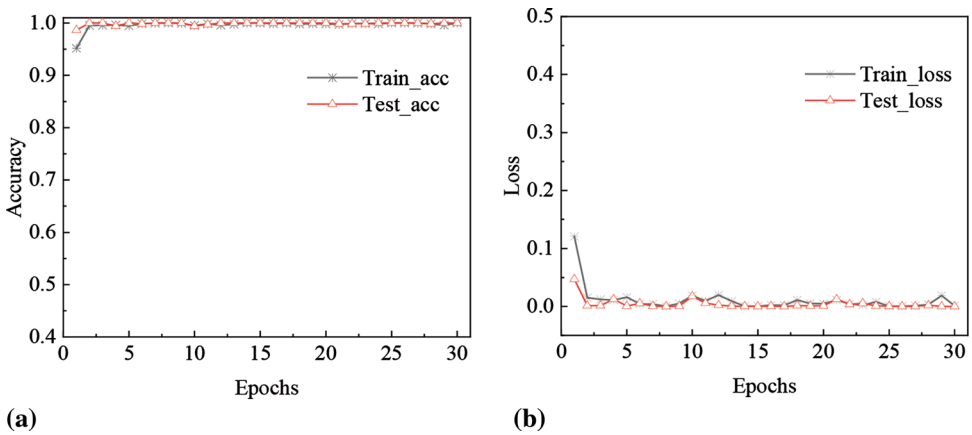

Fig. 18 The accuracy curves and loss curves on CWRU. (**a**) Accuracy curves of training and testing sets and (**b**) Loss curves of training and testing sets

accuracies of the other four methods. The results validate the effectiveness of the proposed method. In addition, the testing time of the STFT-CNN model is 0.21 s in each epoch, which is less than that of the ResNet18, ResNet34, and BiLSTM models (i.e., 0.69 s, 1.07 s, 0.35 s, respectively). It is only 0.09 s longer than the LeNet, but the

Table 7 Comparison with other methods

Methods	Accuracy, %	Testing time(s)
LeNet	97.65	0.12
ResNet18	99.25	0.69
ResNet34	99.79	1.07
BiLSTM	96.15	0.35
STFT-CNN	99.96	0.21

accuracy is much higher, which is acceptable. This experiment also demonstrates that the proposed method is good at diagnosing rolling bearing faults.

Conclusion

In this paper, we propose a fault diagnosis method based on STFT and CNN. This method strikes the problem of feature loss when extracting features of 1D vibration signals. The original time domain signals are converted to time–frequency images via optimized STFT, and subsequently the images are inputted to 2D-CNN for fault classification and identification. To obtain optimal STFT, the window function types, window widths, and translation overlap widths suitable for the 1D bearing vibration signals are determined. In addition, a new CNN architecture is proposed for use in rolling bearings fault diagnosis. The STFT-CNN framework is constructed with two convolutional layers stacked to increase the nonlinear expression capability. The proposed method is tested on two benchmark bearing datasets, and the results show that STFT-CNN model can achieve high fault identification rates with prediction accuracies of 100% and 99.96%, respectively, outperforming other methods. The experimental results demonstrate the effectiveness of the proposed method and strongly suggest that the proposed method has a great potential in fault diagnosis of rolling bearings and rotating machinery.

Acknowledgments The authors thank the supports from the National Natural Science Foundation of China (Grant No. 62241308) and Technological Innovation Guidance Plan of Gansu Province (Grant No. 22CX8GA130). The authors also sincerely thank the Case Western Reserve University Bearing Data Center and the Machine Failure Prevention Technology Society for supplying fault bearing datasets, and the anonymous reviewers for their constructive suggestions and comments on this paper.

Conflicts of interests The authors declare no competing interests.

References

1. T.F. Zhang, S.Y. Liu, S. Zhang et al., Improved sparse representation of rolling bearing fault feature based on nested dictionary. J. Fail. Anal. and Preven. **22**, 815–828 (2022)
2. X.Q. Zhao, Y.Z. Zhang, An intelligent diagnosis method of rolling bearing based on multi-scale residual shrinkage convolutional neural network. Meas. Sci. Technol. **33**, 085103 (2022)
3. I. González-Prieto, M.J. Duran, N. Rios-Garcia et al., Open-switch fault detection in five-phase induction motor drives using model predictive control. IEEE Trans. Ind. Electron. **65**, 3045–3055 (2018)
4. D. Jung, C. Sundstrom, A combined data-driven and model-based residual selection algorithm for fault detection and isolation. IEEE Trans. Control Syst. Technol. **27**, 616–630 (2017)
5. D.C. Zhu, Y.Y. Pan, W.P. Gao, Fault feature extraction of rolling element bearing under complex transmission path based on multiband signals cross-correlation spectrum. J. Fail. Anal. and Preven. **22**, 1164–1179 (2022)
6. H.D. Shao, J.S. Cheng, H.K. Jiang et al., Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. Knowl. Based Syst. **188**, 105022 (2020)
7. Z.W. Gao, C. Cecati, S.X. Ding, A survey of fault diagnosis and fault-tolerant techniques-Part II: fault diagnosis with knowledge-based and hybrid/active approaches. IEEE Trans. Ind. Electron. **62**, 3768–3774 (2015)
8. Z.W. Gao, C. Cecati, S.X. Ding, A survey of fault diagnosis and fault-tolerant techniques-Part I: fault diagnosis with model-based and signal-based approaches. IEEE Trans. Ind. Electron. **62**, 3757–3767 (2015)
9. T. Jin, C. Yan, C. Chen et al., Light neural network with fewer parameters based on CNN for fault diagnosis of rotating machinery. Measurement. **181**, 109639 (2021)
10. J.Y. Jiao, M. Zhao, J. Lin et al., A comprehensive review on convolutional neural network in machine fault diagnosis. Neurocomputing. **417**, 36–63 (2020)
11. X. Wang, D. Mao, X. Li, Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. Measurement. **173**, 108518 (2021)
12. J.S.L. Senanayaka, H.V. Khang, K.G. Robbersmyr, Toward self-supervised feature learning for online diagnosis of multiple faults in electric powertrains. IEEE Trans. Ind. Inform. **17**, 3772–3781 (2021)
13. W. Zhang, G.L. Peng, C.H. Li et al., A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors. **17**, 425 (2017)
14. O. Abdeljaber, O. Avci, S. Kiranyaz et al., Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. J. Sound Vib. **388**, 154–170 (2017)
15. L.Y. Su, L. Ma, N. Qin et al., Fault diagnosis of high-speed train bogie by residual-squeeze net. IEEE Trans. Ind. Inform. **15**, 3856–3863 (2019)
16. H. Wang, Z.L. Liu, D.D. Peng et al., Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. IEEE Trans. Ind. Inform. **16**, 5735–5745 (2020)
17. Z.B. Zhao, T.F. Li, J.Y. Wu et al., Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study. ISA Trans. **107**, 224–255 (2020)
18. O. Janssens, V. Slavkovikj, B. Vervisch et al., Convolutional neural network based fault detection for rotating machinery. J. Sound Vib. **377**, 331–345 (2016)

