Real-time Intelligent Diagnosis of Co-Frequency Vibration Faults in Rotating Machinery Based on Lightweight-Convolutional Neural Networks

Xin Pan (✉ panxin@mail.buct.edu.cn)

Beijing University of Chemical Technology <https://orcid.org/0000-0002-1969-7368>

Xiancheng Zhang

Beijing University of Chemical Technology

Zhinong Jiang

Beijing University of Chemical Technology

Guangfu Bin

Hunan University of Science and Technology

Research Article

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

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1207913/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title page

Real-time Intelligent Diagnosis of Co-Frequency Vibration Faults in Rotating Machinery Based on Lightweight-Convolutional Neural Networks

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.

Tel: +86-010-64443098; E-mail: panxin@mail.buct.edu.cn

Xian-Cheng 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. E-mail: zhangxclife@163.com

Zhi-Nong 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.

Tel: +86-010-64446043; E-mail: jiangzhinong@263.net

Guang-Fu 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.

E-mail: abin811025@163.com

Corresponding author: Xin Pan E-mail: panxin@mail.buct.edu.cn

ORIGINAL ARTICLE

Real-time Intelligent Diagnosis of Co-Frequency Vibration Faults in Rotating Machinery Based on Lightweight-Convolutional Neural Networks

Xin Pan^{1,2} • Xian-Cheng Zhang¹ • Zhi-Nong Jiang^{1,2} • Guang-Fu Bin³

Received June xx, 201x; revised February xx, 201x; accepted March xx, 201x

© Chinese Mechanical Engineering Society and Springer-Verlag Berlin Heidelberg 2017

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 difficult to identify, and existing intelligent methods require more hardware conditions and exclusive time-consuming. Therefore, lightweight-Convolutional Neural Networks (LW-CNN) algorithm is proposed in this paper to achieve real-time fault diagnosis purpose. For the sliding window data augmentation method, the important parameters are discussed and verified by simulated and experimental signals. 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 the process from data acquisition to fault diagnosis. It is verified by experiments that the LW-CNN and sliding window method are used with high accuracy and real-time performance.

Keywords: Co-frequency vibration • Real-time diagnosis • LW-CNN • Data augmentation

1 Introduction

✉ Xin Pan
panxin@mail.buct.edu.cn

Zhi-Nong Jiang
jiangzhinong@263.net

¹ Beijing Key Laboratory of Health Monitoring and Self-recovery for High-end Mechanical Equipment, Beijing University of Chemical Technology, Beijing 100029, P R China

² Key Laboratory of Engine Health Monitoring-Control and Networking (Ministry of Education), Beijing University of Chemical Technology, Beijing 100029, P R China

³ College of Mechanical Engineering, Hunan University of Science and Technology, Xiangtan, 411201, P R China

Vibration dominated by the rotational vibration component (referred to as the co-frequency vibration) is the most common form of vibration faults in rotating equipment. According to statistical data, co-frequency vibration accounts for about 70% of rotating equipment vibration failure format. When serious vibration fault occurs, if the vibration signal is measured in the co-frequency components dominate, it is usually done on site balancing or the rotor is removed to the factory. If the true cause of vibration fault isn't found, the vibration fault can't be eliminated after reassembly, which not only wastes precious production time. Even more, sometimes even produces new faults due to improper assembly, which complicates the problem. Meanwhile, with the development of intelligent equipment, one of the inevitable needs is self-diagnosis of the equipment [1,2]. Real-time fault diagnosis is the prerequisite for self-recovery regulation mechanisms such as active balancing and automatic balancing algorithms and shafting alignment algorithm to suppress and location excessive vibration [3,4].

The fault types with the co-frequency characteristics include unbalance, misalignment, shaft bending, and slight loosening faults, etc. In engineering scenarios, it is easy to ascribe the co-frequency vibration fault to unbalance, which makes the co-frequency vibration fault become simple to distinguish the fault state but difficult to distinguish the fault category. In terms of methods, fault diagnosis of rotating machinery can be divided into three categories, that is, vibration model, signal processing and intelligent algorithm[5]. The methods based on vibration models, such as using 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 different equipment vary greatly and

partly randomness. 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 takes a long time in loop iteration.

