Research Article

Defect Diagnosis of Gear-Shaft Bearing System Based on the OWF-TSCNN Composed of Wavelet Time-Frequency Map and FFT Spectrum 1

Peng Dai ^{1,2,3} JianPing Wang ^{1,2,3} Lulu Wu,^{1,2,3} ShuPing Yan,^{1,2,3} FengTao Wang ^{1,2,3} and Linkai Niu⁴

¹School of Mechanical Engineering, Anhui Polytechnic University, Wuhu 241000, China

²Anhui New R&D Institutions of Human-machine Interaction and Collaboration, Wuhu 241000, China

³Anhui Key Laboratory of Advance Numerical Control and Servo Technology, Wuhu 241000, China

⁴College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China

Correspondence should be addressed to JianPing Wang; tjqhd@163.com

Received 19 September 2021; Revised 9 November 2021; Accepted 22 December 2021; Published 4 March 2022

Academic Editor: Nibaldo Rodríguez

Copyright © 2022 Peng Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the defect diagnosis of the gear-shaft-bearing system with compound defects, the generated vibration signals are complicated. In addition, the information acquired by a single sensor is easily affected by uncertain factors, and low diagnostic accuracy is caused when traditional defect diagnosis methods are used, which cannot meet the high-precision diagnosis requirements. Therefore, a method is developed to identify the defect types and defect degrees of the gear-shaft-bearing system efficiently. In this method, the vibration signals are collected using multiple sensors, the dual-tree complex wavelet and the optimal weighting factor (OWF) methods are used for the data layer fusion, and the preprocessing is realized through wavelet transform and FFT. A learning model based on two-stream CNN composed of 1D-CNN and 2D-CNN is established, and the obtained wavelet time-frequency map and FFT spectrum are used as the input. Then, the trained features from the output of the connected layer are classified by the SVM. Compared with the OWF-1DCNN and OWF-2DCNN models, the time consumption of the OWF-TSCNN model is increased by 14.5%–26.6%, and the convergence speed of the network is decreased. However, its accuracy reaches 100% and 99.83% in the training set and test set, and the loss entropy and over-fitting rate are also greatly reduced. The feature extraction ability and generalization ability of the OWF-TSCNN model are increased, reaching 100% diagnosis accuracy on different defect types and defect degrees, which is more suitable for defect diagnosis of the gear-shaft-bearing system.

1. Introduction

The gear-shaft-bearing system is the most commonly used rotation transmission mechanism, and its health is directly related to the reliable operation of the entire equipment [1]. Due to the harsh and complex working environment, the localized defects of gears and bearings are easily caused [2]. Moreover, after there is a single-point localized defect on the system, the generated impulse force is transferred between the rigid bodies, which may affect other components, and the compound defects of the system will be caused [3]. In addition, with the continuous operation of the equipment, the defect area will be expanded under the long-term action of

the impulse force, which caused the defect degree to be deepened, and the performance and life of the system will be further reduced [4, 5]. Therefore, the research on the defect diagnosis of the gear-shaft-bearing system is of great significance for ensuring the safe and reliable operation of the equipment.

Since deep learning (DL) was proposed, it has been attracted the attention of many researchers and applied in computer vision [6], text processing [7], speech recognition [8], and other fields. Because the corresponding relationship between vibration signals and the operating state of the system can be determined by deep learning, it is suitable to process complex high-dimensional data, so it has been well

applied and developed in the field of defect diagnosis [9]. A defect diagnosis model of gear pairs based on deep sparse autoencoder was proposed by Qu [10], and in this model, the extracted diversity features were fused. And through experiments, it was concluded that this model had good performance in defect diagnosis of gearbox due to the strong generalization ability and robustness. The vibration time-domain signal of rolling bearing was transformed into a two-dimensional time-frequency image by Verstraete et al. [11], which was used as the input of the improved convolutional neural network (CNN), and the defect diagnosis with a high accuracy rate was realized by adaptively extracting features and performing feature classification. Zhang et al. [12] used the frequency spectrum of the vibration signal as the input of the convolutional neural network, and then, the defect types and defect degrees of the bearing were accurately identified. Pan et al. [13] built a model based on the time-domain improved algorithm in the second-generation wavelet transform. The resampling and normalization were combined to keep the signal samples consistent at different speeds, and defect types of bearing were diagnosed by this method. The time-domain information and frequency-domain information of the signal were combined by Zhao et al. [14] and used as the input of the gated loop unit. The average feature of the center offset was used as the additional input of the network for realizing the defect diagnosis and life prediction of the rotating machinery. An adaptive stacked convolutional neural network was proposed by Qian et al. [15]. In this model, an adaptive layer was used in the algorithm to overcome the translation error and the boundary problem caused by the traditional convolutional network, so the defect classification of bearing was realized under small data samples.

In the actual process of measuring the vibration signal, noise is an unavoidable interference, so the measured signal will inevitably interfere with different degrees of noise, which brings great difficulty to the feature extraction [16]. In order to reduce the interference of noise, many methods have been proposed by researchers. The time-frequency map of the original signal was used by Lu et al. [17] as the input of a 4-layer convolutional neural network, which realizes the defect classification of bearings in the range of signal-to-noise ratio of 10 dB to 50 dB. The residual network was applied by Zhang et al. [18] for defect diagnosis of bearings and realized defect diagnosis in the range of 0 dB to 8 dB signal-to-noise ratio. Qiao et al. [19] proposed an adaptive multiscale convolutional neural network for extracting multiscale features from the original signal and realized defect diagnosis in the range of -3 dB to 7 dB signal-to-noise ratio. The length loss was used by Liu et al. [20] for enhancing the adaptability of the nonlinear noise reduction autoencoder of the gated neural network to noise, and multisensor data were fused in the proposed algorithm for realizing defect diagnosis in the range of 1 dB to 10 dB signal-to-noise ratio.

