applied sciences

MDPI

Article

A Variable-Scale Attention Mechanism Guided Time-Frequency Feature Fusion Transfer Learning Method for Bearing Fault Diagnosis in an Annealing Kiln Roller System

Yu Xin [1,*], Kangqu Zhou [1], Songlin Liu [1] and Tianchuang Liu [2]

[1] College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China; ywg6789@cqut.edu.cn (K.Z.); 331904724@stu.cqut.edu.cn (S.L.)

[2] College of Intelligent Manufacturing and Automotive, Chongqing Vocational College of Transportation, Chongqing 402260, China; 17702301650@163.com

* Correspondence: yu_xin@cqut.edu.cn

Featured Application: The roller bearing and the through shaft bearing are the rotating and load-bearing components of the annealing kiln roller system, which operate at different locations and under different working conditions. Their health status is significant for maintaining the stable operation of the glass production line. In order to improve the efficiency of bearing condition monitoring, this paper proposes a variable-scale attention mechanism guided time-frequency feature fusion transfer learning method, which is used for bearing fault diagnoses at different installation locations in the annealing kiln roller system. It effectively achieves the intelligent diagnosis of roller bearing and through shaft bearing faults in the annealing kiln roller system.

Abstract: Effective real-time health condition monitoring of the roller table and through shaft bearings in the annealing kiln roller system of glass production lines is crucial for maintaining their operational safety and stability for the quality and production efficiency of glass products. However, the collected vibration signal of the roller bearing system is affected by the low rotating frequency and strong mechanical background noise, which shows the width impact interval and non-stationary multi-component characteristics. Moreover, the distribution characteristics of monitoring data and probability of fault occurrence of the roller bearing and through shaft bearing improve the difficulty of the fault diagnosis and condition monitoring of the annealing kiln roller system, as well as the reliance on professional experience and prior knowledge. Therefore, this paper proposes a variable-scale attention mechanism guided time-frequency feature fusion transfer learning method for a bearing fault diagnosis at different installation positions in an annealing kiln roller system. Firstly, the instinct time decomposition method and the Gini–Kurtosis composed index are used to decompose and reconstruct the signal for noise reduction, wavelet transform with the Morlet basic function is used to extract the time-frequency features, and histogram equalization is introduced to reform the time-frequency map for the blur and implicit time-frequency features. Secondly, a variable-scale attention mechanism guided time-frequency feature fusion framework is established to extract multiscale time-dependency features from the time-frequency representation for the distinguished fault diagnosis of roller table bearings. Then, for through shaft bearings, the vibration signal of the roller table bearing is used as the source domain and the signal of the through shaft bearing is used as the target domain, based on the feature fusion framework and the multi-kernel maximum mean differences metric function, and the transfer diagnosis method is proposed to reduce the distribution differences and extract the across-domain invariant feature to diagnose the through shaft bearing fault speed under different working conditions, using a small sample. Finally, the effectiveness of the proposed method is verified based on the vibration signal from the experimental platform and the roller bearing system of the glass production line. Results show that the proposed method can effectively diagnose roller table and through shaft bearings' fault information in the annealing kiln roller system.

Citation: Xin, Y.; Zhou, K.; Liu, S.; Liu, T. A Variable-Scale Attention Mechanism Guided Time-Frequency Feature Fusion Transfer Learning Method for Bearing Fault Diagnosis in an Annealing Kiln Roller System. *Appl. Sci.* **2024**, *14*, 3434. https://doi.org/10.3390/app14083434

Academic Editor: Qizhi Xu

Received: 7 March 2024
Revised: 15 April 2024
Accepted: 17 April 2024
Published: 18 April 2024

Keywords: intelligent fault diagnosis; roller system of annealing kiln; transfer learning; Inception

1. Introduction

The roller system of annealing kilns is a crucial component of the glass production line for carrying and hauling glass products. The diagram of the annealing kiln roller system is shown in Figure 1. The bearings of the annealing kiln roller system include two types: roller and through shaft bearings. Roller and through shaft bearings are SKF-1218K double-row self-aligning ball bearings, which are seated in bearing seats to ensure the stability and smooth operation of the annealing kiln roller system. The roller bearings in a production line are used for loading and pulling glass products at the speed of 16 r/min, and the through shaft bearing is used to transmit power at the speed of 12 r/min. As the core components of the roller system, the stable operation and health status of roller and through shaft bearings significantly affect the quality of glass products [1]. Therefore, the timely health condition monitoring and abnormal diagnostic of the roller table and through shaft bearings are essential for intelligently maintaining the operational stability and safety for the quality and production efficiency of the glass product line. However, the condition monitoring signal of the roller bearing system is affected by the low rotating frequency and strong mechanical background noise, which shows the width impact interval and non-stationary multi-component characteristics. Moreover, the difference in the rotating speed, load, and working condition between the roller bearing and through shaft bearing obviously increases the difficulty of the real-time intelligent fault diagnosis and condition monitoring of the annealing kiln roller system. And the professional knowledge and prior knowledge of the working condition of the roller bearing system also limit the efficiency of identifying abnormal states. Thus, exploring new vibration signal preprocessing and intelligent transfer diagnostic methods is essential for health monitoring and fault detecting for the annealing kiln roller system.

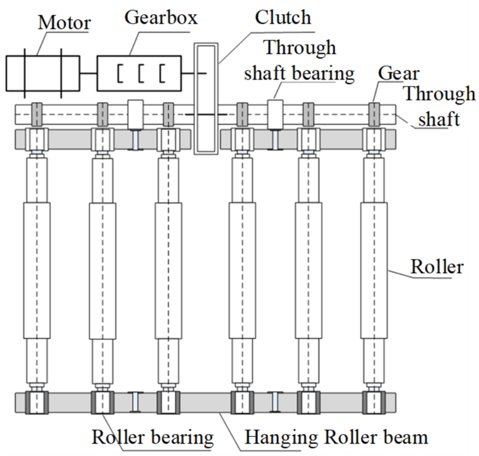


Figure 1. Diagram of the structure of the roller table system of the annealing kiln.

Affected by the working environment, the condition monitoring vibration signal of the bearing the annealing kiln roller system inevitably contains multiple non-stationary components and mechanical background noise. Hence, it is a prerequisite for improving the accuracy of fault diagnostics and the effectiveness of health monitoring to effectively decompose and filter out the noise component in the bearing vibration signal. Focused on denoising and filtering methods based on signal decomposition, classical fault diagnosis methods, such as ensemble empirical mode decomposition [2,3], local mean decomposition [4], and variational mode decomposition [5] methods, were wildly used to filter out background noise from nonlinear and non-stationary vibration signals. However, these methods cannot adaptively decompose the non-stationary vibration signal for separating meaningful information based on the time-scale, and endpoint effects, modal aliasing, and noise interference inevitably affect the decomposition results [6–8]. In order to effectively avoid these drawbacks, a novel method named intrinsic time scale decomposition (ITD) was introduced to decompose the vibration signal into several amplitude and frequency demodulation proper rotation single components [9,10]. On the basis of linear transform, the ITD method can decompose the signal into several proper rotation components, which are mono-components in nature and suitable for calculating the instantaneous frequency and amplitude. Moreover, ITD defines the instantaneous amplitude and frequency of signals on the basis of single wave analysis, thus overcoming the limitations of modal aliasing and noise interference and providing an approach to the decomposition of vibration signals into several mono-components for further demodulation analysis. Hence, considering the strong noise interference and low speed and width impact interval characteristics in the bearing vibration signal of the annealing kiln roller system, the ITD method is introduced to decompose the signal into several components with different frequencies, and the composed index with kurtosis [11] and the Gini index [12] are used to select the demodulation frequency component with the optimal pulse and sparse characteristics of fault information. The selected components are used to reconstruct the denoised signal for further identification. To improve the time-frequency representation ability of bearings under low-speed operating condition information, which have strong nonlinearity and non-stationary characteristics under complex working conditions, continuous wavelet transform (CWT) [13] is used to extract the time-frequency representation with different scales of denoised signals, and histogram equalization is used for processing the time-frequency image through non-linear straightening, reassigning image pixel values, and improving the image contrast.