The intelligent algorithm has the characteristics of strong adaptability and high processing precision, and is combined with the signal processing method to make the intelligent algorithm more practical. Support vector machine (SVM) and convolutional neural network (CNN) are two intelligent algorithms commonly used in fault diagnosis. Linear SVM is a kind of binary classification model whose basic model is defined as a linear classifier with the largest interval in the feature space, which has the advantages of wide applicability and fast training. Pei et al. proposed a joint optimization of SVM and nondominated sorting genetic algorithm II for the target, and the performance of the jointly optimized classifier was 99% or even more[14]. Ahmad et al. suggested 2 new parameters for axis misalignment that occurs during overload, and used SVM to do fault diagnosis, which has high accuracy in practical use[15]. The classical CNN has been widely used in image processing and fault diagnosis field. Wang et al. diagnosed motor bearing vibration fault signals under variable loads[16] and Li et al. transferred a small samples of bearing fault diagnosis to a different but related task[17], both of which adopted CNN as the classifier. At present, gearbox and bearing fault diagnosis are the main fault diagnosis objects of rotating machinery [11, 18, 19], and these fault signal features are accompanied by shock signals.

Most of the fault diagnosis studies using deep learning methods are aimed at fault signals with impact characteristics, while the real-time diagnostic studies for co-frequency faults has not been reported at home and abroad. The co-frequency faults of rotating machinery, such as unbalance, misalignment and looseness and other faults, are often with single spectral characteristics, vibration without shock, and the vibration energy values are close to each other under the same working conditions, which are difficult to distinguish by common spectrograms. Misalignment fault, for example, as described in Section 2.1, the causes of the fault are diverse, but the misalignment fault is clear from the vibration mechanism[8]. Meanwhile, it is always more than one fault specifically to different scenarios of different rotor system

damping, stiffness, installation conditions, so the effect of spectrum graph is not obvious. The classification method uses machine learning and deep learning methods, which can extract the signal features of these co-frequency faults in a high-dimensional space or deep network. In engineering scenarios, most of the data in the collection process is normal data, but the fault data is relatively small, so rational data augmentation is also necessary. The original vibration time domain signal is intercepted by sliding window processing and can be directly used for classifier. In this paper, the significant parameters in sliding window data augmentation method are discussed and quantitatively analysed.

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 same-frequency faults and its evaluation method are proposed, as well as the method of data pre-processing. In Section 4, the performance of real-time intelligent diagnosis method based on LW-CNN is verified with designed experiments. Finally, the conclusion is in Section 5.

2 Rotating machinery failure mechanism and intelligent algorithms

The common rotating equipment 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 to interact with fluid materials, motor rotor affected by electromagnetic field, and coupling. In short, the rotor-bearing sealing system is affected by centrifugal force, gravity, magnetic force, and interaction between rigid bodies, 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 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 poor alignment, bearing problems, improper clearance of bearing shell or other reasons. The co-frequency vibration excitation force mainly come from the radial rotation vector caused by rotor shaft system, bearing failure, interaction between rotor and stator or transmission parts, shell vibration and resonance, etc. The locations of rotor system unbalance, misalignment and looseness fault are shown in Figure 2.

Figure 1 Rotating machinery co-frequency vibration fault diagnosis

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

2.1 Comparison faults mechanism and actual faults signal

Unbalance: Typical fault of radial rotation vector caused by the rotor itself is unbalanced fault. When the rotor has unbalance fault, i.e., the center of mass and center of 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 ω force, the force as shown in Eq.(1), in which e is eccentricity.

$$F = me\omega^2 \quad (1)$$

Misalignment: Rotating machinery needs couplings to transmit forces and moments from the power end to the load end and between multiple rotors. The axis is often a certain misalignment, when the rotor is running under load.

Generally, the rotor is always to varying degrees of center misalignment and angle misalignment, coupling itself can overcome a certain degree of misalignment. However, if the misalignment exceeds a certain level, it will affect the performance of the bearing.

Figure 3 The trajectory of O'

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

Then, the trajectory of O' is expressed as follows.

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

Expression of exciting force.

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

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. In actual conditions, it is more cases of both misalignments.

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. Loosening fault, very small unbalance and misalignment can lead to great co-frequency vibration of the rotor support system. Loose rotor support parts make the connection stiffness and mechanical damping decrease, assuming that the gap is Δ , the mass converted to the rotor disc is recorded as C_0 , the rotor's equation of motion is as follow.

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

Where x is the disc mass displacement, 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 a non-linear system, the vibration response contains the co-frequency signal, and $2X$, $3X$ and other high harmonics and low harmonics, in which X is 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 not satisfactory. As shown in the Figure 4. In the spectrum diagram, the high amplitude frequency is basically concentrated in power frequency and $2X$, and the variation of vibration amplitude at different measuring points under the same fault and at the same measuring point under different faults is basically irregular.