For most of the research on defect diagnosis of rotating machinery, the main focus is on the gear pairs or bearings, and the gear-shaft-bearing system is not considered comprehensively. What is more, the defect diagnosis is ignored

by lots of researchers when there are compound defects in the system and the deepening of the defect degree. The information acquired by a single sensor is easily affected by uncertain factors, and low diagnostic accuracy is caused when traditional defect diagnosis methods are used, which cannot meet the high-precision diagnosis requirements. Therefore, a defect diagnosis method based on the OWF-TSCNN model of the gear-shaft-bearing system is proposed in this paper, and the compound defects and deepening of the defect degree of gears and bearings are considered. This method is mainly divided into 5 steps: signal acquisition, noise reduction, multisource information fusion, feature extraction of two-stream CNN model, and classification of SVM. The performance of the proposed method has been greatly improved, which can meet the requirements of high-precision defect diagnosis and improve the reference for defect diagnosis of gear-shaft-bearing systems in engineering practice.

2. Defect Diagnosis Model

The defect diagnosis model framework establishment of the gear-shaft-bearing system is shown in Figure 1. Multiple sensors are used to simultaneously collect the vibration information at different positions of the gearbox. The dual-tree complex wavelet transform is used for de-noising the original signal. Then, the optimal weighting factor method is used to fuse the signals for the data layer fusion. The wavelet time-frequency map and frequency spectrum of the signal can be obtained after the fused signals are processed by the wavelet transform and fast Fourier transform (FFT). Particularly, the wavelet time-frequency map of the signals after wavelet transformation is a 2D image group, and after the FFT, the frequency spectrum of the signal can be obtained, which belongs to the 1D signal. Then, the signal acquisition and preprocessing operation are completed.

The wavelet time-frequency map and frequency spectrum are used as the input of 2D-CNN and 1D-CNN, respectively, and thus, the two-stream CNN diagnostic model is composed. After the learning and training of 1D-CNN and 2D-CNN models, the extracted feature information is spliced in the fully connected layer. And then, the SVM is used to classify the feature from the output of the fully connected layer.

2.1. Dual-Tree Complex Wavelet Transform for De-Noising.

If the signal with noise is directly used as the input of the learning model, a learning model with a complex structure needs to be built to resist the interference of noise factors, which decreases the accuracy of the diagnosis system and takes a longer time for diagnosis [21, 22]. In order to accelerate the calculation of the deep learning model and improve the accuracy of the diagnostic system, the noise reduction process is separated from the feature information extraction process, so that the signal input to the learning model is more consistent with the measured object.

The traditional discrete wavelet packet transform is often used for signal noise reduction [23], but the phenomenon of

FIGURE 1: Defect diagnosis framework establishment of the gear-shaft-bearing system.

frequency aliasing will be generated when the signal is decomposed and reconstructed, and the feature extraction will be disturbed in the next step [24]. For this problem, the dual-tree complex wavelet transform (DT-CWT) is introduced, which not only retains the advantages of the complex wavelet transform but also has the advantages of translation invariance and complete reconstruction [25]. The process of decomposition and reconstruction of dual-tree complex wavelet transform is shown in Figure 2.

According to the theory of wavelet transform, one-dimensional complex wavelet transform can be expressed as

$$\varphi_t = \varphi_h(t) + i\varphi_g(t), \quad (1)$$

where $\varphi_h(t)$ and $\varphi_g(t)$ are two orthogonal or biorthogonal real wavelets and i is an imaginary unit.

The wavelet coefficients $dI_j^{Re}(n)$ and scale coefficients $cI_j^{Re}(n)$ of the real part tree are as follows:

$$\begin{aligned} dI_j^{Re}(n) &= 2^{j/2} \int_{-\infty}^{+\infty} x(t) \varphi_h(2^j t - n) dt, \quad j = 1, 2, 3, \dots, J, \\ cI_j^{Re}(n) &= 2^{j/2} \int_{-\infty}^{+\infty} x(t) \varphi_h(2^j t - n) dt, \end{aligned} \quad (2)$$

where $x(t)$ is the input time series signal.

The wavelet coefficients $dI_j^{Im}(n)$ and scale coefficients $cI_j^{Im}(n)$ of the imaginary tree are as follows:

$$\begin{aligned} dI_j^{Im}(n) &= 2^{j/2} \int_{-\infty}^{+\infty} x(t) \varphi_g(2^j t - n) dt, \quad j = 1, 2, 3, \dots, J, \\ cI_j^{Im}(n) &= 2^{j/2} \int_{-\infty}^{+\infty} x(t) \varphi_g(2^j t - n) dt. \end{aligned} \quad (3)$$

The dual-tree complex wavelet transform is composed of two wavelet transforms. Then, the wavelet coefficients $d_j^{\varphi}(n)$ and scale coefficients $c_j^{\varphi}(n)$ of the dual-tree complex wavelet are as follows:

$$\begin{aligned} d_j^{\varphi}(n) &= dI_j^{Re}(n) + dI_j^{Im}(n), \quad j = 1, 2, 3, \dots, J, \\ c_j^{\varphi}(n) &= cI_j^{Re}(n) + cI_j^{Im}(n). \end{aligned} \quad (4)$$