19. S. Zhang, S.B. Zhang, B.N. Wang et al., Deep learning algorithms for bearing fault diagnostics-a comprehensive review. IEEE Access. **8**, 29857–29881 (2020)
20. M. Bhadane, K.I. Ramachandran, Bearing fault identification and classification with convolutional neural network. International Conference on Circuit, Power and Computing Technologies (ICCPCT), (Kollam, India, 2017).
21. D.T. Hoang, H.J. Kang, Rolling element bearing fault diagnosis using convolutional neural network and vibration image. Cogn. Syst. Res. **53**, 42–50 (2019)
22. Q.B. Wang, B. Zhao, H.B. Ma et al., A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion. J. Mech. Sci. Technol. **33**, 2561–2571 (2019)
23. L. Wen, X.Y. Li, L. Gao et al., A new convolutional neural network-based data-driven fault diagnosis method. IEEE Trans. Ind. Electron. **65**, 5990–5998 (2018)
24. B.X. Zhao, Q. Yuan, Improved generative adversarial network for vibration-based fault diagnosis with imbalanced data. Measurement. **169**, 108522 (2021)
25. D. Verstraete, A. Ferrada, D.E. López et al., Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. Shock Vib. **2017**, 1–17 (2017)
26. Z.Y. Zhu, G.L. Peng, Y.H. Chen et al., A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis. Neurocomputing. **323**, 62–75 (2019)
27. H.F. Tao, P. Wang, Y.Y. Chen et al., An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks. J. Franklin Inst. **357**, 7286–7307 (2020)
28. Y. Lecun, L. Bottou, Y. Bengio et al., Gradient-based learning applied to document recognition. P. IEEE. **86**, 2278–2324 (1998)
29. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, (Lake Tahoe, USA, 2012)
30. C. Szegedy, W. Liu, Y.Q. Jia et al., A going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Boston, USA, 2015)
31. K.M. He, X.Y. Zhang, S.Q. Ren et al., Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Las Vegas, USA, 2016)
32. M. Sandler, A. Howard, M.L. Zhu et al., MobileNetV2: inverted residuals and linear bottlenecks. In IEEE conference on computer vision and pattern recognition (CVPR). (Salt Lake City, USA, 2018)
33. X. Zhang, S. Liu, L. Li et al., Multiscale holospectrum convolutional neural network-based fault diagnosis of rolling bearings with variable operating conditions. Meas. Sci. Technol. **32**, 105027 (2021)
34. J.L. Yang, T.Y. Gao, S.D. Jiang et al., Fault diagnosis of rotating machinery based on one-dimensional deep residual shrinkage network with a wide convolution layer. Shock Vib. **2020**, 1–12 (2020)
35. W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the case western reserve university data: a benchmark study. Mech. Syst. Signal Process. **64–65**, 100–131 (2015)
36. X.J. Guo, L. Chen, C.Q. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement. **93**, 490–502 (2016)
37. M. Gan, C. Wang, C.A. Zhu, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mech. Syst. Signal Process. **72–73**, 92–104 (2016)
38. Y.G. Lei, F. Jia, J. Lin et al., An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans. Ind. Electron. **63**, 3137–3147 (2016)
39. X. Li, W. Zhang, Q. Ding et al., Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. J. Intell. Manuf. **31**, 433–452 (2020)
40. H. Wang, Z.L. Liu, Y.P. Ge et al., Self-supervised signal representation learning for machinery fault diagnosis under limited annotation data. Knowl. Based Syst. **239**, 107978 (2022)
41. Y. Xu, Z.X. Li, S.Q. Wang et al., A hybrid deep-learning model for fault diagnosis of rolling bearings. Measurement. **169**, 108502 (2021)
42. S. Ayas, M.S. Ayas, A novel bearing fault diagnosis method using deep residual learning network. Multimed. Tools Appl. **81**, 1–17 (2022)
43. E.A. Bechhoefer, Quick introduction to bearing envelope analysis MFPT Data (available at: www.mfpt.org/fault-data-sets)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.