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

2.2 Intelligent diagnosis algorithm

CNN and SVM are classical classification methods in fault diagnosis filed. Different convolutional structures, convolutional kernels number, convolutional kernels 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 to form different improved-SVM [11,21]. The following is a brief description of CNN and SVM principles and signal processing.

2.2.1 CNN and LW-CNN

CNN, a kind of neural network containing convolution computation, get 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 2-dimensional convolution is defined as follow.

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

For an 1 layer 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

$$z_{u,v}^l = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} y_{i+u, j+v}^{l-1} \cdot k_{rot, u-i, v-j}^l \cdot \delta(i, j) \quad (6)$$

$$\delta(i, j) = \begin{cases} 1, & 0 \leq i, j \leq n \\ 0, & \text{otherwise} \end{cases}$$

Using average pooling method, the weight of each convolution kernel is λ^{l+1} , and the output of the pooling layer with convolution add the bias is Eq.(7)

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

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

The pooling layer has 2 functions, one is to allow further feature extraction meanwhile, 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 number of parameters and better generalization performance of the model compared to previous neural networks.

CNN was widely used in computer vision and natural language recognition, also prevalent in the field of fault diagnosis in. 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 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, the shorter the training time, the shorter the prediction time of the trained network, so the lightweight neural network should be taken into consideration.

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. In the structure shown in Fig.4, the 5 hidden layers of CNN are, 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).

Figure 5 Signal processing and LW-CNN structure

2.2.2 SVM

Assume that the input space and feature space of SVM are two different spaces, where the input space is a Euclidean space or discrete set and the feature space is a Euclidean space or Hilbert space. The linear separable SVM and linear SVM assume that the elements of these two spaces correspond one-to-one and map the inputs in the input space to the feature vectors in 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 ϕ from the input space to the feature space H , the kernel function satisfies Eq.(8).

$$K(x, z) = \phi(x) \cdot \phi(z) \quad (8)$$

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).

$$f(x) = \omega^T \varphi(x) + b \quad (9)$$

Figure 6 Margin in SVM

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

$$\begin{aligned} \min_{\omega, b} \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i (\omega^T \varphi(x_i) + b) \geq 1, i = 1, 2, \dots, m \end{aligned} \quad (10)$$

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).

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

After solving α by the above equation, ω and b can be found subsequently. The solution leads to the hyperplane equation.

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

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 space 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 low number of parameters. With the development of machine learning, SVM as a common classification method has been used extensively, and for more detailed derivation of formulas in SVM in Ref. [28].

3 Real-time diagnosis method of co-frequency faults

3.1 The basic principle

As mentioned above, the actual same-frequency vibration faults do not have obvious characteristics in the 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 itself in deep learning methods is a black-box process. The same 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.

Figure 7 Co-frequency vibration fault diagnosis diagram

The traditional offline fault diagnosis often relies on specialized knowledge to extract signal or mechanism features, but these feature extractions often require an iterative process which is needed for each diagnosis process.

In contrast, real-time fault algorithms are more time consuming during the training process, 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 $S_1, S_2, \dots, S_n$, the data augmentation process is performed on the S_i . The rotor speed n and the sampling frequency (Fs) during the experiment, then the row and column size of each sample is N_f . The sample size N_f^2 is the nearest positive integer of sampling points per full turn at rotate speed n .

$$\left\lceil \frac{60F_s}{n} \right\rceil \leq N_f^2 \quad (13)$$

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 moving steps and after several iterations plenty of data can be generated 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.

$$Sig.i = A_i \sin(2\pi f_i + \varphi_i) + d_i + n_i \quad (14)$$

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 correlation analysis done for the signals under different RWW 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 RWW.

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

Table 2 Simulated signal settings

	f_i /Hz	A_i	φ_i	d_i	SNR/dB	Sampling points	Fs/Hz
Sig.1	30	2	0°	0	15	16384	25600
Sig.2	30	4	90°	2	10	16384	25600

Figure 8 Schematic of data augmentation by sliding window processing method

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

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

3.3 Real-time performance evaluation method

To verify whether the proposed algorithm is suitable for 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 proposed 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 original data points; b.) Add one sample feature size(L_2) each time based on a certain number from 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.

Figure 11 Schematic diagram of real-time data addition method

4 Experimental verification

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

Figure 12 Flowchart of diagnosis process

4.1 Experiment preparation

This experimental verification is mainly verified by the rotor test bench shown in Figure 13. The test bench is driven by a motor, and the first-order critical speed of the actual test bench rotor is about 2700 rpm. 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 shows more clearly the location of fault and sensor position. The experimental data collected by BH7000. The misalignment fault is simulated by using the same idea as the micrometre, i.e., the handle is screwed in to adjust the misalignment quantity; unbalance fault is to add a fixed mass of counterweight block to the 0 phase; loose fault is to loosen the screw on the bearing support side of the rear end, the test bench fault simulation is shown in Figure 14.