Then, the wavelet coefficients $d_j(t)$ and scale coefficients $c_j(t)$ of the dual-tree complex wavelet transform are reconstructed, as follows:

$$\begin{aligned} d_j(t) &= 2^{(j-1)/2} \left[\sum_{-\infty}^{+\infty} dI_j^{Re}(n) \varphi_h(2^j t - n) + \sum_{-\infty}^{+\infty} dI_j^{Im}(n) \varphi_g(2^j t - n) \right], \\ c_j(t) &= 2^{(j-1)/2} \left[\sum_{-\infty}^{+\infty} cI_j^{Re}(n) \varphi_h(2^j t - n) + \sum_{-\infty}^{+\infty} cI_j^{Im}(n) \varphi_g(2^j t - n) \right]. \end{aligned} \quad (5)$$

2.2. Multisource Information Fusion. The information collected by multiple sensors is fully utilized to expand the coverage of the system. Through reasonable use of these

FIGURE 2: Decomposition and reconstruction of the DT-CWT.

kinds of information, multiple redundant or complementary information is combined according to certain criteria for obtaining a description that is consistent with the tested object [26]. Moreover, the interference of environmental noise and other uncertain factors on the target signal can be effectively resisted by the use of multisensor information fusion to preprocess the vibration signal.

The optimal weighting factor method is used to fuse the signals in the data layer, as shown in Figure 3.

The measured values of each sensor are $x_1, \dots, x_i, \dots, x_n$, and the variances of the signal measured by n sensors are, respectively, assumed to be $\sigma_1, \dots, \sigma_i, \dots, \sigma_n$, which are independent of each other and are unbiased estimates of X . The weighting factors of each signal are $W_1, \dots, W_i, \dots, W_n$, and then, the fused signal X and the weighting factor satisfy the following equations:

$$\begin{aligned} X &= \sum_{i=1}^n W_i x_i, \\ \sum_{i=1}^n W_i &= 1. \end{aligned} \quad (6)$$

Because $x_1, \dots, x_i, \dots, x_n$ are independent of each other, the mean square error σ^2 of signal X is

$$\begin{aligned} \sigma^2 &= E \left[\sum_{i=1}^n W_i^2 (X - x_i)^2 + 2 \sum_{i=1, j=1}^n W_i W_j (X - x_i)(X - x_j) \right], \\ &= E \left[\sum_{i=1}^n W_i^2 (X - x_i)^2 \right] = \sum_{i=1}^n W_i^2 \sigma_i^2. \end{aligned} \quad (7)$$

It can be seen that the total mean square error σ^2 is a multivariate quadratic function of the weighting factors, so σ^2 must have a minimum value. According to the theory of finding the extreme value from the multivariate function, the weighting factor is obtained as $W_i = 1/[\sigma_i^2 \sum_{i=1}^n (1/\sigma_i^2)]$ when the total mean square error is the minimum, and the corresponding mean square error is $\sigma_{\min}^2 = 1/\sum_{i=1}^n (1/\sigma_i^2)$.

FIGURE 3: Multisource information fusion.

2.3. Construction of Two-Stream CNN. The wavelet time-frequency map of the vibration signal is used as the input of 2D-CNN [27], and the frequency spectrum of the vibration signal is used as the input of 1D-CNN [28], and thus, the two-stream CNN diagnostic model is composed. After a series of convolution and pooling operations of 1D-CNN and 2D-CNN models, the respective extracted features are spliced in the fully connected layer after the training. The time-domain features and frequency-domain features of the signal are simultaneously input to the learning model for feature extraction so that the deeper and more features in the signal can be mined.

2.3.1. Convolution Layer. For the 1D-CNN, the neural network is defined as [29]

$$(k \otimes f)[i] = \sum_{m=0}^W k[m] f[i+m], \quad (8)$$

where $\otimes$ represents the convolution operation, k and f represent the convolution kernel function and convolution input, and W represents the width of the convolution kernel. For the 2D-CNN, the definition in the neural network is as follows:

$$(k \otimes f)[i, j] = \sum_{m=0}^{W_1} \sum_{n=0}^{W_2} k[m, n] f[i+m, j+n], \quad (9)$$

where W_1 and W_2 represent the width of the convolution kernel in different dimensions.

2.3.2. Rectified Linear Unit. A simple piecewise function is used in the ReLU to form an overall nonlinear function, which is defined as follows [30]:

$$\text{ReLU}(x) = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases} \quad (10)$$

2.3.3. Pooling Layer. In the pooling operation, a region is replaced by the overall statistical characteristics of the region and its neighboring regions. The maximum pooling is used in this paper, and it is defined as

$$y_i = \max(r_j^{n \times n} u(n, n)), \quad (11)$$

where y_i is the output value of area j ; $r_j^{n \times n}$ represents the j -th area of $n \times n$ size; and $u(n, n)$ represents a window function, and the window function is taken as 1.

2.3.4. Dropout Layer. The neurons are closed with a certain probability during the neural network training process in the dropout layer, which can prevent the mutual adaptation between neurons. For each subnetwork, its probability distribution $p(y|x, u)$ can be defined by the mask u , so the integration of subnetworks can be defined as follows:

$$\sum_u p(u) p(y|x, u), \quad (12)$$

where $p(u)$ represents the probability of sampling mask u during training. In this paper, the neurons in the network are set to zero with a probability of 0.5.

