


ORIGINAL ARTICLE

Open Access



# Real-Time Intelligent Diagnosis of Co-frequency Vibration Faults in Rotating Machinery Based on Lightweight-Convolutional Neural Networks

Xin Pan<sup>1,2\*</sup> , Xiancheng Zhang<sup>1</sup>, Zhinong Jiang<sup>1,2</sup> and Guangfu Bin<sup>3</sup>

## Abstract

The co-frequency vibration fault is one of the common faults in the operation of rotating equipment, and realizing the real-time diagnosis of the co-frequency vibration fault is of great significance for monitoring the health state and carrying out vibration suppression of the equipment. In engineering scenarios, co-frequency vibration faults are highlighted by rotational frequency and are difficult to identify, and existing intelligent methods require more hardware conditions and are exclusively time-consuming. Therefore, Lightweight-convolutional neural networks (LW-CNN) algorithm is proposed in this paper to achieve real-time fault diagnosis. The critical parameters are discussed and verified by simulated and experimental signals for the sliding window data augmentation method. Based on LW-CNN and data augmentation, the real-time intelligent diagnosis of co-frequency is realized. Moreover, a real-time detection method of fault diagnosis algorithm is proposed for data acquisition to fault diagnosis. It is verified by experiments that the LW-CNN and sliding window methods are used with high accuracy and real-time performance.

**Keywords** Co-frequency vibration, Real-time diagnosis, LW-CNN, Data augmentation

## 1 Introduction

Co-frequency vibration, which is dominated by the rotational vibration component, is the most common form of vibration defects in rotating machinery. Co-frequency vibration accounts for approximately 70% of rotating equipment vibration failures, according to statistical data. When a severe vibration defect arises and the vibration

signal is measured to be dominated by co-frequency components, balancing is often performed on-site or the rotor is sent to the factory. If the underlying source of a vibration defect is not identified, the vibration fault cannot be eradicated after reassembly, which not only wastes production time but is also inefficient. In addition, incorrect assembly might introduce new flaws, which further compounds the problem. With the growth of intelligent equipment, self-diagnosis of the equipment has become a necessity [1, 2]. Self-recovery regulatory systems such as active balancing and automatic balancing algorithms and shafting alignment algorithm to suppress and locate excessive vibration require real-time fault diagnostics [3, 4].

A variety of faults will increase the co-frequency vibration, such as imbalance, misalignment, and looseness. In engineering contexts, it is simple to attribute co-frequency vibration faults to unbalance, making it easy to identify the fault state but challenging to identify the

\*Correspondence:

Xin Pan  
panxin@mail.buct.edu.cn

<sup>1</sup> Beijing Key Laboratory of Health Monitoring and Self-recovery for High-End Mechanical Equipment, College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

<sup>2</sup> Key Laboratory of Engine Health Monitoring-Control and Networking (Ministry of Education), College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

<sup>3</sup> College of Mechanical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China



© The Author(s) 2024. **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/>.

fault category. Methods for rotating equipment problem diagnosis can be split into three categories: Vibration model, signal processing, and intelligent algorithm [5]. The methods based on vibration models, such as employing the finite element method to establish the dynamics equations of the rotor-bearing-foundation systems [6, 7], and identifying faults by calculating the residual forces of faults [8], but the kinetic parameters of various equipment vary greatly and are partially random. Fault diagnosis methods based on signal processing, such as the empirical mode decomposition (EMD) algorithm, decomposing vibration signals into intrinsic mode functions (IMFs) [9], and ensemble empirical mode decomposition (EEMD) [10], need to be improved in different application scenarios [11–13], and the process of extracting IMFs is time-consuming in loop iteration.

Combining the intelligent algorithm's adaptability and processing precision with the signal processing approach makes the algorithm more applicable. Support vector machine (SVM) and convolutional neural network (CNN) are two intelligent defect diagnosis algorithms. Linear SVM is a type of binary classification model whose fundamental model is a linear classifier with the biggest interval in the feature space, which has the advantages of broad applicability and quick training. Pei et al. [14] introduced a joint optimization of SVM and non-dominated sorting genetic algorithm II for the target, and the performance of the jointly optimized classifier was at least 99.9%. Ahmad et al. [15] proposed two new factors for axis misalignment that occurs during overload and employed SVM for fault identification, which has a high degree of accuracy in practice. In the fields of image processing and fault diagnostics, the traditional CNN has found widespread application. Both Wang et al. [16] and Li et al. [17] detected motor bearing vibration fault signals under changing loads using CNN as a classifier. Nowadays, gearbox and bearing fault diagnosis comprise the majority of rotating equipment fault diagnosis [11, 18, 19], and these fault signal characteristics are accompanied by shock signals.

The majority of fault diagnosis studies utilizing deep learning approaches focus on fault signals with impact characteristics, whereas real-time diagnostic investigations for co-frequency faults have not been published domestically or internationally. The co-frequency faults of rotating machinery, such as unbalance, misalignment, and looseness, are frequently characterized by a single spectral characteristic, vibration without shock, and vibration energy values that are close to one another under the same working condition, making them difficult to differentiate using conventional spectrograms. As explained in Section 2.1, the reasons of misalignment fault are different, but the misalignment fault is

evident from the vibration mechanism [8]. In the meantime, there is always more than one problem particular to distinct rotor system damping, stiffness, and installation conditions, thus the effect of the spectrum graph is obscure. The classification method employs machine learning and deep learning techniques to extract the signal characteristics of these co-frequency defects from a high-dimensional space or a deep network. In engineering settings, the majority of the data collected throughout the collection process is normal data, while the fault data is rather tiny, necessitating the need for logical data augmentation. The initial vibration signal in the time domain is intercepted by sliding window processing and can be utilized immediately for classifier. Significant sliding window data augmentation method parameters are explored and statistically analyzed in this article.

Lightweight-convolutional neural networks (LW-CNN) is proposed in this paper, which minimizes the number of layers and parameters to accomplish real-time diagnosis. The CNN structure consists primarily of input and output layers as well as an intermediate convolutional-pooling layer. The LW-CNN suggested in this study, however, reduces the number of convolutional pooling layers while maintaining accuracy.

This paper is organized as follows: In Section 2, a summary of the mechanism of co-frequency faults in rotating machinery is made; meanwhile, the principle of the co-frequency fault diagnosis algorithm LW-CNN and SVM is introduced. In Section 3, the method of real-time diagnosis of co-frequency faults and its evaluation method are proposed, as well as the method of data pre-processing. In Section 4, developed experiments are used to validate the performance of the real-time intelligent diagnosis approach based on the LW-CNN. Finally, the conclusions are in Section 5.

## 2 Rotating Machinery Failure Mechanism and Intelligent Algorithms

The common rotating equipments such as pumps, compressors, fans, etc., are composed of four parts: rotor system, bearing support, seals and other accessories. The rotor is supported by the bearing in the process of movement, the impeller interact with fluid materials, the motor rotor is affected by electromagnetic field and coupling. In short, the rotor-bearing system is affected by centrifugal force, gravity, magnetic force, and interaction between rigid bodies and fluid. The causes of the co-frequency vibration are complex and often result from the comprehensive action of many factors. The common causes are as shown in Figure 1.

The co-frequency vibration of rotating machinery is not always caused by rotor unbalance but also may be due to

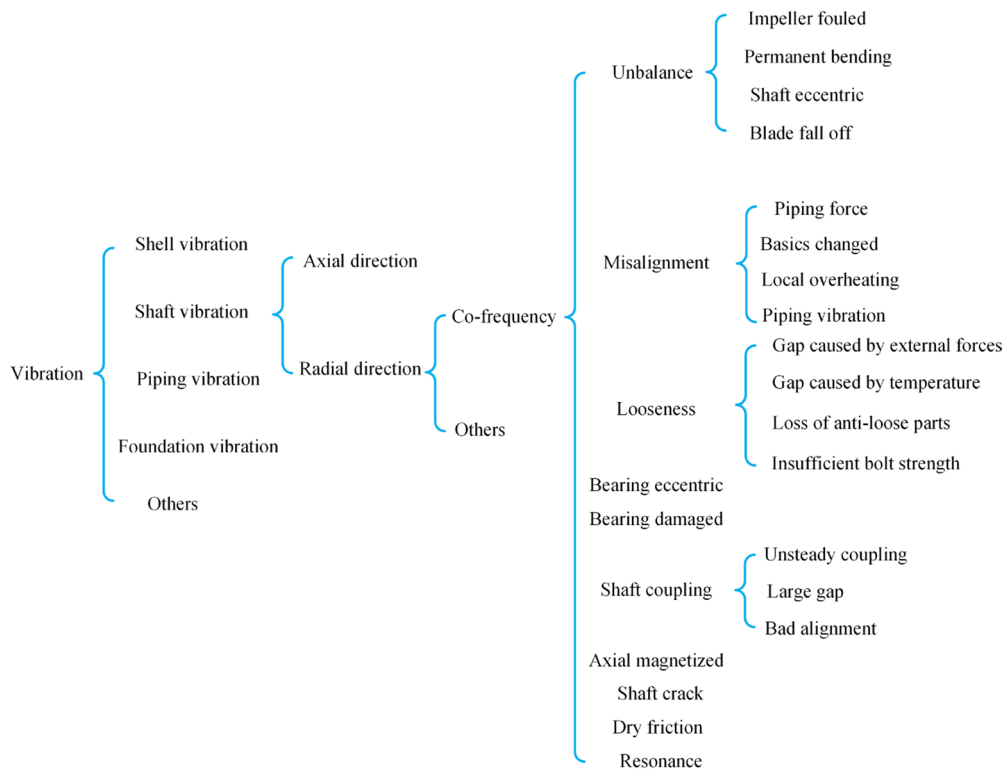


Figure 1 Typical vibration of rotating machinery

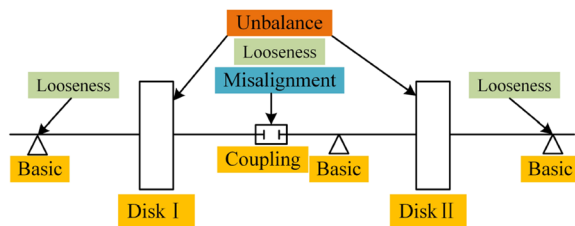


Figure 2 Schematic diagram of the location of unbalance, misalignment and looseness fault in a rotor system

poor alignment, bearing problems, improper clearance of bearing shell, or other reasons. The co-frequency vibration excitation force mainly comes from the radial rotation vector caused by rotor shaft system, bearing failure, interaction between the rotor and stator or transmission parts, shell vibration and resonance, etc. The locations of the rotor system unbalance, misalignment and looseness fault are shown in Figure 2.