The actual faults often do not occur isolation, and the same type of failure also has different causes of failure, in most case, one fault is dominant and others coexist. Hence the dynamic balance and alignment of the test rig should be completed before fault data collection, regard as experimental data only if the test bench vibration value in the acceptable range. After doing each fault simulation test, no longer repeat the dynamic balance of the test bench and alignment experiments, but will remove the fault and continue to do the next 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 the experimental unbalance not too large lead to obvious distinction with other faults, the unbalance weight is deliberately applied to the side of the motor. The parameters of each fault in this experiment are shown in Table 3 and experiment is conducted at a constant speed of 1800 rpm (slightly below the first-order critical speed). The sampling frequency (F_s) sampling points (N_s) are 25.6 kHz and 16384, respectively. The amount of data collected is shown in Table 4.

Figure 13 Rotor failure test bench (a. Actual rotor test bench b. Schematic diagram of rotor test bench)

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

Table 3 Experimental fault type, magnitude setting

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

Table 4 Summary of fault data collected from experiments

Type	Data size
Misalignment	4*64*16384
Unbalance	4*47*16384
Looseness	4*43*16384
Normal	4*25*16384

Data format: No. measuring points * No. sampling periods * N_s

4.2 Experimental data processing

The data are sampled from different faults, different measurement points, the same time point, and their time domain waveforms are shown in Figure 15. 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 diagram is as illustrated in Section 2.1, therefore, based on time-domain and frequency spectral, which show that the four faults in the experiment are little different in practice.

Figure 15 Time domain waveforms of different fault types (a. unbalance b. misalignment c. looseness d. normal, CH1~CH4 are measurement points 1ha, 1va, 2ha, 2va ,respectively)

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

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 are used using 3/4 and 1/2 RWW, and the validation set is used using 1/3RWW 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 RWW are shown in Figure 17, Figure 18.

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

(a)

(b)

(c)

(d)

Figure 17 Autocorrelation coefficients for different types of faults at different RWW (a. unbalanced b. misaligned c. looseness d. normal)

(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)

From Figure 17 and 18, it can be concluded that the different fault data generated has good differentiation in 2-dimensional space. The similarity between different faults is in weak similarity or even in very weak, while the similarity between the same faults is in medium to high similarity. The relationships between unbalanced and loose faults in Figure 18(b) is high, mainly because they have the same fault magnitude and similar fault characteristics. Therefore, the data set after data augmentation can be used for further diagnosis.

(a)

4.3 Fault diagnosis verification

The experiments were performed on Python 3.7.10, in a 64bit Window environment with an 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 Table 6 and Table 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 itself to be lightweight, and the algorithm needs to be able to maintain a high correct rate even with little data. To address this practical problem, this experiment does the following in the fault diagnosis process.

- The difference between the two CNNs is that the Multi-CNN has more network layers and more parameters. In order to compare the different layers of CNN network, two CNN are designed respectively as LW-CNN and multi-CNN. Thereinto, multi-CNN has more hidden layers and more parameters.
- The algorithm does complete training, testing and validation at a gradual and uniform increase of the data volume from 25% to 100% of the total data volume.
- To avoid accidental error, the algorithms need to 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, as a common classification evaluation parameter in machine learning, is used to do a more comprehensive evaluation. For a specific binary classification problem, the samples are labelled with positive or negative labels (1, or 0), and the algorithm outputs a predicted label after training, compared with the actual label, the labelled results have four cases as shown in Figure 19, which are recorded as True Positive (TP), False Positive (FP), False Positive (FP), 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)~(18).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1score = \frac{2Precision \cdot Recall}{Precision + Recall} \quad (18)$$

		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

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 Figure 20 to Figure 23. In Table 8, the overall accuracy of SVM is above 85%, and 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 the increase of data volume. At 25% data volume, the accuracy of LW-CNN is above 95%, and most of which are at 99% level with excellent performance and the F1 score is above 0.98 on average. By analyzing Figure 20 to Figure 23, the fault diagnosis accuracy of CNN under the four data amounts is basically above 95%, which has excellent performance. However, the SVM, the imbalance and misalignment diagnosis accuracy are basically below 85% with 100% data volume or less. In terms of parameter number, the LW-CNN has 1.3 million, and the multi-layer CNN has 3.37million in comparison.