2.3.5. SVM. The output of the fully connected layer is extracted to construct the SVM classification. The Gaussian function is selected as the kernel function of the SVM, and the decision function $f(x)$ model is [31]

$$f(x) = \text{sgn} \left[\sum_{n=1}^e y_n \lambda_n K(x, x_n) + b \right], \quad (13)$$

$$K(x, x_n) = \exp \left(-\frac{x - x_n^2}{2\sigma^2} \right),$$

where $K(x, x_n)$ is the basis kernel function, which represents the inner product of the test sample x and the n -th training sample x_n mapped to the feature space; λ_n is the Lagrangian factor of the n -th training sample; and σ is the bandwidth of the kernel function.

2.4. Network Structure Parameters. As shown in Table 1, the 1D-CNN network model is set to 8 layers. There are an input layer, 2 convolutional layers, 2 pooling layers, and a flatten layer. Similarly, the network structure of the 2D-CNN is set to 8 layers, including 2 convolutional layers and 2 pooling layers, as shown in Table 2. The BN layer is added before the activation function to normalize the input, which can solve the influence of the offset and increase of the input data.

After the two-stream CNN model is composed, the fully connected layer (FC) is added after the flatten layer combines the trained features of the 1D-CNN and 2D-CNN models, and then, the outputs of 1D-CNN and 2D-CNN are stretched and spliced. In the FC layer, FC-1 is connected to the output of the flatten-1D layer in the 1D-CNN model and the flatten-2d layer in the 2D-CNN model through 120 neurons, and all local features are merged into global features, which produces 120 feature maps of size 1×1 . The FC-2 layer is connected to the output of the FC-1 layer through 84 neurons, and the output layer is connected to the output of the FC-2 layer through 10 neurons. The SVM is used to divide the data samples into 10 categories, which correspond to the 10 defect types in the gear-shaft-bearing system. The settings for the FC layer are shown in Table 3.

3. Case Analysis of the Gear-Shaft-Bearing System

3.1. Experimental Equipment. The parts with localized defects of gears and bearings are shown in Figure 4. There are rectangular spalling defects on the high-speed gear, and the cross-sectional sizes of the defects are $1 \text{ mm} \times 7 \text{ mm}$, $2 \text{ mm} \times 13 \text{ mm}$ and $4 \text{ mm} \times 17 \text{ mm}$, and the depth of the defects is all 0.5 mm. There are crack defects at the outer raceway and roller of cylindrical bearing, and the cross-sectional sizes of the defects are $0.2 \text{ mm} \times 0.2 \text{ mm}$, $0.5 \text{ mm} \times 0.5 \text{ mm}$, and $0.8 \times 0.8 \text{ mm}$. And there are fracture defects on the cage, and the distance of the fracture section is 0.2 mm, 0.5 mm, and 0.8 mm.

The test bench of the gear-shaft-bearing system is shown in Figure 5. The input speed of the motor is controlled to 1800 rpm through the control system, and the load of the magnetic powder brake to the system is 25 N•m. The main parameters of the gear pair and cylindrical roller bearing types are shown in Table 4.

3.2. Multisource Information Composition. The distribution and serial number of bearings in the gear-shaft-bearing system are given in Figure 6. The four acceleration sensors shown in Figure 5 are used to collect the vibration signal of the gearbox, and the corresponding measuring points are the bearing end covers. After a set of experiments, 4 sets of different vibration signals can be obtained, so that the multisource signals of the gear-shaft-bearing system are composed, as shown in Table 5.

3.3. Experiment Grouping. 10 sets of experiments are completed by the use of the parts with a localized defect in Figure 4. In the experiment, the defects are all set on bearing 1, while bearing 2, bearing 3, and bearing 4 are all healthy. The defect types and defect degrees of the gear-shaft-bearing system in each set of experiments are shown in Table 6. In the table, 'T' indicates that the part is healthy, and 'F' indicates that there is a localized defect on the part, and the defect degrees are indicated by the values in brackets.

TABLE 1: 1D-CNN network parameter settings.

Serial number	Layer	Convolution kernels/step	Convolution kernels	Output size	Activation function
1	Input layer			1×433	
2	Conv-1D-1	$1 \times 5, 1$	6	$6@1 \times 429$	ReLU
3	Pooling-1D-1	$1 \times 3, 3$	6	$6@1 \times 143$	
4	BN-1D-1			$6@1 \times 143$	
5	Conv-1D-2	$1 \times 3, 1$	16	$16@1 \times 141$	ReLU
6	Pooling-1D-2	$1 \times 3, 3$	16	$16@1 \times 47$	
7	BN-1D-2			$16@1 \times 47$	
8	Flatten-1D			1×752	

TABLE 2: 2D-CNN network parameter settings.

Serial number	Layer	Convolution kernels/step	Convolution kernels	Output size	Activation function
1	Input layer			$64 \times 64 \times 3$	
2	Conv-2D-1	$5 \times 5, 1$	6	$6@60 \times 60$	ReLU
3	Pooling-2D-1	$2 \times 2, 2$	6	$6@30 \times 30$	
4	BN-2D-1			$6@30 \times 30$	
5	Conv-2D-2	$5 \times 5, 1$	16	$16@26 \times 26$	ReLU
6	Pooling-2D-2	$2 \times 2, 2$	16	$16@13 \times 13$	
7	BN-2D-2			$16@13 \times 13$	
8	Flatten-2D			1×1024	

TABLE 3: Fully connected layer.