### 2.1 Comparison Faults Mechanism and Actual Faults Signal

**Unbalance:** Typical fault of radial rotation vector caused by the rotor itself is unbalance fault. When the rotor has an unbalance fault, i.e., the center of mass and rotation do not coincide, the dynamic unbalance caused by

centrifugal force always has obvious response in each rotation. The intense response at the rotor bending critical speed, which is also related to the unbalance position. The mass eccentric  $m$  causes the bearing to suffer the same rotational periodic  $\omega$  force, as shown in Eq. (1), where  $e$  is the eccentricity.

$$F = me\omega^2. \tag{1}$$

**Misalignment:** Rotating machinery needs couplings to transmit forces and moments. The axis often has a certain misalignment, when the rotor is running under loads. In general, rotors always have center and angular misalignment, and couplings can overcome misalignment to a lesser extent. However, if the misalignment exceeds a certain level, the performance of the bearing will be affected.

Any point in the coupling moves around the center of the shaft from  $a$  to  $b$  by an angle  $\omega't$ , the offset is  $\Delta y$ .

Then, the trajectory of  $O'$  is expressed as follows and is shown in Figure 3.

$$\begin{aligned} x &= \frac{\Delta y}{2} \sin(\omega't - \varphi') = \frac{\Delta y}{2} \sin(2\omega t - 2\varphi), \\ y &= \frac{\Delta y}{2} \cos(\omega't - \varphi') = \frac{\Delta y}{2} \cos(2\omega t - 2\varphi). \end{aligned} \tag{2}$$

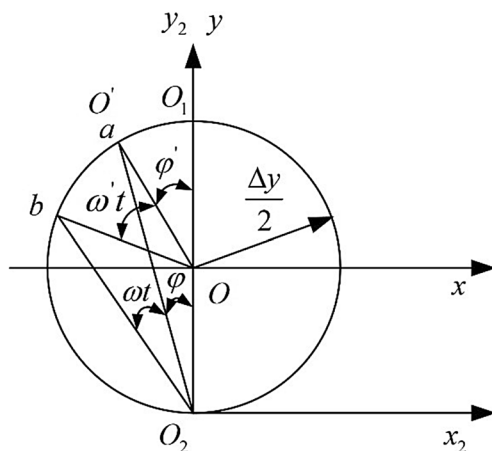


Figure 3 Trajectory of  $O'$

Expression of exciting force:

$$\begin{aligned} F_x &= 2m\Delta y\omega^2 \sin(2\omega t - 2\varphi), \\ F_y &= 2m\Delta y\omega^2 \cos(2\omega t - 2\varphi). \end{aligned} \tag{3}$$

In the case of only central misalignment, the radial vibration frequency caused by the misaligned excited rotor is twice the rotational frequency, with a large of harmonic components.

Looseness: It is a kind of fault caused by the existence of gaps or insufficient rigidity of the connection at the joint surface of the rotor support system, resulting in low mechanical damping and excessive vibration in operation. Looseness, unbalance, and misalignment faults can all lead to co-frequency vibrations in the rotor system. Loose rotor support parts make the connection stiffness and mechanical damping decrease, assuming that the gap is  $\Delta$ , the mass converted to the rotor disc is recorded as  $C_0$ , the rotor's equation of motion is as follows:

$$\begin{aligned} M\ddot{x} + k\dot{x} &= F(x) + Q_x, \\ F(x) &= \begin{cases} kC_0, & x > C_0, \\ kx, & -C_0 < x < C_0, \\ -kC_0, & x < -C_0, \end{cases} \end{aligned} \tag{4}$$

where  $x$  is the displacement of disc mass,  $k$  is the rotor support stiffness, and  $M$  is the disc mass.

Harmonic analysis of the above equation shows that the rotor support system is non-linear, the vibration response contains the co-frequency signal, and  $2X$ ,  $3X$  and other high and low harmonics, in which  $X$  is the rotational frequency.

However, although all the above analyses provide a reference direction for the co-frequency fault diagnosis, the actual collected co-frequency fault signals are unsatisfactory. As shown in Figure 4, the high amplitude frequency is basically concentrated in power frequency and  $2X$ . The amplitude changes of different measurement points under the same fault and the amplitude changes of the same measurement point under different faults are irregular.

## 2.2 Intelligent Diagnosis Algorithm

CNN and SVM are classical classification methods in the fault diagnosis field. Different convolutional structures, convolutional kernels number, convolutional kernel steps, the size of fully connected layers, and different normalization methods, etc., form different CNN. SVM combines other probabilistic conditions [20] or, after combining multiple SVMs, forms different improved SVM [11, 21]. The following briefly describes CNN and SVM principles and signal processing.

### 2.2.1 CNN and LW-CNN

CNN, a kind of neural network containing convolution computation, gets feature maps by convolution window sliding. After that, the feature map used to represent input data features will be obtained, then the pooling calculation is carried out, which calculate through a sliding window without weight.

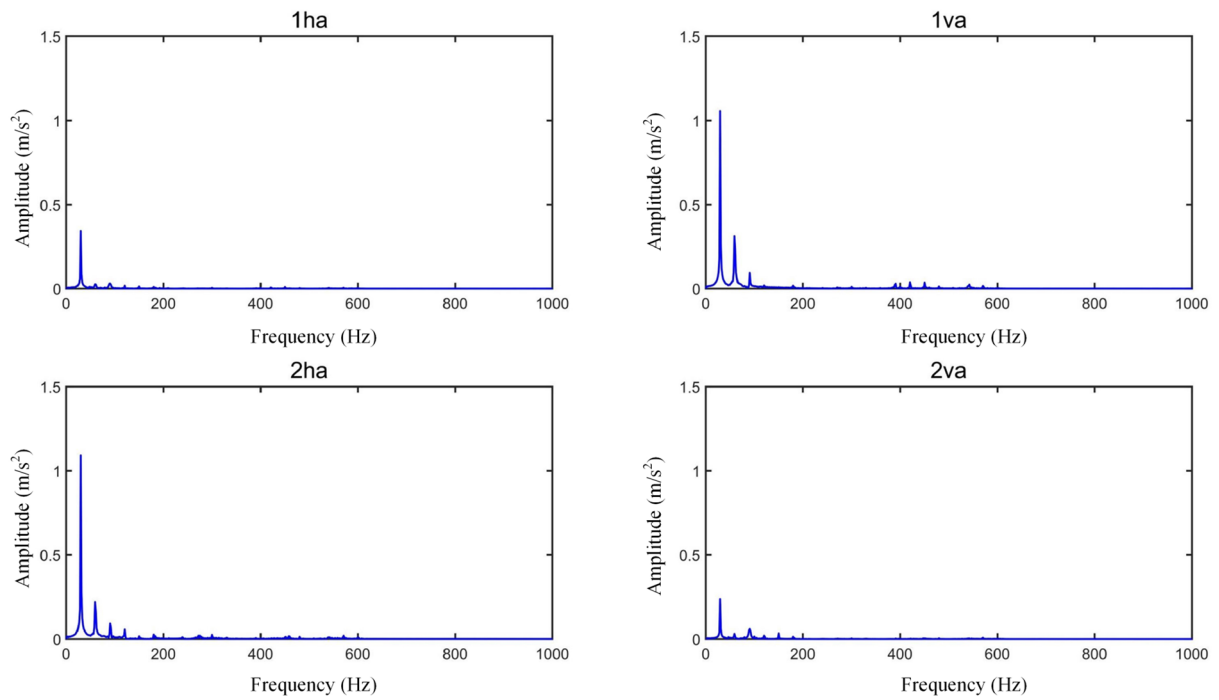
The 2D convolution is defined as follows:

$$z(u, v) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} x_{i,j} \cdot k_{u-i, v-j}, \tag{5}$$

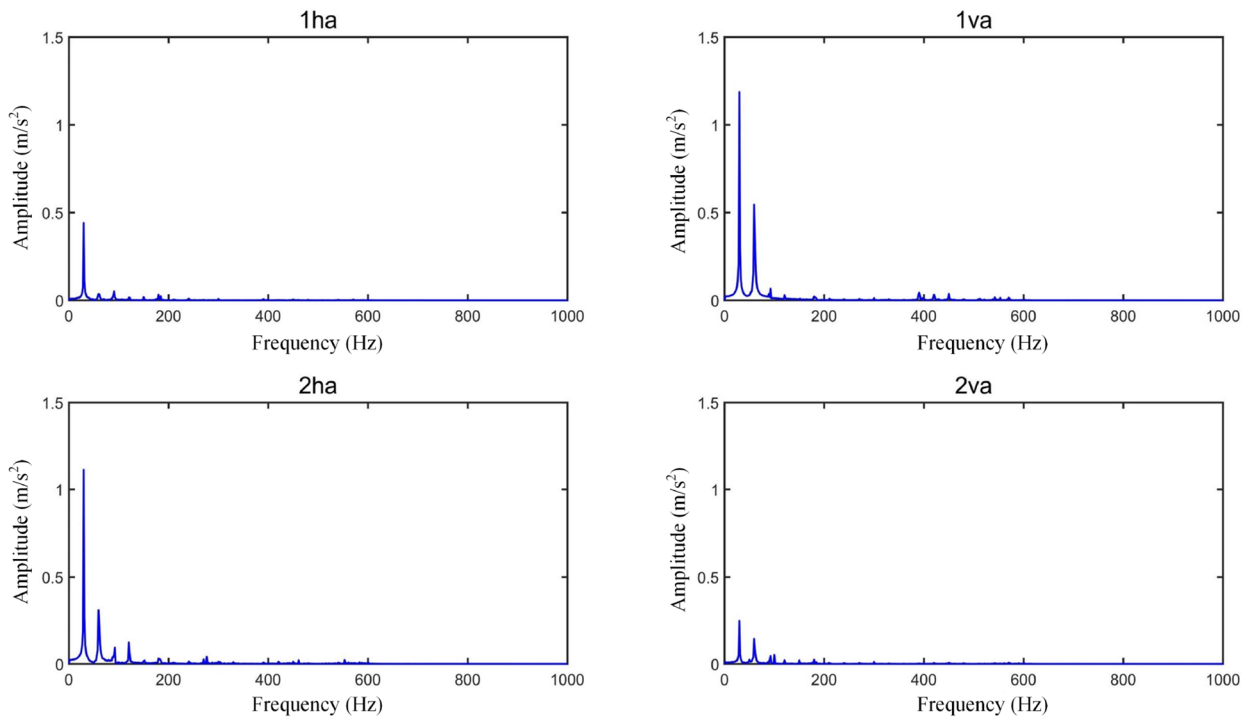
where  $z$  is the result of the convolution of  $x$  on  $k(u, v)$ ,  $i$  and  $j$  are the two directions respectively. For an  $l$  convolutional layer, given the input feature map size is  $(m \times m)$ , the convolutional kernel size is  $(n \times n)$ , and the output of the layer is:

$$\begin{aligned} z_{u,v}^l &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} y_{i+u, j+v}^{l-1} \cdot k_{rot, u-i, v-j}^l \cdot \delta(i, j), \\ \delta(i, j) &= \begin{cases} 1, & 0 \leq i, j \leq n, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

Using average pooling method, the weight of each convolution kernel is  $\lambda^{l+1}$ , and the output of the pooling layer with convolution adds the bias as shown in Eq. (7):