Based on the denoised signal, reducing the reliance on professional experience and prior knowledge for bearing abnormal detection and intelligent diagnosis have significant advantages for improving the performance of intelligently maintaining and the condition monitoring of the glass production line. Deep learning methods have been widely utilized in the field of mechanical fault diagnosis [14]. Focused on the multi-scale and dependency characteristics, Li [15] combined Inception with an attention mechanism to extract the multi-scale features of bearing faults and found that the recognition effect using the Inception module is significantly better than that using a single CNN. Qiao [16] used CNN and long short-term memory (LSTM) to extract features under variable load and noise conditions and found that the recognition accuracy of the proposed method is higher than that using CNN or LSTM alone. Shi [17] proposed a novel deep neural network based on bidirectional-convolutional LSTM to determine the type, location, and direction of planetary gearbox faults by automatically and simultaneously extracting spatial and temporal features from both vibration and rotational speed measurements. However, these methods focused on the static data and their features and not so much on time-varying non-stationary data. The complex deep features in the time dependency, high-dimensional, and noisy real-world vibration signal cannot be adaptively learned with the shallow model, which constructed only several numbers of non-linear operations, and it could effectively model such complex data. Therefore, considering the multi-scale feature-extraction ability of Inception and the time-state reliance representation ability of LSTM, an variable scale attention mechanism-guided time-frequency feature fusion was constructed to further extract deep differential and abnormal features and improve the effectiveness of the bearing fault diagnosis.

As the crucial transmission part, the health condition of the through shaft bearings can also significantly affect the stability of the glass production line. But, affected by the rotating speed and working environment, the failure probability and vibration signal characteristics are apparently different from the roller table bearings. Therefore, when diagnosing roller bearing faults, how to accurately diagnose the through shaft bearing faults simultaneously is the key to effectively monitoring the health condition of the roller table system of the annealing kiln. As an excellent novel deep learning method, transfer learning aims to apply existing knowledge in evaluating and recognizing similar features between source and target domain data with different distribution characteristics [18,19]. As a nonparametric distance metric, MMD maps data to a reproducing Hilbert kernel space to evaluate the distribution characteristics and distance; therefore, the selection of kernel functions is crucial [20,21]. And this method not only performs well on data with few or insufficient samples but also has good identification performance based on cross-domain distribution difference data of the source and object domain. Therefore, based on the feature fusion network, considering the similarities and differences between annealing kiln roller and through shaft bearings, the transfer learning method is adopted to achieve domain invariant feature extraction and cross domain diagnosis, with the Multi Kernel Maximum Mean Difference (MK-MMD) as a distance indicator, roller bearing vibration data as the source domain, and through shaft bearing vibration data as the target domain, and an intelligent transfer diagnosis model for through shaft bearing faults is constructed.

To monitor and diagnose the real-time health condition of the bearing in the annealing kiln roller system in glass production lines, this paper proposes a time-scale data-driven vibration signal adaptive denoising and intelligent transfer diagnosis method based on a variable-scale attention mechanism-guided time-frequency feature fusion transfer learning method for the abnormal state detection and fault diagnosis of roller and through shaft bearings in the annealing kiln roller system. First, the strong mechanical background noise in the vibration signal is filtered using ITD and the Gini–kurtosis criteria, and the time-frequency representation of the denoised signal is obtained using CWT, and the time-frequency images are enhanced through histogram equalization. Then, according to Inception and LSTM, an intelligent diagnosis network based on a variable-scale attention mechanism-guided feature fusion framework is constructed to extract multi-scale features from the time-frequency representations and accurately identify the fault information of roller bearings. Furthermore, based on the intelligent diagnosis network, an intelligent transfer diagnosis model for through shaft bearings is constructed using MK-MMD. The vibration signals of the roller and through shaft bearings are used as the source and target domains, respectively, to accurately diagnose the operating condition of through shaft bearings. Finally, two kinds of vibration signal data are conducted to verify the proposed method for bearing faults in the annealing kiln roller bearing system. The results show that the proposed method can efficiently extract the bearing fault features and accurately identify the health condition of the bearings in the annealing kiln roller bearing system.

The rest of this paper is organized as follows. Section 2 reviews the concept of the Inception module, LSTM network, and MK-MMD method briefly. In Section 3, the proposed variable-scale attention mechanism-guided feature fusion module is introduced firstly in detail; then, the neural network framework and procedure of intelligent fault diagnosis and transfer diagnosis of the annealing kiln roller bearing system are descripted comprehensively. After that, the experiment data from the test rig are used to comprehensively verify the effectiveness of this proposed neural network framework and fault diagnosis method in Section 4. And the proposed method is introduced to diagnose measured data from the plant of the annealing kiln roller bearing system. The discussion and conclusion are presented in Section 5.

2. Theory Background and Methods

2.1. Inception Module

In GoogLetNet [22], the Inception module was proposed as the optimization module, with the main idea of using parallel convolutional kernels with different sizes in the convolutional layer to extract multi-scale features from input images and then fusing the extracted features to obtain a good image representation. As a milestone design, the Inception module emerged as a breakthrough for object detection and large-scale visual-recognition-related tasks. This module encapsulates multiple parallel kernel filters with different sizes to extract salient features from objects. The remarkable change in the Inception module is toward increasing the network width, enhancing the adaptability of the network to the scale, and improving the network performance.

In the Inception module, the convolutional kernels with different sizes can be perceived as various receptive fields that connect to enrich the information in each layer. These convolution kernels with different scales (such as 1×1, 3×3, and 5×5) are used to extract features from the input and then output the uniformly distributed features. Among them, the 1×1 convolutional kernel is used to extract outstanding features, the 3×3 convolutional kernels can extract feature information and reduce computational complexity while obtaining feature information on different scales, the 5×5 convolutional kernels are used to reduce the number of parameters to be trained and accelerate the training speed. In each parallel structure, batch normalization and global average pooling are introduced to improve the effectiveness of features, decrease the number of parameters, overcome overfitting, and reduce the computational cost. In the output set, the different features in each branch are concatenated. Consequently, variable-scale features can be extracted from the input data, and excellent performance can be achieved with low computational cost. The typical network structure of Inception V2 [23] is shown in Figure 2.

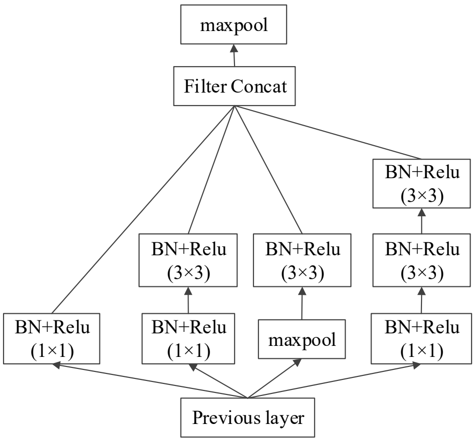


Figure 2. Structure of the Inception V2 module.

2.2. LSTM Network

The LSTM [24] network designs the forget, input, and output gates in the network structure to effectively alleviate gradient vanishing and explosion and has a strong ability to learn the dependency relationship of time-series data. The LSTM network structure is shown in Figure 3. The LSTM network receives the output and cell state from the previous step as the current input. The processing expression is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{5}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{6}$$

where σ represents the activation function, and W and b, respectively, represent the weights and biases of the corresponding processing operations. In the t-th update step, input gate, i, forget get f, output gate, O, and cell state, C are updated by input x, and the hidden state of step $n-1$, $h_{\mathrm{t}-1}$ is the output of previous layer.

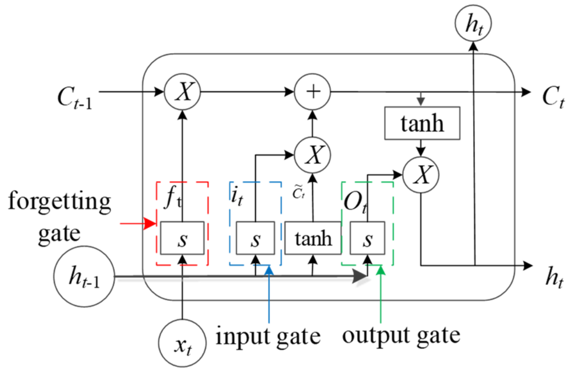


Figure 3. LSTM network structure.