Layer	Step size	Output size	Activation function
Convergent layer		1776	
FC-1	120	120×1	ReLU
FC-2	84	84×1	ReLU
Output layer	10	10×1	

4. Diagnostic Results

The rotating frequency of the shaft is 30 Hz, and the sampling frequency of the acceleration sensor is set to 15 kHz. 1024 data points are regarded as a sample, which corresponds to 2 rotation cycles of high-speed gear. The gear-shaft-bearing system is run for 65 seconds, and a total of about 1,000 samples are generated. 80% of the 1000 samples are randomly selected as the training set, and the rest are used as the test set. The corresponding measuring point is the bearing end cap, and the vibration acceleration signal in the y direction is collected, and this direction is close to the direction of the meshing line of the gear pair.

4.1. Multisource Vibration Signal. The acceleration signal after fusion based on the OWF is shown in Figure 7. It can be seen that the vibration signal generated by the gear-shaft-bearing system with different defect degrees and different defect types is very complicated, and the true defect types and defect degree are difficult to judge. However, the reliability of the acquisition system is enhanced by the redundant data between sensors. When there is a failure or error on one sensor or multiple sensors, the system can still continue to work. The advantages of common or joint operation of multiple sensors are promoted to improve the effectiveness of the system, which expands the coverage of time and space and improves the

reliability of the system. What's more, the interference of noise and other unfavorable factors on the measured object can be resisted effectively.

4.2. Accuracy and Loss Entropy. After the signals are fused, the fused signals are used as the input of the learning model. Four diagnosis models are studied for comparison, and the composition of the 4 diagnosis models is shown in Table 7. The learning rate of the learning model is set to 0.001. The results of the accuracy curve and loss entropy curve are shown in Figure 8.

As shown in Figure 8, the OWF-1DCNN, OWF-2DCNN, VCR-TSCNN, and OWF-TSCNN model can completely converge after 100 iterations. The convergence speed of the OWF-1DCNN model is the smallest, while the OWF-2DCNN model has the fastest convergence speed in the training set and test set before 40 iterations. There is more feature information on the wavelet time-frequency map, which includes not only the frequency-domain information of the signal but also time-domain information, and the increase in the amount of feature information plays a role in sample enhancement. In contrast, the input of the OWF-1DCNN model is the FFT spectrum, which only contains the frequency-domain information of the signal, and the amount of feature information is relatively less.

After the two-stream CNN learning model is composed, the wavelet time-frequency map and the FFT spectrum of the signal need to be trained at the same time, so the amount of calculation is increased, and the convergence speed is affected. Compared with the 1D-CNN model, more feature information is input to the two-stream CNN model, which makes the convergence speed of the two-stream CNN model faster. However, the computational load is greatly increased due to the increased amount of calculation, which causes the convergence speed of the

FIGURE 4: Parts with localized defects.

FIGURE 5: Test bench of the gear-shaft-bearing system.

TABLE 4: Main parameters of the gear-shaft-bearing system.

Parameter (unit)	Driving gear	Driven gear	Bearings	Types
Modulus m /mm	2.5	2.5	Bearing 1	N 304 ECP
Number of teeth z	23	81	Bearing 2	N 304 ECP
Pressure angle $\alpha/(^{\circ})$	20	20	Bearing 3	N 207 ECP
Tooth width L /mm	26	26	Bearing 4	N 207 ECP
Young's modulus E /GPa	206	206	—	—

FIGURE 6: Distribution and serial number of bearings.

TABLE 5: Multisource information of the gear-rotor-bearing system.

Sensor	Signal	Unit
Sensor 1	y_1 (plumb direction at end cover of bearing 1)	m/s^2
Sensor 2	y_2 (plumb direction at end cover of bearing 2)	m/s^2
Sensor 3	y_3 (plumb direction at end cover of bearing 3)	m/s^2
Sensor 4	y_4 (plumb direction at end cover of bearing 4)	m/s^2

TABLE 6: Experiment grouping.

Parts	Driving gear	Outer raceway	Bearing 1	Cage	Label
			Roller		
Experiments					
Experiment 1	T	T	T	T	1
Experiment 2	F (1 mm $\times$ 7 mm)	F (0.2 mm $\times$ 0.2 mm)	T	T	2
Experiment 3	F (1 mm $\times$ 7 mm)	T	F (0.2 mm $\times$ 0.2 mm)	T	3
Experiment 4	F (1 mm $\times$ 7 mm)	T	T	F (0.2 mm)	4
Experiment 5	F (2 mm $\times$ 13 mm)	F (0.5 mm $\times$ 0.5 mm)	T	T	5
Experiment 6	F (2 mm $\times$ 13 mm)	T	F (0.5 mm $\times$ 0.5 mm)	T	6
Experiment 7	F (2 mm $\times$ 13 mm)	T	T	F (0.5 mm)	7
Experiment 8	F (4 mm $\times$ 17 mm)	F (0.8 mm $\times$ 0.8 mm)	T	T	8
Experiment 9	F (4 mm $\times$ 17 mm)	T	F (0.8 mm $\times$ 0.8 mm)	T	9
Experiment 10	F (4 mm $\times$ 17 mm)	T	T	F (0.5 mm)	10

two-stream CNN model to be lower than the 2D-CNN model. The convergence speed of the established two-stream CNN model is between that of the 1D-CNN and 1D-CNN models, and after the models are converged, the accuracy and loss entropy of the two-stream CNN model are better than the other two.

When different multisource information fusion methods are used, and the fused signals are different, which causes the difficulty of extracting feature information in the model to be different, the convergence speed of the two-stream CNN model is affected. When the OWF method is used as the preprocessing method of vibration signal in the model

FIGURE 7: Vibration signals of the system with localized defects.