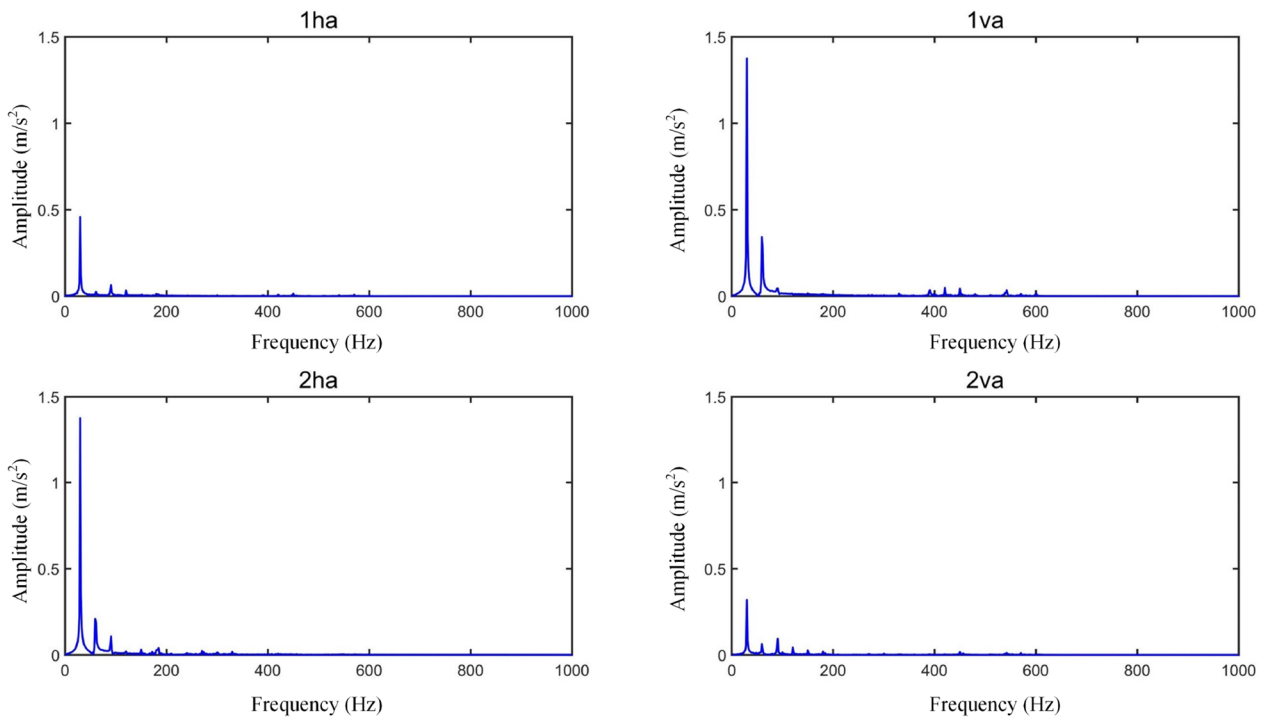


(a)

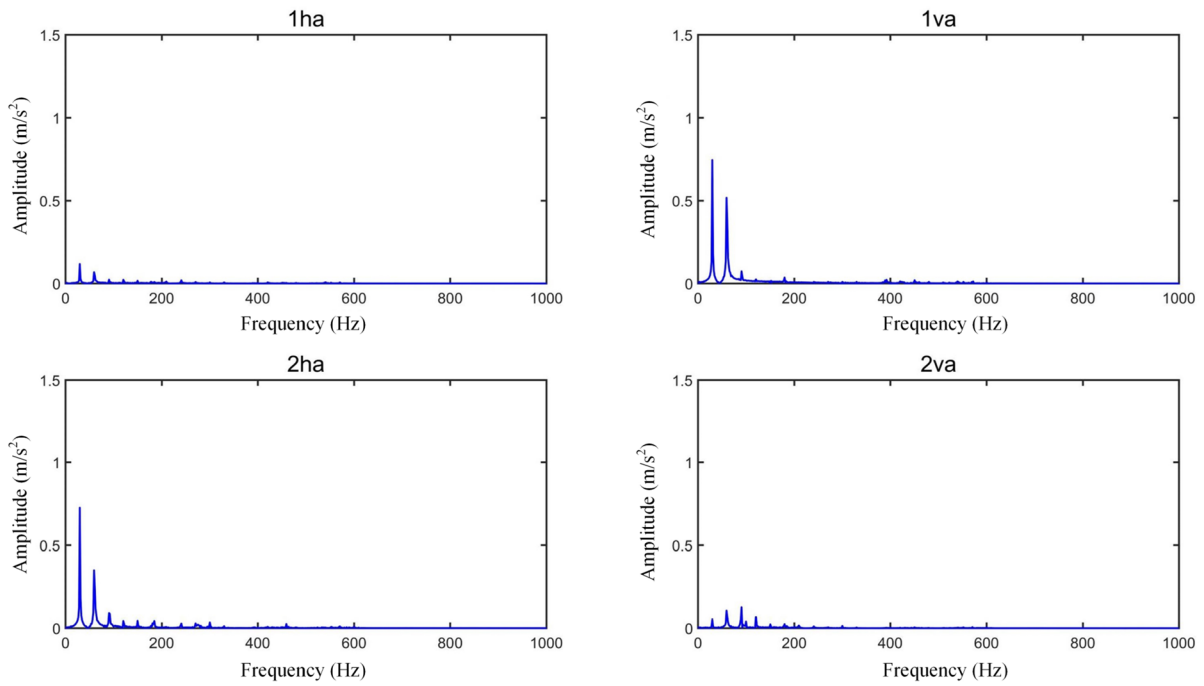


(b)

**Figure 4** Spectrograms of different actual fault signals under different measurement points: (a) Misalignment, (b) Unbalance, (c) Looseness, (d) Normal



(c)



(d)

Figure 4 continued

$$z_{ij}^{l+1} = \lambda^{l+1} \sum_{u=id}^{(i+1)d-1} \sum_{v=jd}^{(j+1)d-1} y_{u,v}^{l+1} + b^{l+1}, \tag{7}$$

$$y_{u,v}^{l+1} = f(z_{ij}^{l+1}).$$

The pooling layer has two functions, one is to allow further feature extraction to reduce the number of weight parameters, and the other is to make the input to the network with translation without distortion. Key differences between CNNs and traditional neural networks are reflected in local connectivity, shared weights, and spatial pooling [22]. These differences also allow CNN to have better translation invariance, fewer parameters, and better generalization performance of the model compared to previous neural networks.

CNN was widely used in computer vision and natural language recognition and are also prevalent in fault diagnosis. In recent years, CNN has made great progress in its development, among which AlexNet, VGGNet, GoogleNet, and ResNet [23–26] are all representative deep networks based on CNN. These methods have hundreds of layers, billions of parameters and require a long training cycle and expensive hardware, thus not acceptable for researchers who want to know test results quickly. The fewer the layers of neural networks, the simpler the structure, the fewer the training parameters, and the shorter the training time, the shorter the prediction time of the trained network, so the lightweight neural network should be considered.

The CNN model mainly consists of convolutional, pooling, and fully connected layers, and Figure 5 shows a LW-CNN model with five hidden layers after the vibration signal is processed. As shown in Figure 4, the 5 hidden layers of CNN are the convolutional layer Conv 1 (first hidden layer), Pooling layer 1 (second hidden layer), convolutional layer Conv 2 (third hidden layer), Pooling layer 2 (fourth hidden layer) and fully connected (FC) layer (fifth hidden layer).

### 2.2.2 SVM

Assume that the input space and feature space of SVM are two different spaces, where the input space is an Euclidean space or discrete set and the feature space is an Euclidean space or Hilbert space. Assume that the elements of the two spaces, linear separable SVM and linear SVM, are in one-to-one correspondence, and map the inputs of the input space to the feature vectors of the feature space.

Rotor fault features are basically nonlinear, and a nonlinear SVM, i.e., a ticked SVM, is required to achieve fault type classification. Assuming that there exists a mapping relationship  $\varphi$  from the input space to the Hilbert feature space, the kernel function satisfies Eq. (8):

$$K(x, z) = \varphi(x) \cdot \varphi(z). \tag{8}$$

The input space corresponds to the Hilbert space, i.e.,  $H$ , so that the original high-dimensional hypersurface in the input space is transformed into a hyperplane in the feature space, thus achieving nonlinear distribution classification [27]. For SVM with kernel functions, the process is essentially the same as for linear SVM, and the presence of the mapping relationship makes the hyperplane equation, as shown in Eq. (9):

$$f(x) = \omega^T \varphi(x) + b. \tag{9}$$

Similar to linear SVM, the optimization model Eq. (10) at this time can be found:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2, \tag{10}$$

$$\text{s.t.}, y_i (\omega^T \varphi(x_i) + b) \geq 1, i = 1, 2, \dots, m,$$

where  $\omega$  is the margin, as shown in Figure 6,  $\varphi$  is the linear expression and  $b$  is the linear expression intercept. To obtain the dual problem under nonlinearity, the inner product of  $x_i$  and  $x_j$  can be equal to the value in the original sample space by the kernel function  $K(x_i, x_j)$ , which greatly reduces the computational difficulty, and the dual problem is shown in Eq. (11):

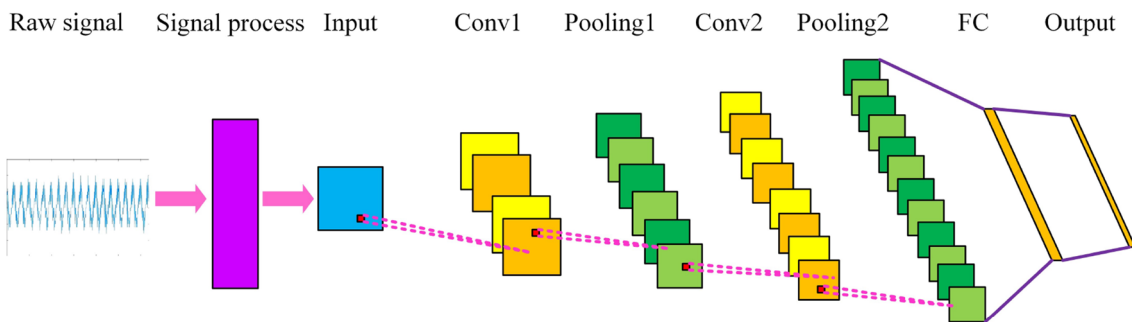


Figure 5 Signal processing and LW-CNN structure

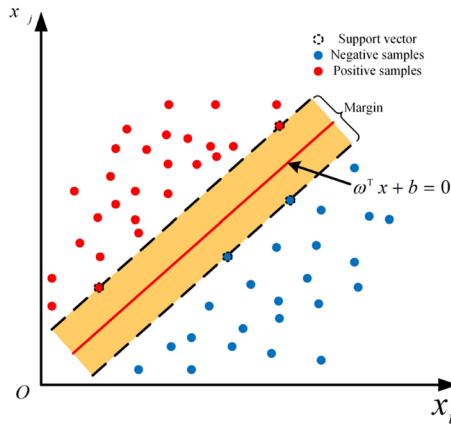


Figure 6 Margin in SVM

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(x_i, x_j), \\ \text{s.t.}, \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & \alpha_i \geq 0, i = 1, 2, \dots, m. \end{aligned} \tag{11}$$

After solving  $\alpha$  by the above equation,  $\omega$  and  $b$  can be found subsequently. The solution leads to the hyper-plane equation:

$$\begin{aligned} f(x) &= \omega^T x + b \\ &= \sum_{i=1}^m \alpha_i y_i \varphi(x)^T \varphi(x_i) + b \\ &= \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b. \end{aligned} \tag{12}$$

From the above calculation process, the appropriate form of  $\varphi(x)$  is crucial for the kernel function  $K$ . It is known from the kernel theorem that any kernel function implicitly defines a reproducing kernel Hilbert feature space. The choice of kernel function causes the biggest difference in SVM. The commonly used kernel functions are linear kernel function, polynomial kernel function, Gaussian kernel function, Laplacian kernel, and Sigmoid kernel function, among which the RBF kernel function is the most commonly used kernel function with strong adaptability and the low number of parameters. With the development of machine learning, SVM as a common classification method has been used extensively, and for a more detailed derivation of formulas in SVM in Ref. [28].