2.3. MK-MMD

MMD [25] is defined as the square of the kernel embedding distance of the data distribution in reproducing the kernel Hilbert space, which is a commonly used non-parametric metric function to evaluate the difference between two distributions. For datasets s and t with sufficient data and different distributions, the MMD distance formula is as follows:

$$MMD^2(s,t) = \left\| \frac{1}{n^s}\sum_{i=1}^{n^s}\varphi(x_i^s) - \frac{1}{n^t}\sum_{j=1}^{n^t}\varphi\left(x_j^t\right) \right\|_{\Re}^2 \tag{7}$$

where $\varphi(\cdot)$ is the mapping function from the feature space to the reproducing kernel Hilbert space, and n^s and n^t represent the number of samples in the dataset s and t, respectively.

When calculating the distances of different distributions in Hilbert space, the multiple kernel functions are used to calculate the MMD [26] as

$$MMD^2(s,t) = \frac{1}{n^s n^s}\sum_{i=1}^{n^s}\sum_{j=1}^{n^s}k\left(x_i^s, x_j^s\right) + \frac{1}{n^t n^t}\sum_{i=1}^{n^t}\sum_{j=1}^{n^t}k\left(x_i^t, x_j^t\right) - \frac{1}{n^s n^t}\sum_{i=1}^{n^s}\sum_{j=1}^{n^t}k\left(x_i^s, x_j^t\right) \tag{8}$$

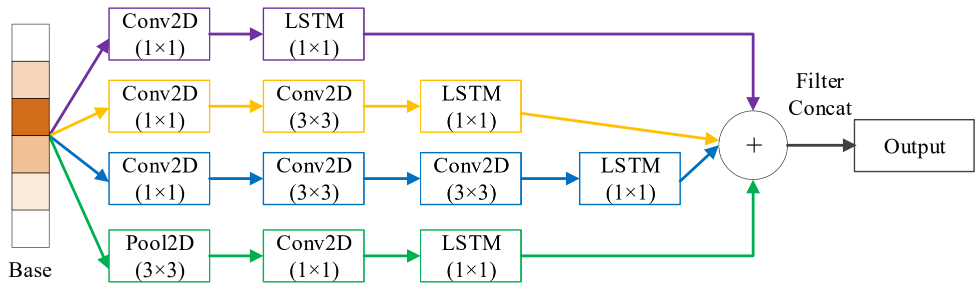
Traditionally, the differentiable Gaussian kernel is used to optimize the gradient descent for the loss function. Considering the probability distributions in different reproducing kernel Hilbert spaces [27], the multiple kernel functions in the MMD method, which can improve the adaptively to the different data and measure the distance between two different distributions to extract good feature maps, are introduced to guarantee the optimal performance of the system. The MK-MMD with multiple Gaussian kernels with different parameters is adopted for superposition calculation and developing a method for optimal kernel selection and improving its effectiveness.

3. Proposed Method

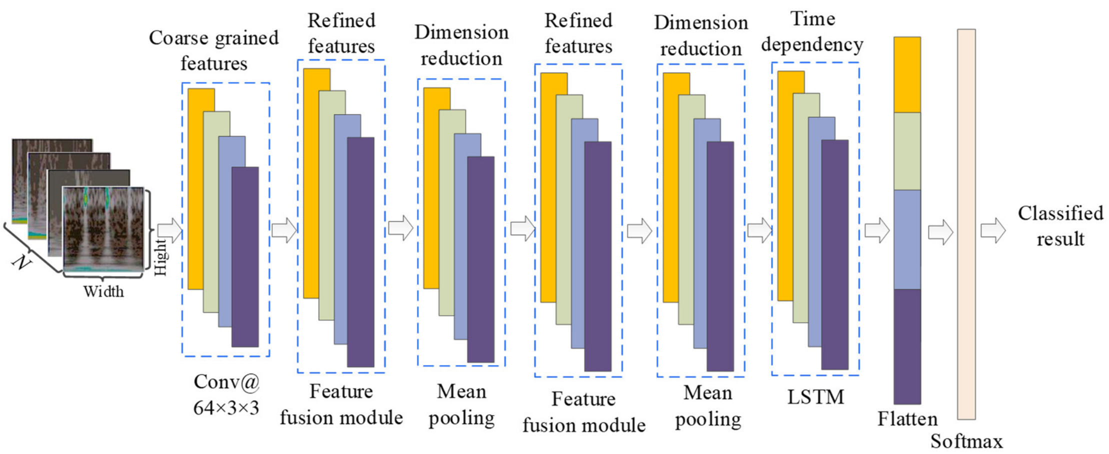
3.1. Variable-Scale Attention Mechanism-Guided Feature Fusion Framework

To accurately extract multi-scale fault features and the temporal dependency relationship from the time-frequency representation of roller bearing monitoring signals, a novel neural network framework with Inception and the LSTM model is established to efficiently diagnose the health of the bearings. The network structure is shown in Figure 4.

First, an enhanced feature fusion module based on Inception V2 is proposed, as shown in Figure 4a. Convolutional kernels with different scales are used to capture inconspicuous differential features from the input data, and the LSTM module is joined with each branch of the Inception module to learn the temporal dependency features. To improve the generalization and fitting ability of feature fusion, batch normalization is used to improve the consistency of the learned features, and the ReLU nonlinear function is used to activate the learned features to prevent the overfitting of the model.



(**a**)



(**b**)

Figure 4. Intelligent fault diagnosis framework for roller bearing based on the variable-scale attention mechanism-guided time-frequency feature fusion module. (**a**) Typical structure of the variable-scale attention mechanism-guided feature fusion module based on Inception and LSTM. (**b**) The intelligent fault diagnosis framework based on the variable-scale attention mechanism-guided feature fusion module.

Second, based on the variable-scale attention mechanism-guided feature fusion module, a novel intelligent fault diagnosis framework for roller bearing is proposed, as shown in Figure 4b. A convolutional layer with 3 × 3 small kernels is used to extract coarse grained features from the time-frequency map to increase the nonlinear expression ability of the network. The optimal activation function and batch normalization are introduced to modify the output. Two stacked feature fusion modules are used as the multi-scale refined feature extractor to further extract differentiation deep representation features, which can increase the overall nonlinear expression ability of the network and reduce network complexity. After each feature fusion module, mean pooling is introduced to reduce the feature dimension. The parameters of the constructed neural network are shown in Table 1.

Table 1. The parameters of the constructed neural network framework for intelligent fault diagnosis.

Layer	Parameter Name	Parameter Size	Output Size
Input layer	/	/	Batch size × 128 × 128
Conv	Kernels	3 × 3	Batch size × 3 × 128 × 128
Feature fusion module 1	Kernel	As in Figure 4a	Batch size × 16 × 64 × 64
Mean Pooling	Mean pooling size	2 × 2	Batch size × 64 × 32 × 32
Feature fusion module 2	Kernels	As in Figure 4a	Batch size × 128 × 16 × 16
Mean pooling	Mean pooling size	2 × 2	Batch size × 256 × 8 × 8

Then, after extracting refined features, the LSTM layer is used to receive the features and further learn the temporal dependencies between them. Subsequently, the Softmax function [28] is introduced to map the learned features into the probability space and select the corresponding labels. Meanwhile, the improved cross entropy function is used as the loss function to train the networks and accelerate the convergence speed of the network.

3.2. Variable-Scale Attention Mechanism-Guided Feature Transfer Learning Framework

On the basis of the feature fusion framework and the transfer learning method and taking the vibration signal of the roller bearing as the source domain and the vibration signal of the through shaft bearings as the target domain, an intelligent transfer diagnosis model for through shaft bearings is constructed. The network structure is shown in Figure 5. MK-MMD utilizes the kernel function to map the source and target domain features obtained from the LSTM layer to the reproduced Hilbert space, solving for the distribution distance between the two domains. The modified cross-entropy function is the loss function for the gradient backpropagation and the object optimization and reduces the distribution differences between the source and target domains. Then, under the cross-condition and using fault label samples, the fault diagnosis of through shaft bearings can be achieved.