TABLE 7: Composition of the diagnostic model.

Diagnosis model	Multisource information fusion	Learning model	Input to the learning model	Symbol
1	Optimal weighting factor	1D-CNN	FFT spectrum	OWF-1DCNN
2	Optimal weighting factor	2D-CNN	Wavelet time-frequency map	OWF-2DCNN
3	Variance contribution rate [32]	Two-stream CNN	FFT spectrum (1D-CNN)	VCR-TSCNN
4	Optimal weighting factor	Two-stream CNN	Wavelet time-frequency map (2D-CNN)	OWF-TSCNN
			FFT spectrum (1D-CNN)	
			Wavelet time-frequency map (2D-CNN)	

established in this paper, the convergence speed of the OWF-TSCNN model is obviously higher than the VCR-TSCNN model.

4.3. Comparison of Different Diagnostic Models. In order to confirm the high efficiency of the diagnosis model established in this paper, the results from different deep learning models after the models are converged are compared, as shown in Table 8.

As shown in Table 8, at the same input signal and learning rate, although the four learning models can all be converged, the accuracy on the training set and the test set is uneven. The training time of OWF-1DCNN and OWF-2DCNN diagnosis model is relatively short, which is 218 s and 234 s, respectively. The two-dimensional wavelet time-frequency map of the vibration signal is trained in the OWF-2DCNN model, the feature information is more abundant than the one-dimensional FFT spectrum, and more time needs to be consumed in each iteration, which causes the total training time to increase by 6.0% compared with the

OWF-1DCNN model. However, its accuracy in the training set and test set is improved, and the loss entropy and over-fitting rate are reduced.

After the two-stream CNN diagnostic model is composed, the training of 1D-CNN and 2D-CNN is working at the same time, and the total calculation amount of the network model is the sum of the two, resulting in an increase in training time, which is about 14.5%–26.6%. The complexity of the two-stream CNN network is increased, which consumes more time for model training. However, the results after training are significantly improved. Its accuracy in the training set is 100%, the accuracy in the test set is more than 99.58%, and the loss entropy and over-fitting rate are greatly reduced.

The training time of the OWF-TSCNN model is reduced by 2.9% compared with the VCR-TSCNN diagnostic model. The learning models used by the two diagnostic models are the same, and the difference in training time is mainly due to the increased complexity of the algorithm structure of the VCR preprocessing method compared to the OWF.

Generally, when different multisensor information fusion methods are used as the way to process the original

FIGURE 8: Accuracy and loss entropy: (a) accuracy of the training set; (b) accuracy of the test set; (c) loss entropy of the training set; (d) loss entropy of the test set.

TABLE 8: Comparison of different diagnostic models.

Learning model	Diagnostic results				
	Accuracy of training set (%)	Accuracy of test set (%)	Loss entropy	Over-fitting rate	Training time (s)
OWF-1DCNN	99.65	96.83	0.0014	1.0801	218
OWF-2DCNN	99.86	97.41	0.00043	1.0453	234
VCR-TSCNN	100	99.58	0.00015	1.0022	276
OWF-TSCNN	100	99.93	0.00016	1.0020	268

vibration signal of the gearbox, the convergence speed and training time of the learning model will be affected. However, after the model is converged, the performance difference of the entire diagnostic model mainly depends on the network structure and performance of the learning model.

4.4. Feature Visualization. The results of the feature extraction from the four diagnosis models are visually analyzed for comparing the feature extraction capabilities, and the t-SNE algorithm [33] is used to nonlinearly map high-dimensional data to low-dimensional space and achieve feature dimensionality reduction. As shown in Figure 9, the feature

information is effectively extracted by the four learning models, and the signal points are gathered from the original chaotic and disordered state. However, it is obvious that after the use of OWF-1DCNN and OWF-2DCNN models, the distance between the signal points is large, and some signal points are still in a divergent state. This shows that there are insufficient capabilities to extract feature information from the signal in the OWF-1DCNN and OWF-2DCNN models, and the high accurate diagnosis requirements cannot be met.

The diagnosis results of the VCR-TSCNN and OWF-TSCNN model are shown in Figures 9(c) and 9(d). It can be seen that the distances of the same types of signal points are relatively small, and the boundaries of the large-class spacing

FIGURE 9: Feature visualization: (a) OWF-1DCNN; (b) OWF-2DCNN; (c) VCR-TSCNN; (d) OWF-TSCNN.

are obvious. This shows that the feature extraction capability of the two-stream CNN composed of 1D-CNN and 2D-CNN has been improved, and the deep-level feature information can be mined again by using the model.

When different multisource information fusion methods are used, the two-stream CNN model has sufficient ability to extract features from different signals. It can also be seen from the feature visualization map that the signal points generated by the VCR-TSCNN and OWF-TSCNN models are basically in a convergent state, which also shows the universal applicability of the two-stream CNN model to different signals. Of course, for different signals, the feature information extracted by two-stream CNN is different, which leads to changes in the position of signal points on the feature visualization map, but the convergence and divergence of signal points are mainly determined by the feature extraction ability of the two-stream CNN model.

4.5. Generalization Ability. The defect classification accuracy of the diagnosis models on the unknown data set is tested for judging the generalization ability of the established

diagnostic model. The test set data are input into the trained model, and the result is given in the form of a confusion matrix after 100 iterations, as shown in Figure 10. The predicted defect labels are shown on the horizontal axis, and the real defect labels are shown on the vertical axis. The accuracy of prediction is indicated by the value on the diagonal line, and the error probability of prediction is indicated by the value on the off-diagonal line.