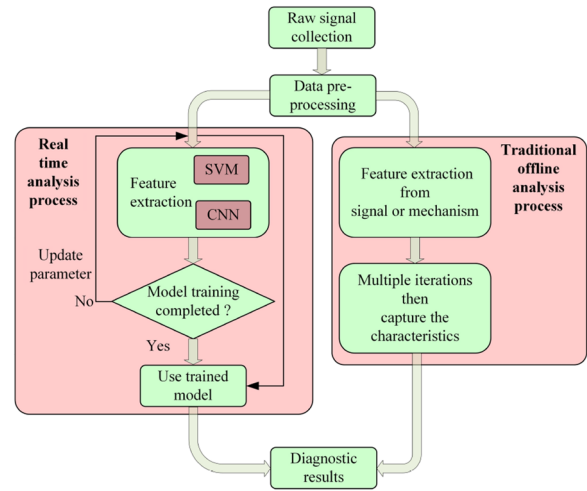


Figure 7 Co-frequency vibration fault diagnosis diagram

### 3 Real-Time Diagnosis Method of Co-Frequency Faults

#### 3.1 Basic Principle

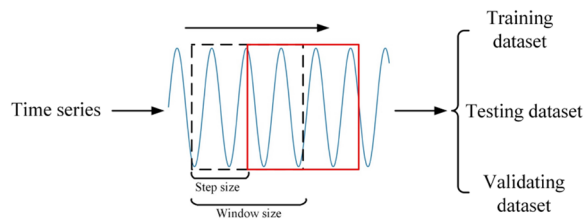
As mentioned above, the actual co-frequency vibration faults do not have obvious characteristics in time and frequency domains, and the types of faults are often not directly determined by spectrum maps. However, deep learning methods are able to extract deep features and implicit features of faults, and these features do not give a specific physical meaning interpretation because the feature extraction process in deep learning methods is a black-box process. The co-frequency fault diagnosis method using the deep learning method is shown in Figure 7. When the raw data is pre-processed, the bad points in the measurement process are removed, and the data is collated into a suitable format and sent to the model for training.

The traditional offline fault diagnosis often relies on specialized knowledge to extract signal or mechanism features. However, these feature extractions often require an iterative process for each diagnosis process. In contrast, real-time fault algorithms are more time-consuming during training, but the diagnosis is very fast after the training is completed and is fully real-time.

#### 3.2 Data Preprocessing

After normalizing the continuous vibration signals  $S_1, S_2, \dots, S_m$ , the data augmentation process is performed on  $S_i$ . The rotor speed is  $n$ , the sampling frequency during the experiment is  $F_s$ , and the row and each sample's row and column size is  $N_f$ . The sample size  $N_f^2$  is the nearest positive integer of sampling points per full turn at the rotate speed  $n$ .





**Figure 8** Schematic of data augmentation by sliding window processing method

**Table 1** Relationship between the correlation coefficient and the degree of correlation

Value of CC	Degree of relevance
$0.8 \leq CC < 1$	High
$0.5 \leq CC < 0.8$	Moderate
$0.3 \leq CC < 0.5$	Low
$0 \leq CC < 0.3$	Very weak

$$\left[ \frac{60F_s}{n} \right] \leq N_f^2 \tag{13}$$

The window slides on the time-domain signals  $S_i$ , and the feature can be ignored when the window slides to the end of the last occurrence of insufficient points. Different move steps and after several iterations can generate a large amount of data to train the model, as shown in Figure 8. In the process of sliding the window to generate data, the size of sliding window step and window size are two important parameters, and the sliding window step is expressed as the ratio of window step and window size (RWW). The correlation coefficient (CC) of the time-domain signal is used to quantitatively characterize the correlation of the data before and after the data augmentation. If the autocorrelation coefficient is close to 1 and the interrelationship is low, the data after data augmentation can be used. From statistical knowledge, the relationship between the magnitude of the correlation coefficient and the strength of the correlation is shown in Table 1.

Two different simulated signals are considered to vary mainly in terms of frequency, amplitude, initial phase

and noise intensity. The simulated Signal  $i$  is characterized by the following equation:

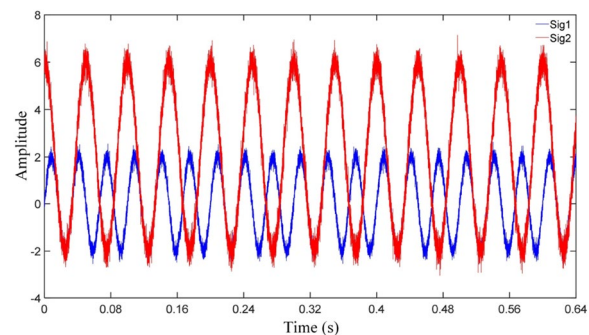
$$\text{Sig. } i = A_i \sin(2\pi f_i + \varphi_i) + d_i + n_i, \tag{14}$$

where  $d_i$  is the offset and  $n_i$  denotes the signal noise, which is expressed by different signal-to-noise ratios (SNR) in the simulation experiment. Two sets of simulated signals, Sig. 1 and Sig. 2, are set up with the parameters shown in Table 2, and time domain plots are shown in Figure 9. The results of the correlation analysis done for the signals under different RWWs are shown in Figure 10. From Figure 10, the autocorrelation coefficients of the two signals under different RWW are above 0.8, and the mutual correlation number is about 0.5, which is medium to low correlation. Therefore, the requirements for the data can be satisfied under different RWWs.

### 3.3 Real-Time Performance Evaluation Method

In order to verify the applicability of the proposed algorithm to online fault diagnosis, the real-time performance of the algorithm needs to be tested. The open-source environment of Python, TensorFlow, and other modules are updated and iterated rapidly to facilitate further improvement of algorithms. In this paper, the algorithm is divided into the following parts: normalization of the original data, transformation into standard data, and the classification process in the classifier.

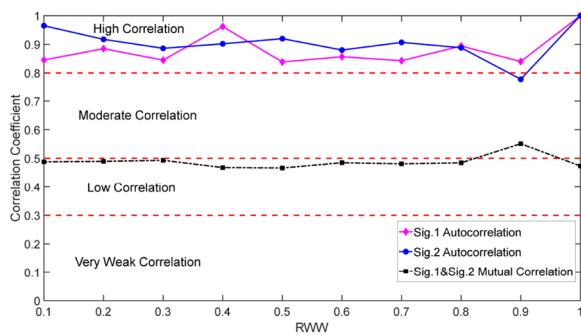
Aiming to test the real-time performance of the algorithm, the number of samples needs to be changed. There are 3 ways to increase the sample points: a) Add one point each time in the sample features based on a certain number from the original data points; b) add one sample



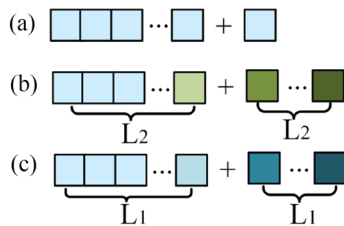
**Figure 9** Time domain mapping of Sig. 1 and Sig. 2

**Table 2** Simulated signal settings

	$f_i$ (Hz)	$A_i$	$\varphi_i$	$d_i$	SNR (dB)	Sampling Points	$F_s$ (Hz)
Sig. 1	30	2	$0^\circ$	0	15	16384	25600
Sig. 2	30	4	$90^\circ$	2	10	16384	25600



**Figure 10** Autocorrelation coefficients and intercorrelation numbers of Sig. 1 and Sig. 2 at different RWW



**Figure 11** Schematic diagram of real-time data addition method: (a) Method a, (b) Method b, (c) Method c

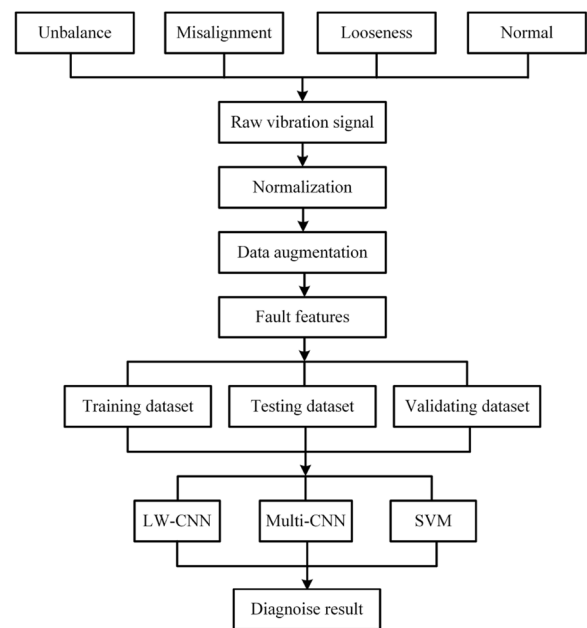
feature size ( $L_2$ ) each time based on a certain number from the original data points; c) add the number of sampling points ( $L_1$ ) in each sampling time. The fault diagnosis time within each sampling time is tested to ensure that the algorithm time is much less than the sampling time. Figure 11 illustrates the above 3 methods.