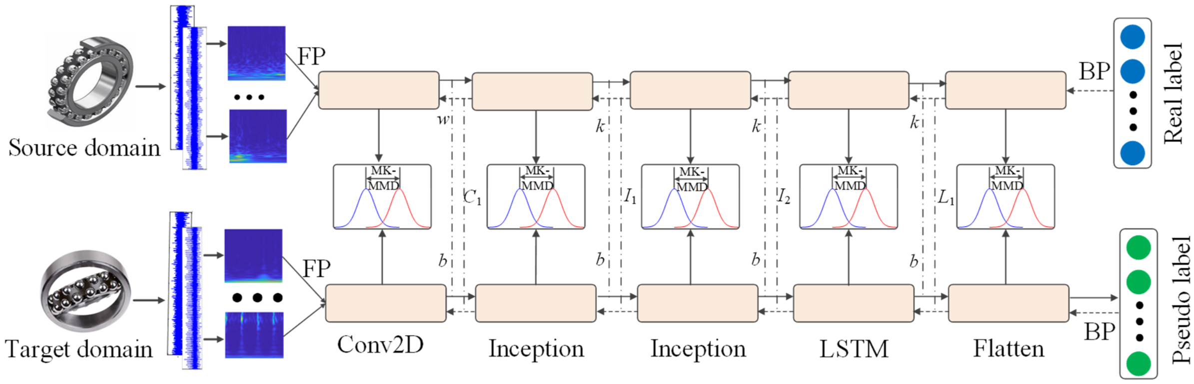


Figure 5. Intelligent transfer learning framework for through shaft bearing.

3.3. Fault Diagnosis Procedure of Roller Table Bearing System of the Annealing Kiln

Based on variable-scale attention mechanism-guided time-frequency feature fusion module and transfer learning, the fault diagnosis procedure of the roller table bearing system in the annealing kiln is as shown in Figure 6.

(1) Accelerometers are used to collect the vibration signal from the bearings in the roller table system of the annealing kiln. The collected vibration signal is decomposed into several CPFs via ITD, and the effective components are selected using the Gini–kurtosis criterion, as shown in Equation (9). Then, the denoised bearing vibration signal can be reconstructed for further analysis. GI can not only measure the impulsiveness of the signal but has a strong robustness against the random impulse noise, which is a great concern for signal processing methods based on other sparsity indexes. Impulsiveness is important but not the exclusive feature of the fault components in the measured fault signal. GI, which is a rare statistical index that can highlight repetitive fault components without prior knowledge, is not fully utilized.

For signal $x(t)$, because Gini and kurtosis are dimensionless indicators, the Gini–kurtosis criterion is defined as

$$\mathrm{Index}_{\mathrm{Gini-Kurtosis}} = \left[1 - 2\sum_{n=1}^{N} \frac{x_{(n)}}{\|\vec{x}\|}\left(\frac{N-n+0.5}{N}\right)\right] \cdot \left[\frac{\langle x^4(t,f)\rangle}{\langle X^2(t,f)\rangle^2} - 2\right] \tag{9}$$

where $x(t,f)$ is the time-frequency envelope of the filtered signal $x(t)$ around f, and $\langle\cdot\rangle$ represents the time-frequency averaging operator. The effective components can be determined based on whether the Gini–kurtosis index is larger than parameter η.

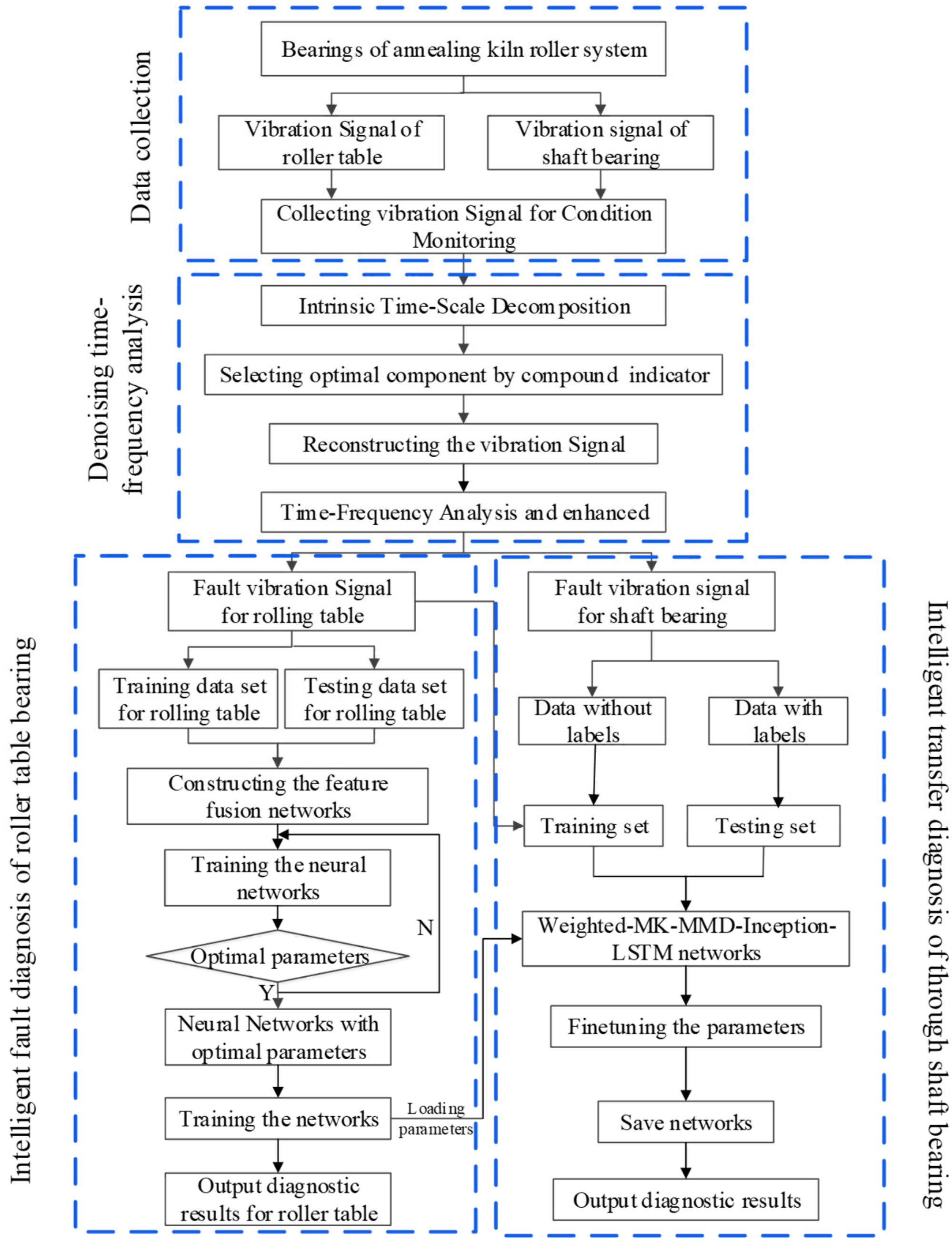


Figure 6. Fault diagnosis process for bearings in the roller table system of annealing kiln.

(2) The time-frequency representation of the denoised signals is obtained using the CWT method with the Morlet wavelet basis function; then, histogram equalization can enhance the time-frequency map.

(3) The time-frequency maps are organized as roller bearing datasets and input into the variable-scale attention mechanism-guided time frequency feature fusion neural network of the roller bearing to intelligently diagnose, train, and obtain the optimal hyperparameters of this network.

(4) Based on the optimal neural network and hyper-parameters, training is completed and the roller bearing intelligent diagnosis network is saved, accurately identifying the health condition of the roller bearing.

(5) Combining the MK-MMD method, an intelligent transfer diagnosis network for through shaft bearings is established, and the parameters of the trained optimal feature fusion intelligent diagnosis network are loaded.

(6) Using the roller and through shaft bearing datasets as the source and target domains, respectively, the roller and labeled through shaft bearing data as the training set, and the unlabeled through shaft bearing data as the test set, the transfer network is trained, the network parameters are fine-tuned by repressing part of the layer, and the through shaft bearing diagnosis results finally generated.

4. Case Study and Verification

4.1. Introduction to Experiment

The experimental data are from a rolling bearing fault experimental platform, as shown in Figure 7a, and the bearing used for testing is an N205EM cylindrical roller bearing. The single point faults were processed on the bearing using electric discharge technology, including the inner ring, outer ring, and roller element faults, as shown in Figure 7b–e. The fault diameters were 0.3, 0.6, and 1 mm, respectively. The tested bearing is installed inside the bearing seat, and three-axis acceleration sensors 1A313E are installed on the seat, as shown in Figure 7. The sampling frequency of the vibration signals is 20 kHz. To simulate the different working conditions, the tested bearing load was set to 0, 30, and 60 kg, and the driving motor speed was set to 800, 1200, and 1500 r/min. The vibration signal of each fault diameter was collected by our vibration testing equipment. According to the speed conditions, datasets A, B, and C are used to verify the effectiveness of intelligent transfer learning.