From Figures 10(a) and 10(b), it can be seen that the defect types and defects degrees of the gear-shaft-bearing system can be diagnosed by OWF-1DCNN and OWF-2DCNN model. Among them, the performance of the OWF-1DCNN model is the worst. Although the training time is the shortest, its generalization ability is not ideal. It not only has a deviation in the recognition of the defect degrees but also has a larger error rate in the defect types. The training time of the OWF-2DCNN model is longer than that of the OWF-1DCNN model. However, the generalization ability is significantly better, and 100% accuracy is achieved in the identification of defect types. It is also found that the OWF-2DCNN model is not optimal for defect diagnosis of the gearbox, and the ability to identify the defect degrees needs

FIGURE 10: Confusion matrix of defect classification: (a) OWF-1DCNN; (b) OWF-2DCNN; (c) VCR-TSCNN; (d) OWF-TSCNN.

to be improved, especially in the diagnosis of vibration signals generated by the gearbox with localized defects on bearing rollers, and the maximum error rate is 4.4%. This may be due to the fact that the rollers are not running in a pure rolling manner in the bearing raceways, and there are also a large number of sliding motions. When there is a localized defect on the roller, the shock generated is not obvious, resulting in a high error rate of the diagnostic model.

The diagnosis results of the VCR-TSCNN and OWF-TSCNN model are shown in Figures 10(c) and 10(d), and it can be seen that under different defect degrees and defect types, it maintains 100% accuracy. The accuracy has been improved to meet the requirements of the modern machinery industry for defect diagnosis.

Due to the time-domain features and frequency-domain features of the signal being used at the same time, the advantages of the 1D-CNN and 2D-CNN models are integrated, and the convergence speed and training time of the two-stream CNN model are not lost too much. In addition,

the performance of the diagnostic model is greatly improved. After comprehensive consideration, it can be concluded that the OWF-TSCNN model is the most suitable for defect diagnosis of gear-shaft-bearing systems.

5. Conclusion

A defect diagnosis model of the gear-shaft-bearing system is established in this paper. In this model, multisource information fusion method and two-stream CNN learning model are used, and then, a signal processing route is formulated. The dual-tree complex wavelet is used to denoise the signal, which separates the signal denoising process from the signal learning and training process. Multiple sensors are used to detect the running status of the system at the same time, and 1D-CNN and 2D-CNN models are combined to establish a two-stream CNN learning model. The vibration signals generated by the system with different defect types and different defect degrees are diagnosed, and the diagnosis results are as follows:

- (1) The reliability of the sensor system is enhanced by using the multisensor information fusion method, which enhances the sensor system's ability to resist interference from other unfavorable factors such as noise, and makes the collected vibration signals more reliable. When different multisource information fusion methods are used, the convergence speed of the two-stream CNN model is also different. The OWF method is used as the pre-processing method of vibration signal in the model established in this paper, and its convergence speed is obviously higher than the VCR method. The training time of the OWF-TSCNN model is reduced by 2.9% compared with the VCR-TSCNN diagnostic model.
- (2) The two-stream CNN learning model is built on the basis of the 1D-CNN and 2D-CNN model, and the time domain and frequency-domain features of the signal are used as the input at the same time for improving the accuracy of diagnosis. It is found that although the time consumption of the OWF-TSCNN model is increased by 14.5%–26.6%, the convergence speed of the network is decreased. However, its accuracy reaches 100% and 99.83% in the training set and test set, and the loss entropy and over-fitting rate are also greatly reduced. The feature extraction ability and generalization ability of the OWF-TSCNN model are increased, reaching 100% diagnosis accuracy on different defect types and defect degrees of the gear-shaft-bearing system.

Data Availability

The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare they have no conflicts of interest.

Authors' Contributions

Peng Dai, Jianping Wang, and Fengtao Wang contributed to the conception of the study; Peng Dai, Lulu Wu, and Shuping Yang contributed significantly to analysis and manuscript preparation; Peng Dai and Jianping Wang performed the data analyses and wrote the manuscript; Fengtao Wang and Linkai Niu helped to perform the analysis with constructive discussion.

Acknowledgments

This study was fully supported by a grant from the National Natural Science Foundation of China (No. 52005003) and Wuhu Science and Technology Projects (No. 2020yf53).