### 4 Experimental Verification

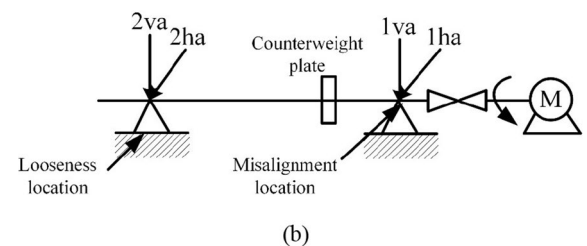
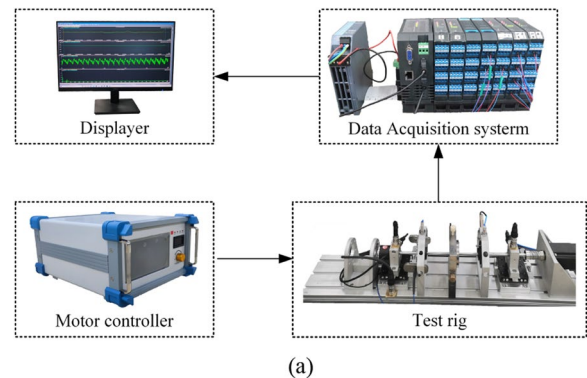
Specific experiments are used to verify the above process. The experimental procedure is shown in Figure 12.

#### 4.1 Experiment Preparation

The rotor test bench shown in Figure 13 mainly verifies this experimental verification. The test bench is driven by a motor, and the first-order critical speed of the actual test bench rotor is about 2700 r/min. Both ends of the rotor are supported by sliding bearings, and various faults are simulated on the experimental bench. A simple diagram of the test bench is shown in Figure 13(b), which clearly shows the location of the fault and sensor position. The experimental data was collected by BH7000. Misalignment fault is simulated and measured by adjusting a micrometer screw at the coupling position; unbalance fault is to add a fixed mass of counterweight block to the 0 phase; loose fault is to loosen the screw on the



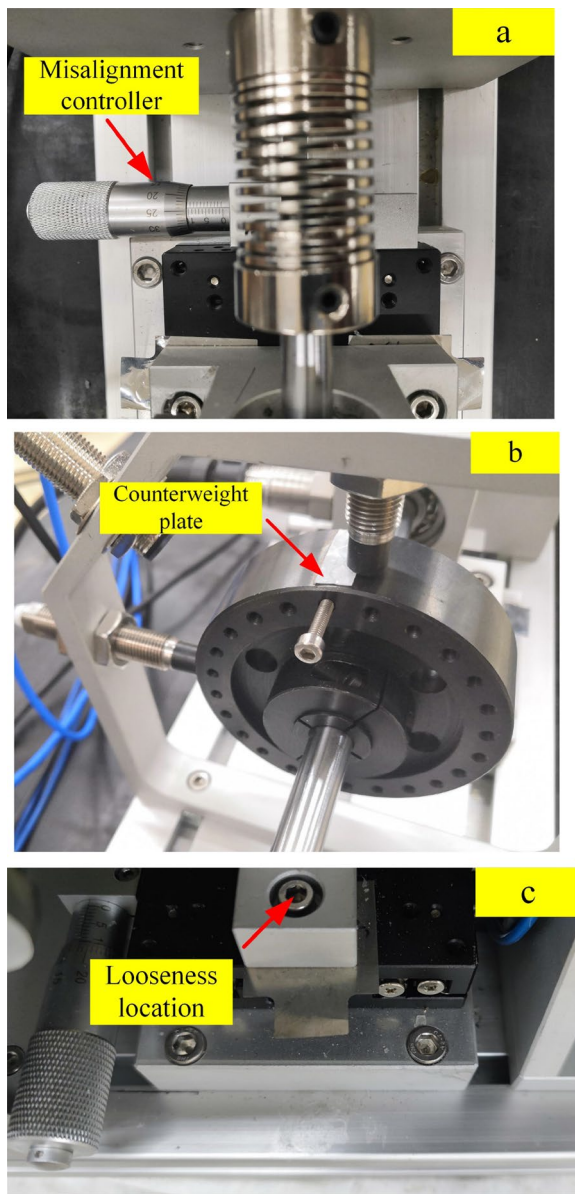
**Figure 12** Flowchart of diagnosis process



**Figure 13** Rotor failure test bench: (a) Actual rotor test bench, (b) Schematic diagram of rotor test bench

bearing support side of the rear end, the test bench fault simulation is shown in Figure 14.

The actual faults often do not occur in isolation, and the same type of failure also has different causes of failure; in most cases, one fault is dominant, and others coexist.



**Figure 14** The actual form of fault simulation used in this experiment: (a) Misalignment, (b) Unbalance, (c) Looseness

Hence the dynamic balance and alignment of the test rig should be completed before fault data collection, regarded as experimental data only if the test bench vibration value is in the acceptable range. After each fault simulation test, the fault is removed instead of repeating the dynamic balancing and alignment experiments on the test bench, and the trial continues with the following fault. By this operation, each major fault contains more than one faults.

Misalignment fault point is set close to the coupling position because the vibration intensity at the free end is greater when the rotor vibrates, also in loose fault experiment. For unbalance fault, the same unbalance is added to the middle of the shaft position to provoke greater vibration. To ensure that the unbalance fault in the experiment is not significantly different from other faults, the unbalance fault is applied at the motor side. The parameters of each fault in this experiment are shown in Table 3 and experiment is conducted at a constant speed of 1800 r/min (slightly below the first-order critical speed). The sampling frequency  $F_s$  and sampling points  $N_s$  are 25.6 kHz and 16384, respectively. The amount of data collected is shown in Table 4.

#### 4.2 Experimental Data Processing

After processing the data sampled from different faults, different measurement points, and the same time point, the time domain waveforms are shown in Figure 15, where CH1–CH4 respectively represent sensors 1ha, 1va, 2ha, 2va. The time domain waveforms are filtered by sampling moving average, filtering higher frequency. The time domain waveform from filtered data are shown in Figure 16. The frequency domain plot is shown in Section 2.1. Based on the time domain and frequency spectrum, the 3 faults are highly similar.

As mentioned in the previous section, the sliding window method is adopted for data argumentation, the window size in this experiment is 900. The training and test sets use using 3/4 and 1/2 RWW, and the validation set uses using 1/3 RWW for one of the measurement points, so the size of the resulting data set is shown in Table 5. The autocorrelation coefficients and intercorrelation numbers under different RWWs are shown in Figures 17 and 18.

**Table 3** Experimental fault type, magnitude setting

Label	Rotor state	Speed (r/min)	Fault name	Sampling rate	Sampling points
0	Misalignment	1800	0.75 mm	16384	25600
1	Unbalance	1800	$2g \angle 0^\circ$	16384	25600
2	Looseness	1800	1 turn	16384	25600
3	Normal	1800	–	16384	25600

**Table 4** Summary of fault data collected from experiments

Type	Data size
Misalignment	$4 \times 64 \times 16384$
Unbalance	$4 \times 47 \times 16384$
Looseness	$4 \times 43 \times 16384$
Normal	$4 \times 25 \times 16384$

Data format: No. measuring points  $\times$  No. sampling periods  $\times$  Ns

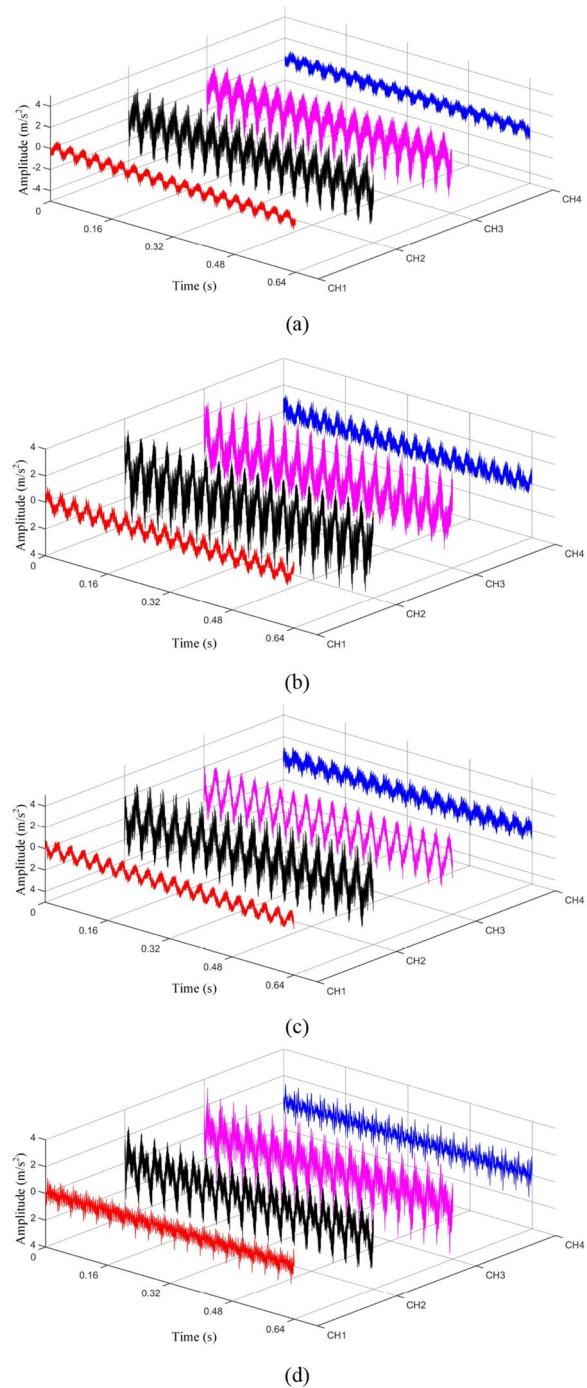
From Figures 17 and 18, it can be concluded that the augmented data with different faults are well-distinguished in 2D space. The similarity between different faults is weak or even extremely weak, while the similarity between the same faults is in medium and high similarity. In Figure 17, the autocorrelation coefficient of the 2va measurement point is always slightly lower than the other measurement points. This is because, compared to the other measurement points, the 2va is further away from the fault position, the fault signal transmission path is more complex and the fault vibration signal is attenuated, resulting in a smaller signal autocorrelation coefficient. The high correlation coefficient values between the unbalance fault and the loose fault in Figure 18(b) are mainly because both have similar fault magnitude and characteristics. Therefore, the data set after data augmentation can be used for further diagnosis.

**4.3 Fault Diagnosis Verification**

The experiments were performed on Python 3.7.10, in a 64-bit Window environment with a RTX 2080Ti GPU. The CNN is performed on Keras 2.3.1, TensorFlow-gpu2.2.0 environment, and the SVM uses scikit-learn0.24.1. The SVM and two different CNN structure parameters and specific parameter values are shown in Tables 6 and 7.

In the actual fault diagnosis process, the amount of fault data is often small, and the suddenness of the fault requires the diagnosis method to be lightweight. The algorithm needs to be able to maintain a high correct rate even with little data. In order to solve this practical problem, the following work was done in this experiment during the fault diagnosis. As can be seen from Table 7, compared to the multi-CNN, the LW-CNN reduces one layer of convolutional pooling and keeps the other parameter structure unchanged. The final total parameters are reduced from 3.37 million to 1.3 million.