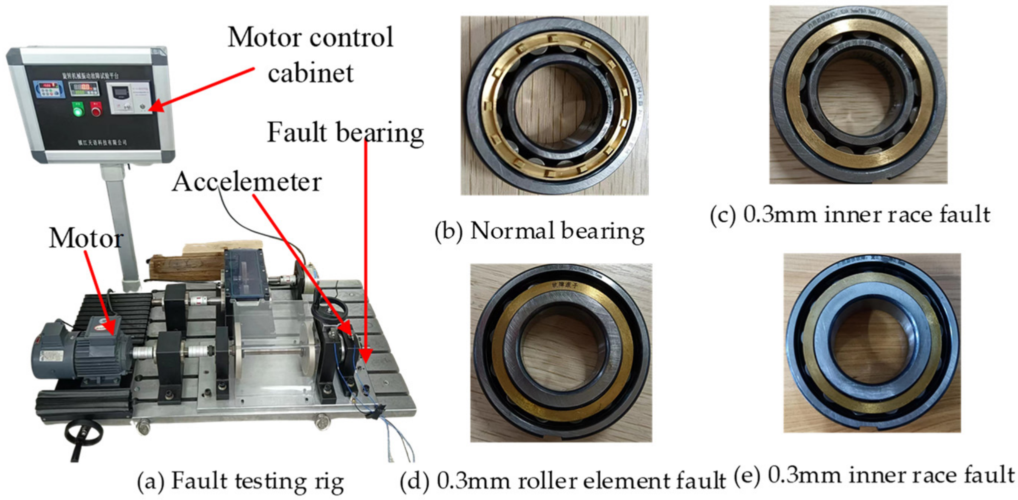


Figure 7. Testing rig of rolling bearing faults and fault bearings.

In this test, the bearing's rotation frequency ranged from 800 to 1500 r/min. Between 800 and 1500 points were sampled in each revolution. To better illustrate the vibration characteristics of each revolution, 1200 points were selected as one sample with a step size of 256. Furthermore, in order to improve and enhance the effectiveness and resolution of the time-frequency analysis, each sample has 1200 points, which provides better time-frequency characteristics. Four hundred samples were collected from each fault vibration signal, and a total of 5200 samples were obtained. Considering the cross-validation experiment, each dataset was divided into the training and testing sets at a 3:1 ratio. Thus, the training set has a total of 3900 samples, and the testing set has a total of 1300 samples. The detailed information of this experimental data is shown in Table 1. As an example, the fault vibration signals with a damage diameter of 0.3 mm and 800 r/min on different components are shown in Figure 8a–d. According to the proposed signal decomposition and reconstruct method, the denoised vibration signals of different faults are shown in Figure 8e–h. The comparison of the signal-to-noise (SNR) ratio of the signals in Table 2 shows that the SNR is 12.45, and the root mean square is 0.44, indicating a good noise-reduction effect. Then, after applying the CWT and histogram equalization methods, 400 enhanced time-frequency maps were obtained for each state. The time-frequency representations before and after the enhancement of the bearing signal are shown in Figure 9. The enhanced time-frequency maps have good clustering and a strong local feature prominence. Here, the computer configuration is the Inter Intel (R) Core (TM) i7-10510U CPU and 16GB RAM in window 11 and Matlab 2022b.

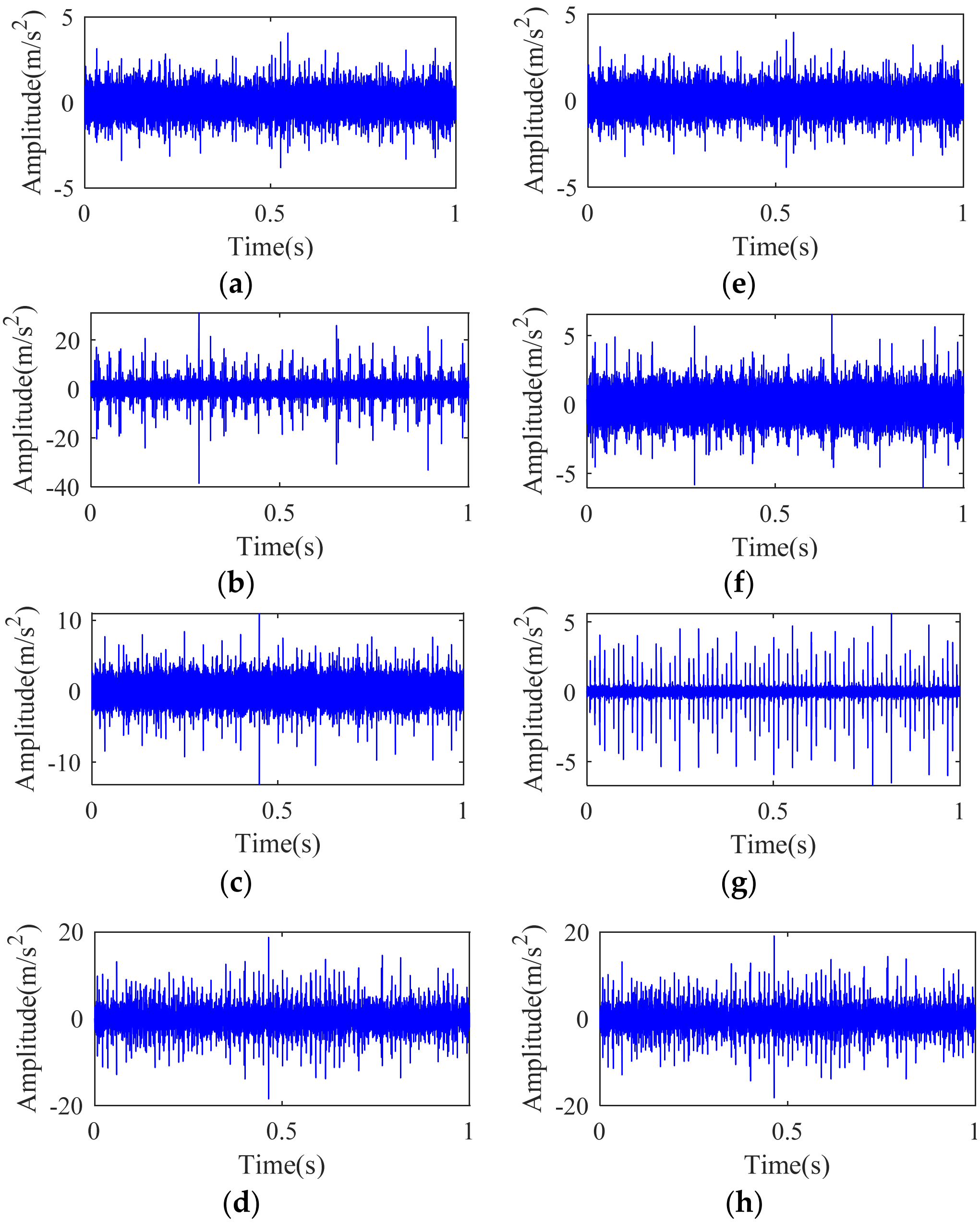
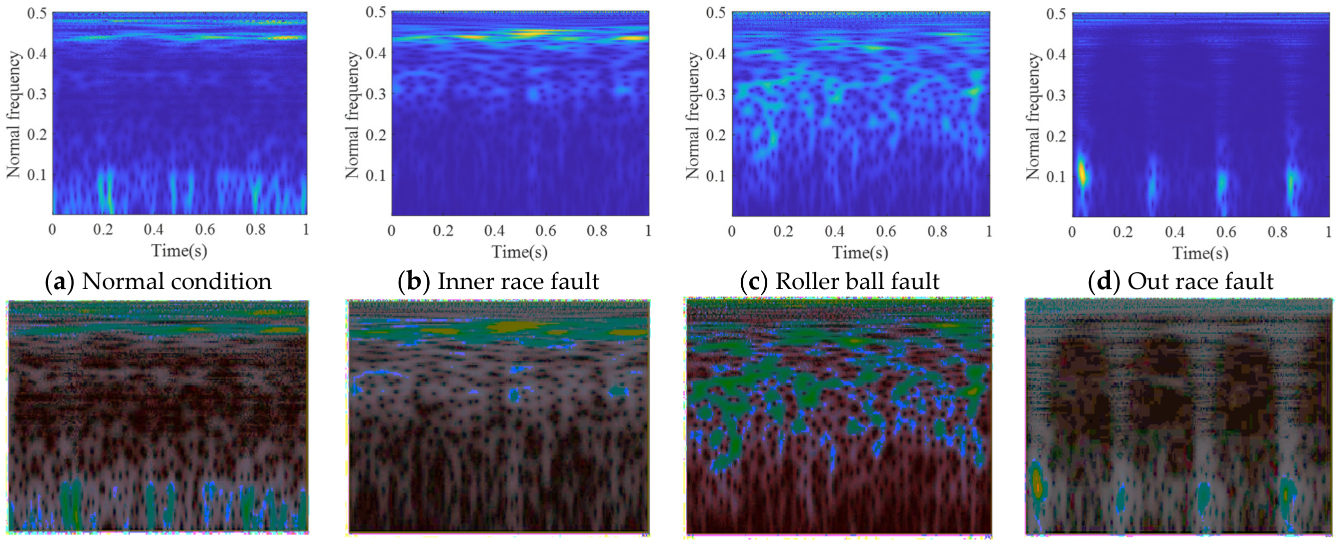