References

- [1] W. Pan, X. Li, L. Wang, and Z. Yang, "Nonlinear response analysis of gear-shaft-bearing system considering tooth contact temperature and random excitations," *Applied Mathematical Modelling*, vol. 68, pp. 113–136, 2019.
- [2] R. Sun, C. Song, C. Zhu, X. Yang, and X. Li, "Computational study of pitting defect influence on mesh stiffness for straight beveloid gear," *Engineering Failure Analysis*, vol. 119, Article ID 104971, 2021.
- [3] N. Sawalhi and R. B. Randall, "Simulating gear and bearing interactions in the presence of faults," *Mechanical Systems and Signal Processing*, vol. 8, no. 22, pp. 1924–1951, 2008.
- [4] H. Jiang and F. Liu, "Mesh stiffness modelling and dynamic simulation of helical gears with tooth crack propagation," *Meccanica*, vol. 55, no. 6, pp. 1215–1236, 2020.
- [5] V. Patil, V. Chouhan, and Y. Pandya, "Geometrical complexity and crack trajectory based fatigue life prediction for a spur gear having tooth root crack," *Engineering Failure Analysis*, vol. 105, pp. 444–465, 2019.
- [6] R. Ghazal, C. AnaMaria, and P. Pierre, "Transfer of learning from vision to touch: a hybrid deep convolutional neural network for visuo-tactile 3D object recognition," *Sensors*, vol. 1, no. 22, p. 113, 2020.
- [7] Y. Liang, H. Li, B. Guo et al., "Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification," *Information Sciences*, vol. 548, pp. 295–312, 2021.
- [8] A. Krug, M. Ebrahimzadeh, J. Alemann, J. Johannsmeier, and S. Stober, "Analyzing and visualizing deep neural networks for speech recognition with saliency-adjusted neuron activation profiles," *Electronics*, vol. 11, no. 10, 1350 pages, 2021.
- [9] L. Yu, Z. Wang, and Z. Duan, "Detecting gear surface defects using background-weakening method and convolutional neural network," *Journal of Sensors*, vol. 2019, Article ID 3140980, 13 pages, 2019.
- [10] Y. Qu, "Detection of pitting in gears using a deep Sparse autoencoder," *Applied Sciences*, vol. 5, no. 7, 2017.
- [11] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings," *Shock and Vibration*, vol. 2017, Article ID 5067651, 17 pages, 2017.
- [12] W. Zhang, F. Zhang, W. Chen, Y. Jiang, and D. Song, "Fault state recognition of rolling bearing based fully convolutional network," *Computing in Science & Engineering*, vol. 21, pp. 55–63, 2018.
- [13] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, "Lifting net: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification," *IEEE Transactions on Industrial Electronics*, vol. 6, no. 65, pp. 4973–4982, 2017.
- [14] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2017.
- [15] W. Qian, S. Li, J. Wang, Z. An, and X. Jiang, "An intelligent fault diagnosis framework for raw vibration signals: adaptive overlapping convolutional neural network," *Measurement Science and Technology*, vol. 29, no. 9, Article ID 095009, 2018.
- [16] Y. Liu, "Improved wavelet packet denoising algorithm using fuzzy threshold and correlation analysis for chaotic signals," *Transactions of the Institute of Measurement and Control*, vol. 6, no. 43, pp. 1394–1403, 2021.

- [17] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, no. 32, pp. 139–151, 2017.
- [18] W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Transactions*, vol. 95, pp. 295–305, 2018.
- [19] H. Qiao, T. Wang, P. Wang, L. Zhang, and M. Xu, "An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions," *IEEE Access*, vol. 7, pp. 118954–118964, 2019.
- [20] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018.
- [21] H. Liu, H. Wang, J. Bin et al., "Efficient noise reduction for the free induction decay signal from a proton precession magnetometer with time-frequency peak filtering," *Review of Scientific Instruments*, vol. 91, no. 4, Article ID 045101, 2020.
- [22] X. Song, "A bearing fault diagnosis model based on CNN with wide convolution kernels," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–16, 2021.
- [23] M. Schimmack and P. Mercorelli, "A wavelet packet tree denoising algorithm for images of atomic-force microscopy," *Asian Journal of Control*, vol. 4, no. 20, pp. 1367–1378, 2018.
- [24] D. Strömbergsson, P. Marklund, K. Berglund, and P. E. Larsson, "Bearing monitoring in the wind turbine drivetrain: a comparative study of the FFT and wavelet transforms," *Wind Energy*, vol. 23, no. 6, pp. 1381–1393, 2020.
- [25] P. Navdeep and S. Jain, "Design and implementation of a robust noise removal system in ECG signals using dual-tree complex wavelet transform," *Biomedical Signal Processing and Control*, vol. 63, Article ID 102212, 2021.
- [26] Z. Zhao, Q. Li, Z. Zhang et al., "Combining a parallel 2D CNN with a self-attention Dilated Residual Network for CTC-based discrete speech emotion recognition," *Neural Networks*, vol. 141, pp. 52–60, 2021.
- [27] T. Jin, C. Yan, C. Chen, and Z. Yang, "Light neural network with fewer parameters based on CNN for fault diagnosis of rotating machinery," *Measurement*, vol. 181, 2021.
- [28] X. Wang, D. Mao, and L. I. Xiaodong, "Bearing Fault Diagnosis Based on Vibro-Acoustic Data Fusion and 1D-CNN Network," *Measurement*, vol. 173, Article ID 108518, 2021.
- [29] L. Liu, R. Deng, and C. Lian-Kuan, "47-kbit/s RGB-LED-based optical camera communication based on 2D-CNN and XOR-based data loss compensation," *Optics Express*, vol. 23, no. 27, pp. 33840–33846, 2017.
- [30] B. Carlo, M. Malatesta Enrico, and Z. Riccardo, "Properties of the geometry of solutions and capacity of multilayer neural networks with rectified linear unit activations," *Physical Review Letters*, vol. 17, no. 123, 2019.
- [31] X. Di-Xiu, "CNN-SVM for microvascular morphological type recognition with data augmentation," *Journal of Medical and Biological Engineering*, vol. 6, no. 36, pp. 755–764, 2016.
- [32] Y. He, H. Li, and Y. Li, "Vibration signal fusion using improved empirical wavelet transform and variance contribution rate for weak fault detection of hydraulic pumps," *ISA Transactions*, vol. 107, pp. 385–401, 2020.
- [33] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *Journal of Machine Learning Research*, vol. 1, no. 15, pp. 3221–3245, 2014.