- a) The difference between the two CNNs is that the Multi-CNN has more network layers and parameters. In order to compare the different layers of the CNN network, two CNN are designed respectively



**Figure 15** Time domain waveforms of different fault types: (a) Unbalance, (b) Misalignment, (c) Looseness, (d) Normal

- as LW-CNN and multi-CNN. Thereinto, multi-CNN has more hidden layers and more parameters.
- b) The algorithm does complete training, testing, and validation at a gradual and uniform increase of the

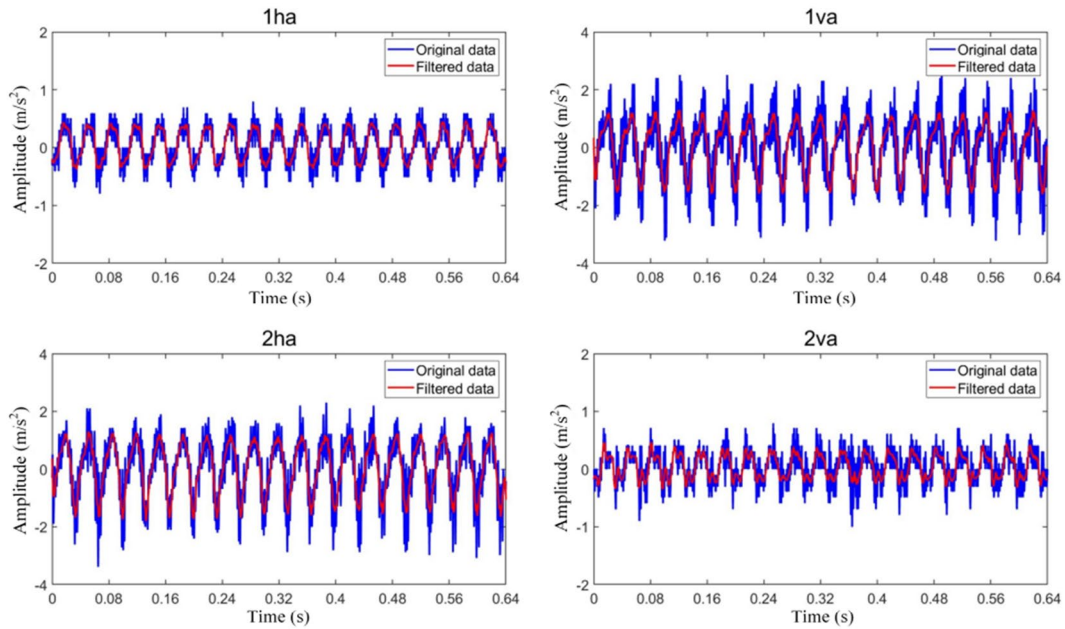


Figure 16 Filtered time domain waveform of each measurement point under unbalance fault

Table 5 Size of dataset after data augmentation

	Dataset size			
	Looseness	Misalignment	Unbalance	Normal
Training	8145 × 900	5979 × 900	5469 × 900	3174 × 900
Testing	3492 × 900	2562 × 900	2343 × 900	1362 × 900
Validating	2343 × 900	2562 × 900	2343 × 900	1362 × 900

data volume from 25% to 100% of the total data volume.

- c) To avoid accidental error, the algorithms must be repeated 10 times on the same data.

For more quantitatively describing the classification performance of different algorithms under different data volumes, the F1 value, a common classification evaluation parameter in machine learning, is used for a more comprehensive evaluation. For a specific binary classification problem, the samples are labeled with positive or negative labels (1 or 0), and the algorithm outputs a predicted label after training. Compared with the actual label, the labeled results have four cases, as shown in Figure 19, which are recorded as true positive (TP), false positive (FP), false negative (FN), false negative (FP), and true negative (TN). The accuracy of the algorithm’s classification (Eq. (15)) and the check-all rate (Eq. (17)) are often two contradictory quantities, and the summed average F1 score of the two (Eq. (18)) is used to do the evaluation, as shown in Eqs. (15), (16), (17), (18):

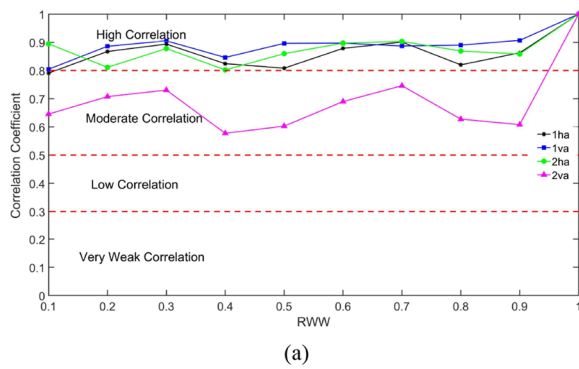
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{15}$$

$$Precision = \frac{TP}{TP + FP}, \tag{16}$$

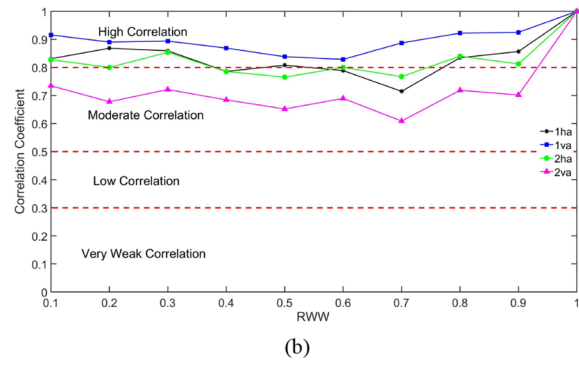
$$Recall = \frac{TP}{TP + FN}, \tag{17}$$

$$F1score = \frac{2Precision \cdot Recall}{Precision + Recall}. \tag{18}$$

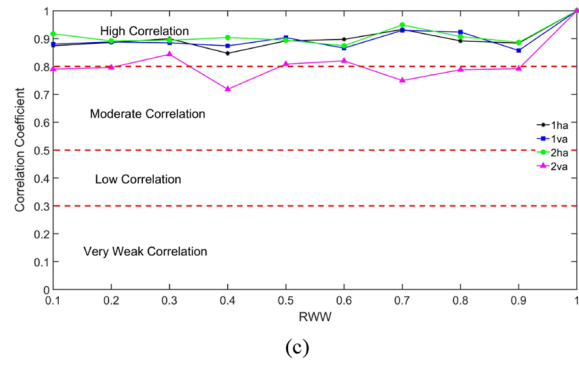
The performance of each method on the training, test, and validation sets is shown in Table 8, and the specific performance on the test set is shown in Figures 20, 21, 22, 23. In Table 8, the overall accuracy of SVM is above 85%. The accuracy performance of SVM improves by about 10% as the data volume increases, and the F1 value also improves by about 10% as the data volume increases. In contrast, the performance of the CNN network is stable with increased data volume. At 25% data volume, the accuracy of LW-CNN is above 95%, most of which are at 99% level with excellent performance, and the F1 score is above 0.98 on average. By analyzing Figures 20, 21, 22, 23, the fault diagnosis accuracy of CNN under the four data amounts is basically above 95%, which has excellent performance. However, in SVM, the imbalance and misalignment diagnosis accuracy are below 85% with 100%



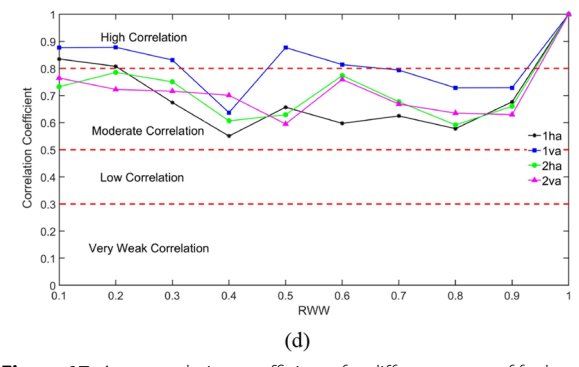
(a)



(b)

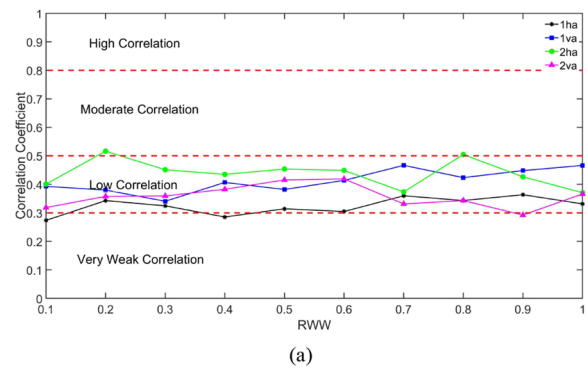


(c)

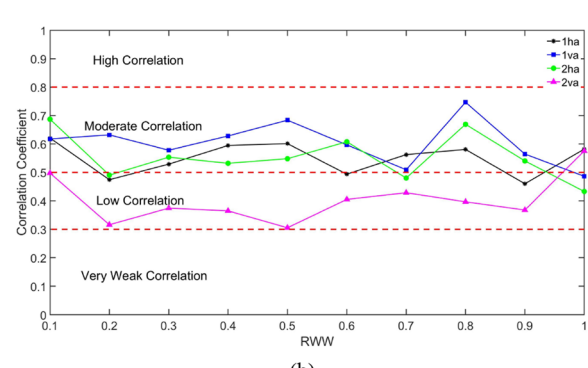


(d)

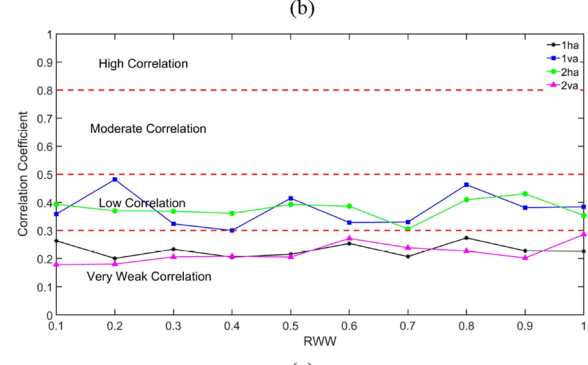
**Figure 17** Autocorrelation coefficients for different types of faults at different RWW: (a) Unbalanced, (b) Misaligned, (c) Looseness, (d) Normal



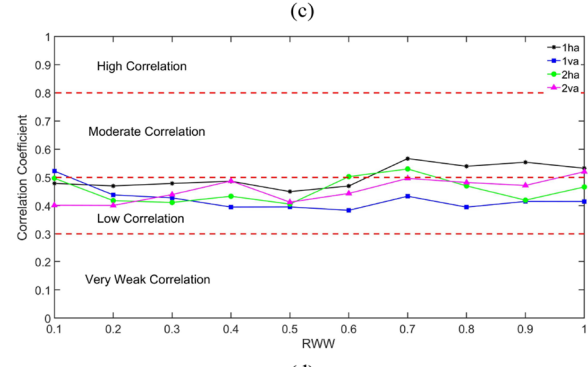
(a)



(b)



(c)



(d)

**Figure 18** Value of interrelationships between different RWW: (a) Misalignment and normal, (b) Unbalance and looseness, (c) Looseness and normal, (d) Misalignment and unbalance

**Table 6** Parameters of machine learning classifier and parameter values

	Parameter	Parameter value
SVM	Regularization parameter	0.3
	Kernel type	RBF
	Kernel coefficient	Scale
	Other parameters	Default

data volume or less. Regarding parameter number, the LW-CNN has 1.3 million, and the multi-layer CNN has 3.37 million in comparison.