Figure 8. Vibration signal of roller table bearing under different health conditions. (**a**) Normal−condition vibration signal of roller table. (**b**) Original vibration signal of inner race fault with 1 mm. (**c**) Original vibration signal of ball fault with 1 mm. (**d**) Original vibration signal of out race fault with 1 mm. (**e**) Denoised normal−condition vibration signal of roller table. (**f**) Denoised original vibration signal of inner race fault with 1 mm. (**g**) Denoised original vibration signal of ball fault with 1 mm. (**h**) Denoised original vibration signal of out race fault with 1 mm.

Table 2. Organization of experimental dataset.

Dataset	Healthy Condition	Fault Diameter/mm	Training Samples	Testing Samples	Label
A/B/C 800 r/min/ 1200 r/min/1500 r/min	Normal state	0	300/300/300	100/100/100	1
	ball faults	0.3	300/300/300	100/100/100	2
		0.6	300/300/300	100/100/100	3
		1.0	300/300/300	100/100/100	4
		1.5	300/300/300	100/100/100	5
	Inner race fault	0.3	300/300/300	100/100/100	6
		0.6	300/300/300	100/100/100	7
		1.0	300/300/300	100/100/100	8
		1.5	300/300/300	100/100/100	9
	Outer race fault	0.3	300/300/300	100/100/100	10
		0.6	300/300/300	100/100/100	11
		1.0	300/300/300	100/100/100	12
		1.5	300/300/300	100/100/100	13



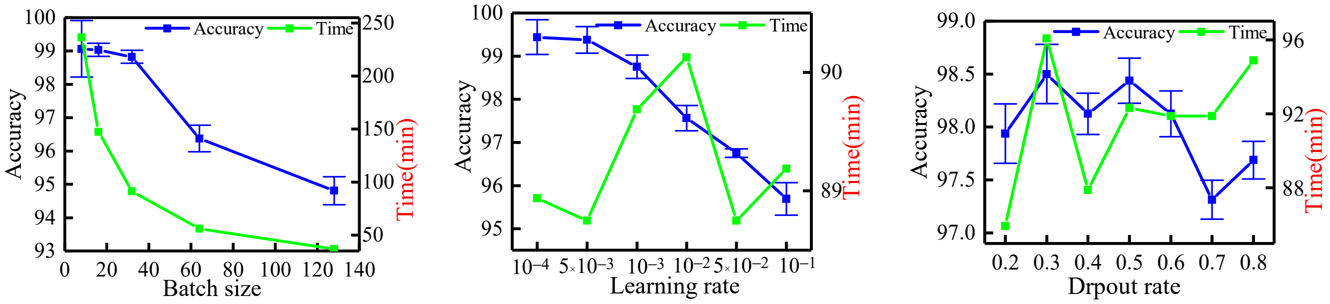
(**a**) Normal condition (**b**) Inner race fault (**c**) Roller ball fault (**d**) Out race fault

(**e**) Normal condition (**f**) Inner race fault (**g**) Roller ball fault (**h**) Out race fault

Figure 9. Time-frequency representation and their histogram equalization for the roller table bearing under different health conditions.

4.2. Experimental Parameter Configuration

For the proposed intelligent fault diagnosis framework, the optimal hyperparameters, the batch size, the learning rate, and the dropout rate greatly affect the training speed, the effectiveness of extracted features, the classification accuracy, and the model robustness [20]. To select the optimal hyperparameters, the impact on the classification accuracy and time consumption of this model under different minimum numbers of batches, the initial learning rate and the dropout rate are shown in Figure 10, and the optimal model parameter configuration was determined. The adaptive moment estimation method is used to optimize the cross-entropy loss function and train the neural network framework. The average recognition accuracy and training time are taken as evaluation indicators, and the results are shown in Figure 10.



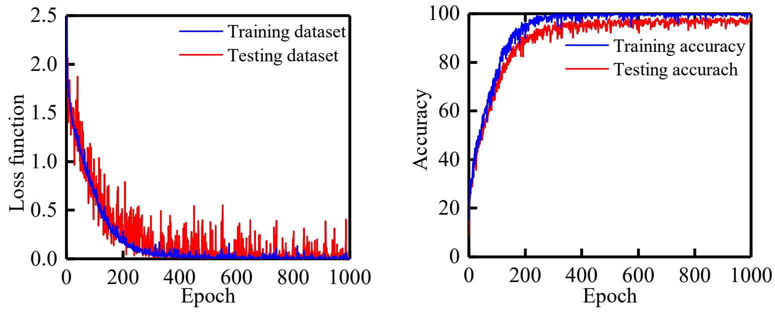
(**a**) Batch size (**b**) Learning rate (**c**) Dropout rate

Figure 10. Influence of hyperparameters on model accuracy.

Figure 10 shows that the batch size has the greatest impact on the accuracy of this model, followed by the learning rate, and the maximum dropout rate has the smallest effect. In Figure 10a, the accuracy and training times decrease as the minimum batch size increases. When the minimum number of batches increases from 16 to 32, the accuracy remains unchanged, although the training time decreases significantly. Therefore, 16 is selected as the optimal batch size. In Figure 10b, the accuracy first increases and then decreases due to the increase in the learning rate, and the training time shows no significant change. The optimal effect is achieved when the learning rate is 0.01. In Figure 10c, the accuracy is decreased when the dropout rate is increased, but the training time has no significant effect on the result. The optimal dropout rate is 0.3.

Hence, the proposed neural network was trained with the optimal hyperparameters, and the typical training curve of the loss function and the training accuracy are shown in Figure 11, with stable fluctuations and network convergence. The fault diagnosis' average accuracy for the roller table bearing reaches 99.90%, and the training error is 0.13%. The confusion matrices of the training and validation sets are shown in Figure 12. The horizontal axis represents the real label, and the vertical axis represents the predicted label. Few samples were misjudged in each of the rolling element and inner circle fault states, and the rest were accurately identified, verifying the recognition accuracy of the proposed variable-scale attention mechanism-guided time-frequency feature fusion network.

In order to verify the anti-noise interference ability and robustness of this proposed method, Gaussian white noise with different decibels is added into the experimental signal. Furthermore, considering to further illustrate the effectiveness of the proposed data preprocessing method in this paper, time-domain, frequency-domain, CWT time-frequency domain data, and the proposed preprocessing data are used to train the constructed neural network, respectively. The results are shown in Figure 13a, the diagnostic accuracies of the proposed method are above 98% under different noise interferences, and they have lower error. Therefore, compared with other input data, the proposed data preprocessing method shows better effectiveness and robustness under different strong background noise environments.



(**a**) Loss function (**b**) Accuracy curve

Figure 11. Training process of loss function and accuracy curve.