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

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 params	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)

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)

After processing the time-domain signal by the sliding window method, the 3 typical co-frequency vibration faults have a low correlation, which improves the discrimination of different vibration fault forms in the signal processing stage, and achieves the effect of data augmentation at the same time. According to the data analysis in Section 4.3.1, it is obvious that CNN performs better when the data volume is small, with better F1 score and better generalization performance. For CNN, although LW-CNN has one less 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 when using lightweight neural networks.

4.4 Real-time analysis of diagnostic methods

As described in the previous section 3.3. The feature size in this experiment is 900, the sampling time is 0.64s, and

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

In fact, 3 methods to increase samples, and the algorithm is repeated 10 times and regard the average time as the final operation time, finally, the results are 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 increase 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. For methods a and b, the data amount added is small. The time change is not visible due to the short time of the algorithm itself, which is regarded as 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 depict, the average processing time is about 0.01s, which can fully meet the real-time requirements of fault diagnosis compared with the actual sampling time of 0.64s.

Figure 24 The running time of the measurement method using different methods of increasing sample points (a, b, c are methods a, b, c respectively)

Considering the possibility that the actual application is equipped with a high-performance GPU is small, the running time of the algorithm on the CPU (4-core i5-1135G7) is measured using method c as shown in Figure 25. The average processing time per data set on the CPU is 0.05s, 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, so the measured time varies from time to time. Anyway, even if the above variations exist, it does not change our conclusion that the algorithm is real-time.

Figure 25 Algorithm runtime on CPU

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: unbalance, 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 to do fault diagnosis for different data sizes, it is found that the classification performance of CNN is better than SVM, where the average classification accuracy of LW-CNN is above 95%, and the use of LW-CNN can also reduce the training parameters and accelerate 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 0.01s on the high-performance GPU and 0.05s on the CPU, both of which can meet the real-time requirements.

This is a public attempt to use data data-driven method to do real-time diagnosis of rotor co-frequency faults. 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.

6 Declaration

Acknowledgements

Not applicable

Funding

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

Availability of data and materials

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

Authors' contributions

The author's contributions are as follows: Xin Pan: writing - review & editing, supervision, funding acquisition, in charge of the whole trial; Xian-Cheng Zhang: conceptualization, methodology, investigation, writing original draft; Zhi-Nong Jiang: investigation, in charge of the whole trial; Guang-Fu Bin: funding acquisition, investigation.

Competing interests

The authors declare no competing financial interests.

Consent for publication

Not applicable

Ethics approval and consent to participate

Not applicable

References

- [1] J J Gao. Artificial Self-recovery and Machinery Self-recovery Regulation System, *Chinese Journal of Mechanical Engineering*, 2018, 83-94. (in Chinese)
- [2] R Liu, B Yang, E Zio, X Chen. 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(1): 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)*, IEEE, 2017, 2: 275-278.
- [5] X Ji, Y Ren, H Tang, J Xiang. DSMT-based three-layer method using multi-classifier to detect faults in hydraulic systems, *Mechanical Systems and Signal Processing*, 2021, 153.
- [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, 025(004), 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, M C Wu, H H Shih, Q Zheng, N.-C Yen, C C Tung, H.H. Liu. 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(01): 1-41.
- [11] S Zhou, S Qian, W Chang, Y Xiao, Y Cheng. 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 W, J Z, C H X, et al. Two-dimensional time series sample entropy algorithm: Applications to rotor axis orbit feature identification[J]. *Mechanical Systems and Signal Processing*, 2021, 147: 107123.
- [13] A Tabrizi, L Garibaldi, A Fasana, S Marchesiello. 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 Ghufuran, Z Pervez, N Ramzan. 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)*, IEEE, 2021, 459-464.
 - [16] X B. Wang, L Luo, L Tang, Z X Yang. 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, P Wang, J Li, Q He. 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, F Pacheco, D Cabrera, J Valente de Oliveira, R E Vásquez. 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, 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, *Computer Science*, 2014.
 - [25] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, A Rabinovich. Going Deeper with Convolutions, *IEEE Computer Society*, 2014.
 - [26] K He, X Zhang, S Ren, J Sun. Deep Residual Learning for Image Recognition, *IEEE*, 2016.
 - [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.

Biographical notes

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. Tel: +86-010-64443098; E-mail: panxin@mail.buct.edu.cn

Xian-Cheng 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.

E-mail: zhangxclife@163.com

Zhi-Nong 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.

Tel: +86-010-64446043; E-mail: jiangzhinong@263.net

Guang-Fu 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.

E-mail: abin811025@163.com