After processing the time domain signal with the sliding window method, the different co-frequency vibration faults have low correlation values with each other. The signal processing stage not only improves the ability to discriminate between different vibration fault forms but also achieves the purpose of data augmentation. According to the data analysis in Section 4.3, it is obvious that CNN performs better when the data volume is small, with a better F1 score and better generalization performance. For CNN, although LW-CNN has one less

		Actual label	
		1	0
Predicted label	1	True Positive	False Positive
	0	False Negative	True Negative

**Figure 19** Filtered time domain waveform of each measurement point under unbalance fault

convolutional layer and pooling layer than multi-CNN, the performance is no worse than multi-CNN. Therefore, the data processing method time domain truncation method produces data features with better adaptability to different faults and can achieve better results using lightweight neural networks.

#### 4.4 Real-Time Analysis of Diagnostic Methods

As described in Section 3.3, the feature size in this experiment is 900, the sampling time is 0.64 s, and the sampling

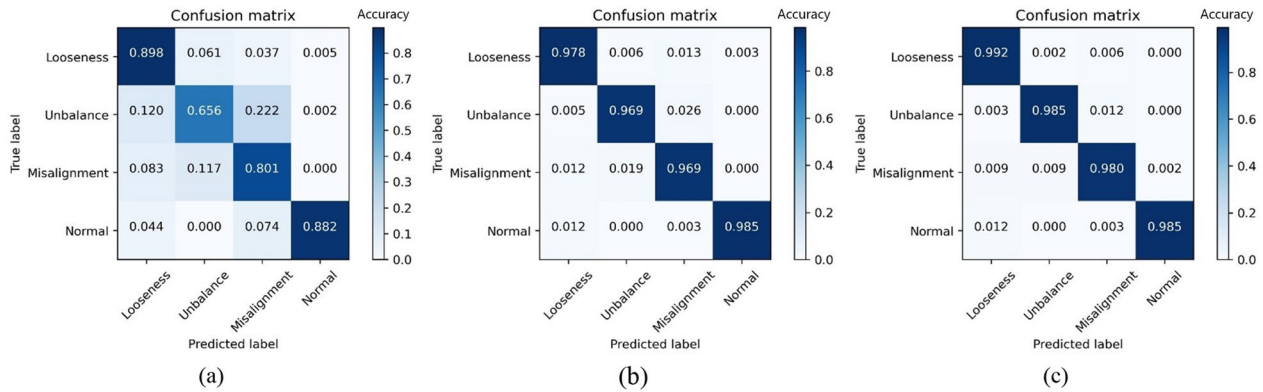
**Table 7** Parameters of convolutional neural network and parameter values

Layer	Parameter	LW-CNN	Multi-CNN
#1 Convolution-pooling	Number of convolution kernel	32	32
	Convolution kernel size	3	3
	Activation function	ReLu	
	Convolution stride	1	
	Pooling size	2	
	Pooling stride	2	
#2 Convolution-pooling	Number of convolution kernel	64	
	Convolution kernel size	3	
	Activation function	ReLu	
	Convolution stride	1	
	Pooling size	3	2
	Pooling stride	2	2
#3 Convolution-pooling	Number of convolution kernel	–	128
	Convolution kernel size	–	3
	Activation function	ReLu	
	Convolution stride	–	1
	Pooling size	–	2
	Pooling stride	–	3
Fully connection	Units	800	1024
	Activation function	ReLu	
Output layer	Units	4	
	Activation function	SoftMax	
Padding	–	Same	
Optimizer	–	Adam	
Loss function	–	Cross entropy	
Total parameters	1.30 million	3.37 million	

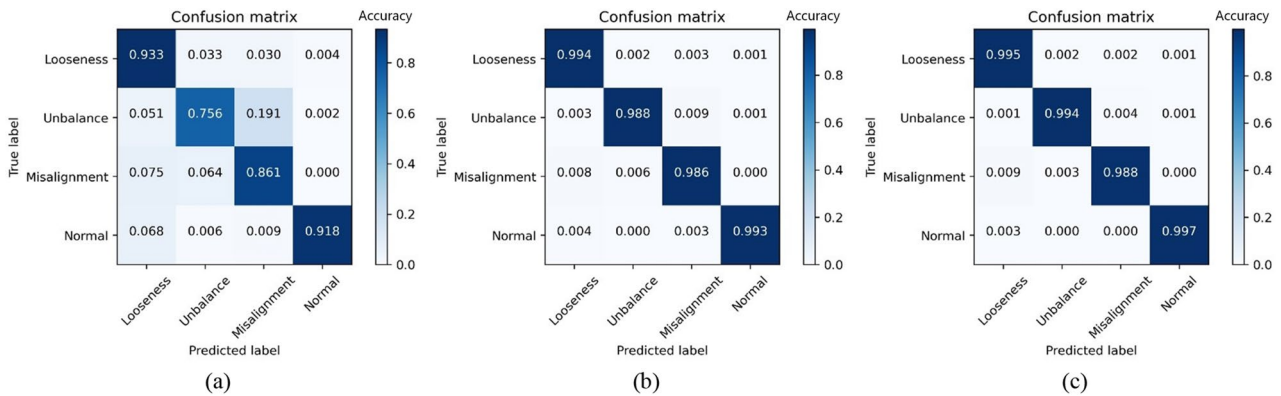
**Table 8** Performance of different methods with different data volumes

Method	Data size			
	25%	50%	75%	100%
SVM	86.2/83.0/83.6/0.82	88.7/89.9/88.3/0.88	93.0/90.1/91.9/0.91	94.5/93.1/93.8/0.93
LW-CNN	99.8/97.9/98.3/0.98	99.7/98.5/98.5/0.98	100/99.1/99.4/0.99	99.9/99.4/99.5/1.0
multi-CNN	100/98.6/99.1/0.99	99.7/99.1/99.2/0.99	99.9/99.6/99.6/1.0	100/99.7/99.8/1.0

Data format: Training accuracy/Testing accuracy/Validating accuracy/F1 score



**Figure 20** Algorithm performance under 25% dataset: (a) SVM, (b) LW-CNN, (c) Multi-CNN



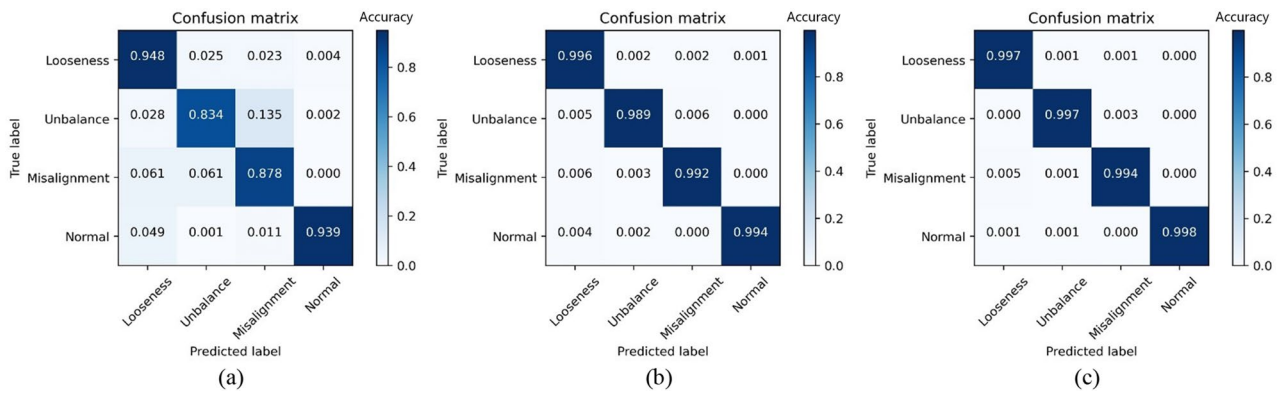
**Figure 21** Algorithm performance under 50% dataset: (a) SVM, (b) LW-CNN, (c) Multi-CNN

points is 16384. Therefore, the 3 methods are tested separately in tensorflow environment as shown in Figure 24.

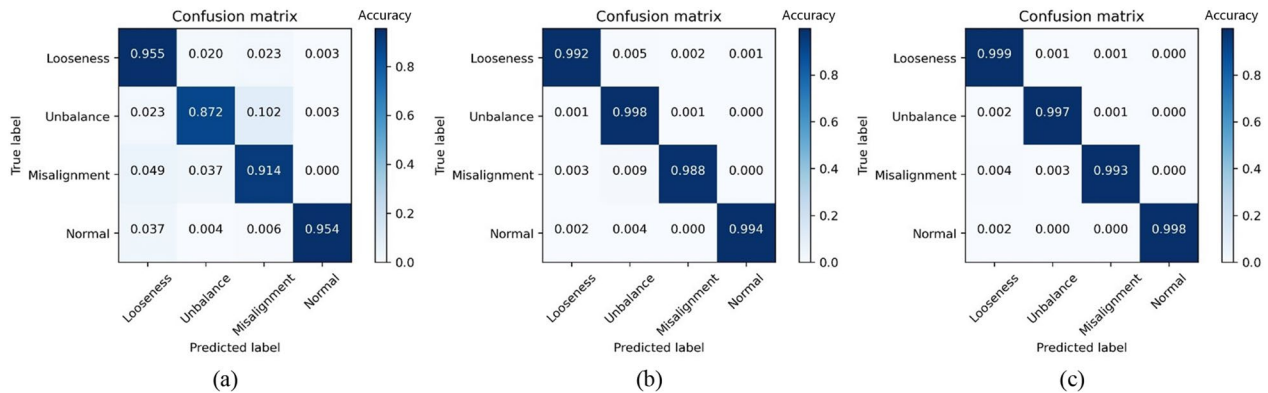
In fact, the sample is increased using the 3 methods before mentioned and repeat 10 times. The final result is shown in Figure 24. For method (a), because the window size and the moving step are fixed in the data processing stage, the amount of standard data entering the classifier does not change even though the number of points increases in a certain range; for method (b), 900 points are added because the size of the reference

window is set, and not much standard data is added after data processing. The data amount added is small for methods (a) and (b). The time change is not visible due to the short time of the algorithm itself, which is regarded as a normal phenomenon because of the short time and the small amount of sample change. Method (c) is considered in one sampling time, which conforms to the actual fault diagnosis process. From Figure 24, the average processing time is about 0.01 s, which can





**Figure 22** Algorithm performance under 75% dataset: (a) SVM, (b) LW-CNN, (c) Multi-CNN



**Figure 23** Algorithm performance under 100% dataset: (a) SVM, (b) LW-CNN, (c) Multi-CNN

fully meet the real-time requirements of fault diagnosis compared with the actual sampling time of 0.64 s.