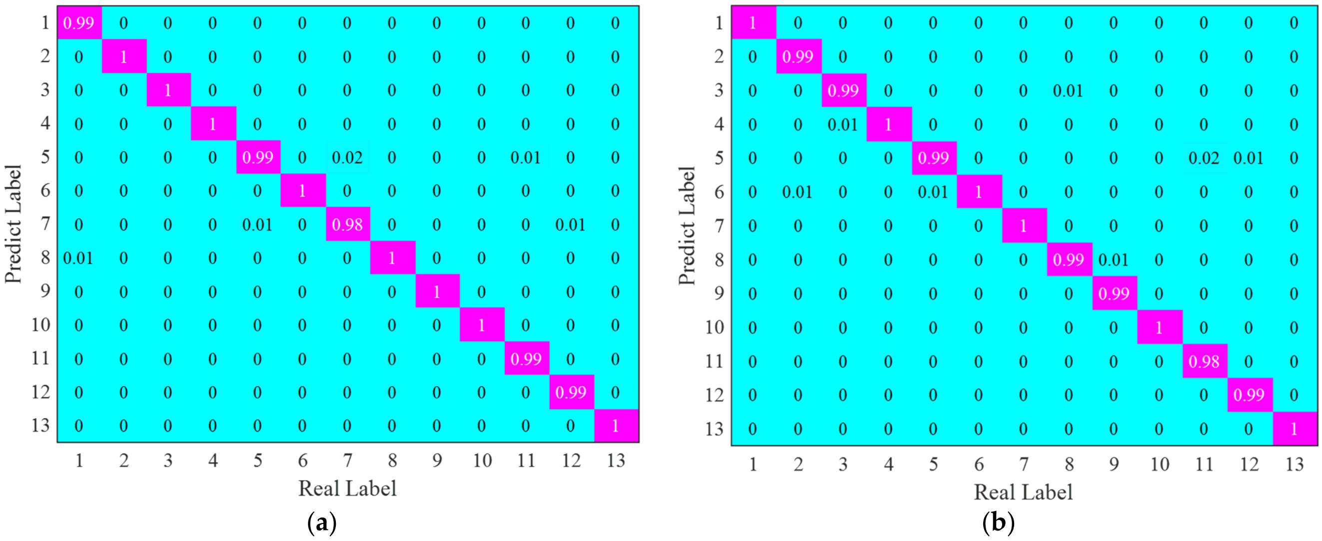


Figure 12. Confusion matrix of training and testing set. (**a**) Confusion matrix of training dataset. (**b**) Confusion matrix of testing dataset.

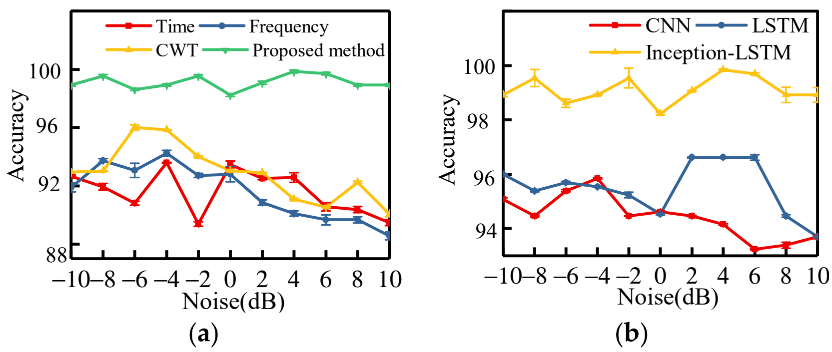


Figure 13. The verification of fault diagnosis effectiveness under different noise interferences. (**a**) The accuracy and error of fault diagnosis for different input data under different noise interferences. (**b**) The accuracy and error of fault diagnosis for different neural network under different noise interferences.

To evaluate the anti-noise interference ability of the proposed neural network, we added Gaussian white noise with varying decibels to the collected fault signal. We compared the performance of the CNN [29], LSTM [16], and proposed methods in identifying fault features, as shown in Figure 13b. The results demonstrate that the proposed method outperforms the CNN and LSTM methods under different levels of noise interference during the 10 training sessions. This highlights the superior anti-noise interference capabilities and robustness of the proposed method. Due to the characteristics of the input data, framework, and parameters of the neural network, the LSTM has better identification performance than the CNN for the noise-added input signal.

4.3. Analysis of Intelligent Transfer Diagnostic Method

To verify the effectiveness of the proposed transfer learning model, the feature fusion network was used as the basic framework, and datasets A, B, and C were selected as the experimental datasets. Using one of datasets A, B, and C as the source domain and the others as the as the target domain, two sets of transfer experiments were transferred from A to B and A to C. Each group was randomly trained 10 times. The average identification accuracy was taken as the evaluation index. Six sets of transfer experiments were conducted.

For Experiment 1, the CNN-LSTM network [30] was selected. For Experiment 2, the DCORAL method was adopted, using Coral as the metric function [27]. For Experiment 3, the dynamic distribution adaptation (DDA) method was adopted, using a Gaussian kernel MMD [31] as the metric function. For Experiment 4, the proposed intelligent transfer method was adopted, using MK-MMD as the metric function. The experimental results are shown in Table 3 and Figure 14. In Experiment 1, the transfer diagnostic displayed the poorest effectiveness. In Experiment 2, the accuracy of transfer diagnosis was 90.52%, and the error was 3.54%. In Experiment 3, the average accuracy was 94.23%, and error was 1.95%. In Experiment 4, using the multi-kernel MMD method, the average accuracy was 99.11%, and a negligible error was obtained. Moreover, the calculation time of the proposed method is slightly longer than other methods. Hence, the proposed method shows good robustness and is more effective than the other methods.

Table 3. Experimental average diagnostic results.

Experiment	Experiment Method	Average Accuracy/%	Error/%	Time/min
Experiment 1	CNN-LSTM	85.67	3.53	86.15
Experiment 2	DCORAL	90.52	3.54	82.46
Experiment 3	DDA	94.23	1.95	92.65
Experiment 4	The proposed method	99.11	0.86	97.92

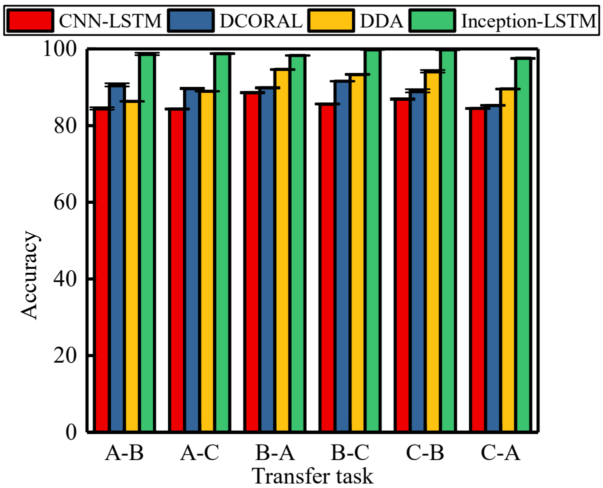


Figure 14. Accuracy rate of fault transfer diagnosis of rolling bearing.

4.4. Fault Diagnosis of Bearings in the Roller Table System of the Annealing Kiln

Here, the proposed method is introduced to diagnose the health condition of the roller and through shaft bearings in the roller table system of the annealing kiln. In a glass production plant, vibration signals of the annealing kiln roller system were collected, as shown in Figure 15. The collected bearing vibration signals were divided into two states: normal and fault. The sampling frequency was 12 KHz, and vibration signals of the roller bearing system were collected using a three-axis acceleration sensor and an NI 9234 data acquisition card.

The collected vibration signals under the normal and fault conditions of the bearings are shown in Figure 16. According to the proposed method, the original signal was decomposed, and the effective component was selected to filter the noise interference. Then, the SNR of the roller bearing vibration signal was 23.11, and the mean square error was 0.52. The SNR of the through shaft bearing vibration signal was 24.63, and the mean square error was 0.96. According to the CWT method, the time-frequency representations are shown in Figure 17, and on the basis of the histogram equalization method, the local features of the enhanced time-frequency map are clear and prominent.

(**a**) (**b**)

Figure 15. Healthy condition monitoring for the roller system of the annealing kiln. (**a**) Vibration signal collection of the roller table bearing (In the red box) (**b**). Vibration signal acquisition of the through shaft bearing (In the red box).