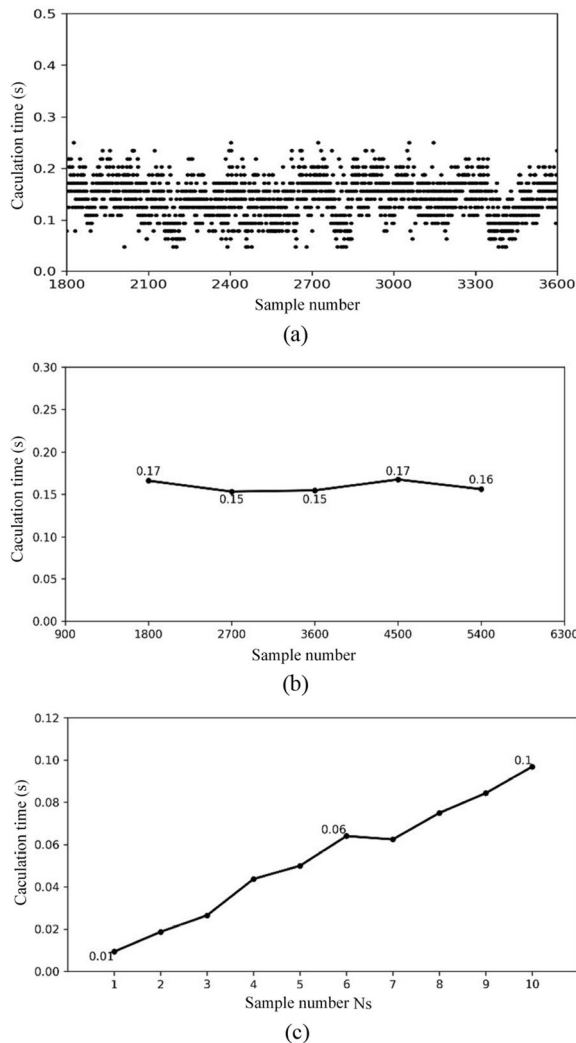
As it was considered unlikely that a high performance GPU would be provided in a real application, the running time of the algorithm was measured on the CPU (4-core i5-1135G7) using method (c), as shown in Figure 25. The average processing time per data set on the CPU is 0.05 s, which still meets the algorithm's real-time requirements. Of course, the running time of the algorithm is also affected by other factors such as programming language, program structure, computer hardware, and computer environment. Hence, the measured time varies from time to time. Even if the above variations exist, it does not change our conclusion that the algorithm is real-time.

## 5 Conclusions

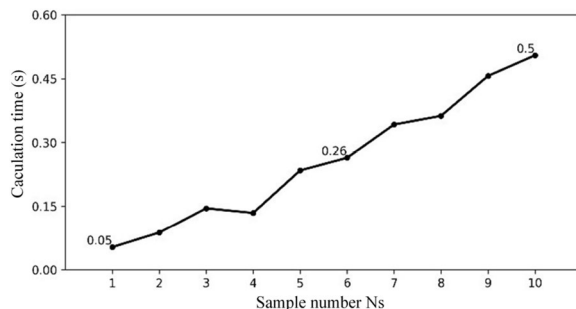
Co-frequency faults account for a large proportion of the actual factory production process. In order to solve the problem that the characteristics of the co-frequency vibration signal are not obvious and the diagnosis is difficult, a real-time fault extraction method based on

LW-CNN is proposed in this paper. The following conclusions verified by experiments are as follows:

- (1) A complete fault diagnosis process is proposed for the three most common co-frequency faults in rotor faults: Imbalance, misalignment, and looseness, and correlation coefficients are proposed for the data enhancement part of it to quantitatively assess the correlation before and after data enhancement.
- (2) Using CNN and SVM for fault diagnosis with different data volumes, the classification performance of CNN is better than SVM. The average classification accuracy of LW-CNN is above 95%, and using LW-CNN can also reduce the training parameters and speed up the training speed.
- (3) The diagnosis time of the LW-CNN-based real-time fault diagnosis method is tested using the measured fault diagnosis time method, and the results show that the time is 0.01 s on the high-performance GPU and 0.05 s on the CPU, both of which can meet the real-time requirements.



**Figure 24** Running time of the measurement method using different methods of increasing sample points (methods (a), (b), (c))



**Figure 25** Algorithm runtime on CPU

This is a public attempt to use data data-driven method to diagnose rotor co-frequency faults in real time. The deep learning method used in this paper is a relatively

simple algorithm, and many more cutting-edge and esoteric algorithms are yet to be further applied in this direction to finally achieve high-performance, fast, and accurate diagnosis of rotor co-frequency faults.

**Authors’ Contributions**

XP and ZJ were in charge of the whole trial; XZ and XP wrote the manuscript and analysed the data; GB and JN assisted with sampling and laboratory analyses. All authors read and approved the final manuscript.

**Funding**

Supported by National Natural Science Foundation of China (Grant Nos. 51875031, 52242507), Beijing Municipal Natural Science Foundation of China (Grant No. 3212010) and Beijing Municipal Youth Backbone Personal Project of China (Grant No. 2017000020124 G018).

**Availability of Data And Materials**

The datasets supporting the conclusions of this article are included within the article.

**Declarations**

**Ethics Approval and Consent to participate**

Not applicable

**Consent for Publication**

Not applicable

**Competing Interests**

The authors declare no competing financial interests.

Received: 4 January 2022 Revised: 24 March 2024 Accepted: 2 April 2024  
Published online: 10 May 2024

**References**

- [1] J J Gao. Artificial self-recovery and machinery self-recovery regulation system. *Chinese Journal of Mechanical Engineering*, 2018: 83-94.
- [2] R Liu, B Yang, E Zio, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 2018, 108: 33-47.
- [3] X Pan, H T He, H Q Wu. Optimal design of novel electromagnetic-ring active balancing actuator with radial excitation. *Chinese Journal of Mechanical Engineering*, 2021, 34: 9.
- [4] G Lai, J Liu, F Zeng. Application of multi-objective genetic algorithm in ship shafting alignment optimization. *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, December 9-10, 2017: 275-278.
- [5] X Ji, Y Ren, H Tang, et al. DSMT-based three-layer method using multi-classifier to detect faults in hydraulic systems. *Mechanical Systems and Signal Processing*, 2021, 153: 107513.
- [6] W M Wang, J J Gao, Z N Jiang. Theoretical and experimental study on the fault self-recovering system of high speed turbo machinery. *Journal of Vibration and Shock*, 2006, 25(4): 30-33.
- [7] M Lal. Modeling and estimation of speed dependent bearing and coupling misalignment faults in a turbine generator system. *Mechanical Systems and Signal Processing*, 2021, 151: 107365.
- [8] A K Jalan, A R Mohanty. Model based fault diagnosis of a rotor-bearing system for misalignment and unbalance under steady-state condition. *Journal of Sound and Vibration*, 2009, 327(3-5): 604-622.
- [9] N E Huang, Z Shen, S R Long, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 1998, 454(1917): 903-995.

- [10] Z Wu, N E Huang. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 2011, 1(1): 1-41.
- [11] S Zhou, S Qian, W Chang, et al. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. *Sensors (Basel)*, 2018, 18(6): 1934.
- [12] H C Wu, J Zhou, C H Xie, et al. Two-dimensional time series sample entropy algorithm: Applications to rotor axis orbit feature identification. *Mechanical Systems and Signal Processing*, 2021, 147: 107123.
- [13] A Tabrizi, L Garibaldi, A Fasana, et al. Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. *Meccanica*, 2015, 50(3): 865-874.
- [14] M Pei, H Li, H Yu. Degradation state identification for hydraulic pumps based on multi-scale ternary dynamic analysis, NSGA-II and SVM. *Measurement Science Review*, 2021, 21(3): 82-92.
- [15] B Ahmad, B K Mishra, M Ghufra, et al. Intelligent predictive maintenance model for rolling components of a machine based on speed and vibration. *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, Republic of Korea, Jeju Island, April 20-23, 2021: 459-464.
- [16] X B Wang, L Luo, L Tang, et al. Automatic representation and detection of fault bearings in in-wheel motors under variable load conditions. *Advanced Engineering Informatics*, 2021, 49: 101321.
- [17] C Li, K Yang, H Tang, et al. Fault diagnosis for rolling bearings of a freight train under limited fault data: few-shot learning method. *Journal of Transportation Engineering, Part A: Systems*, 2021, 147(8): 04021041.
- [18] M Cerrada, R V Sánchez, C Li, et al. A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 2018, 99(15): 169-196.
- [19] B Liu, S Riemenschneider, Y Xu. Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum. *Mechanical Systems and Signal Processing*, 2006, 20(3): 718-734.
- [20] J H Zhong, P K Wong, Z X Yang. Fault diagnosis of rotating machinery based on multiple probabilistic classifiers. *Mechanical Systems and Signal Processing*, 2018, 108: 99-114.
- [21] J D Martínez-Morales, E R Palacios-Hernández, D U Campos-Delgado. Multiple-fault diagnosis in induction motors through support vector machine classification at variable operating conditions. *Electrical Engineering*, 2016, 100: 59-73.
- [22] I Goodfellow, Y Bengio, A Courville. *Deep learning*. Massachusetts: MIT press, 2016.
- [23] A Krizhevsky, I Sutskever, G Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012, 25: 1097-1105.
- [24] K Simonyan, A Zisserman. Very deep convolutional networks for large-scale image recognition, 2014, arXiv: 1409.1556.
- [25] C Szegedy, W Liu, Y Jia, et al. Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Massachusetts, USA, June 7-12, 2015: 1-9.
- [26] K He, X Zhang, S Ren, et al. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, June 26-July 1, 2016: 770-778.
- [27] Z Yin, J Hou. Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. *Neurocomputing*, 2016, 174(22): 643-650.
- [28] A Widodo, B S Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems And Signal Processing*, 2007, 21(6): 2560-2574.

**Xin Pan** born in 1987, is currently an associate professor and master supervisor at *Beijing Key Laboratory of Health Monitoring and Self-Recovery for High-End Mechanical Equipment, Beijing University of Chemical Technology, China*. He received his PhD degree from *Beijing University of Chemical Technology, China*, in 2015. His research interests include vibration monitoring, fault diagnosis and active control.

**Xiancheng Zhang** born in 1997, is currently a graduate student at *Beijing University of Chemical Technology, China*. He received his

bachelor's degree in *Beijing University of Chemical Technology, China*, in 2020. His research interests include fault intelligent diagnosis.

**Zhinong Jiang** born in 1967, is currently a professor and doctoral supervisor at *Key Laboratory of Engine Health Monitoring-Control and Networking (Ministry of Education), Beijing University of Chemical Technology, China*. His research interests include vibration monitoring and fault diagnosis.

**Guangfu Bin** born in 1981, is currently a professor at *Health Maintenance for Mechanical Equipment Key Lab of Hunan Province-Hunan University of Science and Technology, China*. He received his PhD degree from *Beijing University of Chemical Technology, China*, in 2013. His research interests include rotating machinery dynamics and modal analysis, shafting dynamic balance, mechanical dynamic testing.