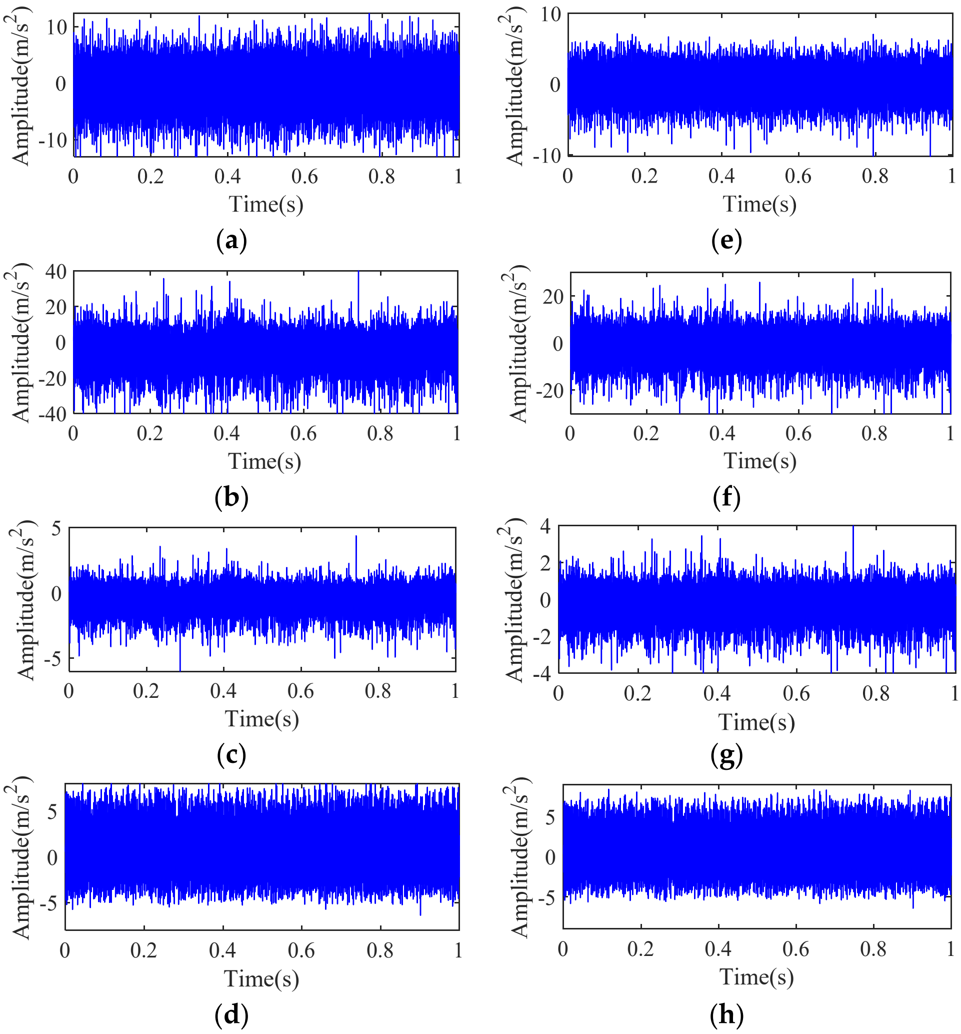


Figure 16. Vibration signals of the roller table system of the annealing kiln. (**a**) Vibration signal of normal roller table bearing. (**b**) Vibration signal of fault roller table bearing. (**c**) Vibration signal of normal through shaft bearing. (**d**) Vibration signal of fault through shaft bearing. (**e**) Denoised vibration signal of normal roller table bearing. (**f**) Denoised vibration signal of fault roller table bearing. (**g**) Denoised vibration signal of normal through shaft bearing. (**h**) Denoised vibration signal of fault through shaft bearing.

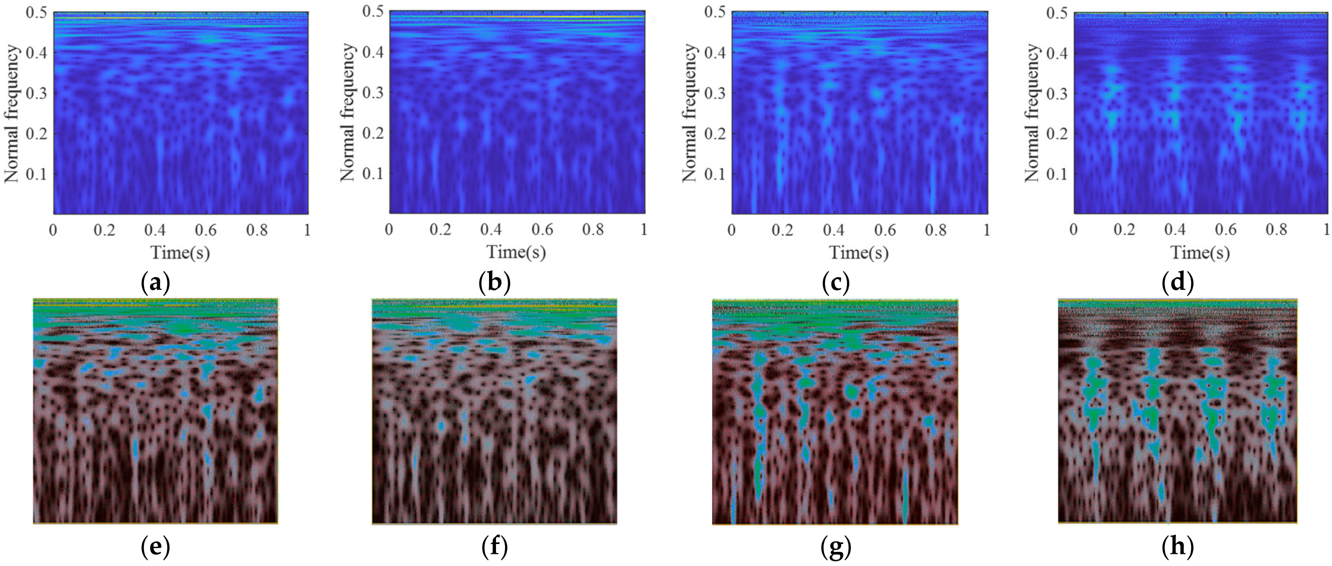


Figure 17. Time-frequency representation and their histogram equalization of roller table and through shaft bearing. (**a**) Normal condition of roller table bearing. (**b**) Fault condition of roller table bearing. (**c**) Normal condition of through shaft bearing. (**d**) Fault condition of through shaft bearing. (**e**) Normal condition of roller table bearing. (**f**) Fault condition of roller table bearing. (**g**) Normal condition of through shaft bearing. (**h**) Fault condition of through shaft bearing.

According to the proposed method, the roller table bearing datasets were organized into 400 labeled samples and divided into the training and testing sets in a 9:1 ratio. The datasets were inputted into the roller bearing intelligent diagnosis network and trained 10 times. The average accuracy of the model reached 99.65%, and the loss function and accuracy curve of roller table dataset is shown in Figure 18. After 500 training iterations, the fluctuation of the accuracy and loss curve converged, and the network was completely trained. Using the roller table bearing datasets as the source domain and the through-shaft bearing dataset as the target domain, the roller table bearing data and 80 unlabeled through shaft bearing data were used as the training set, and 40 labeled through shaft bearing data were used as the test set. The intelligent transfer diagnosis network for through-axis bearings was introduced and trained 10 times. The average accuracy of the model for the through shaft bearing was 99.20%. The training curve is shown in Figure 19, which shows that the accuracy and loss curve fluctuates steadily and ultimately converges. This curve indicates that the proposed diagnostic model is effective and feasible and has high a recognition accuracy. Furthermore, the confusion matrix of the bearing conditions is shown in Figure 20. Few normal and fault roller table bearings were misjudged as normal and fault through shaft bearings, respectively. Comprehensively, the transfer model can effectively identify the fault condition for the bearings in the roller system of the annealing kiln.

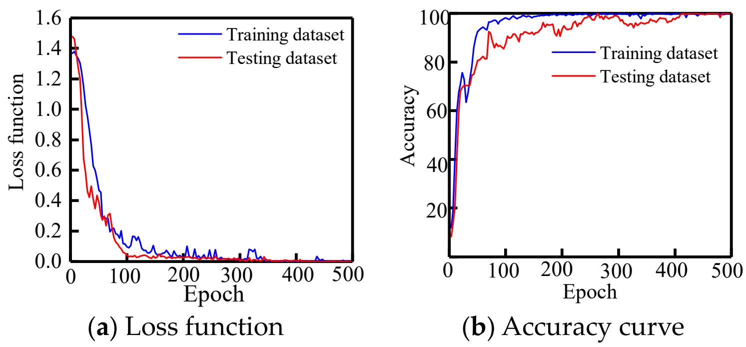
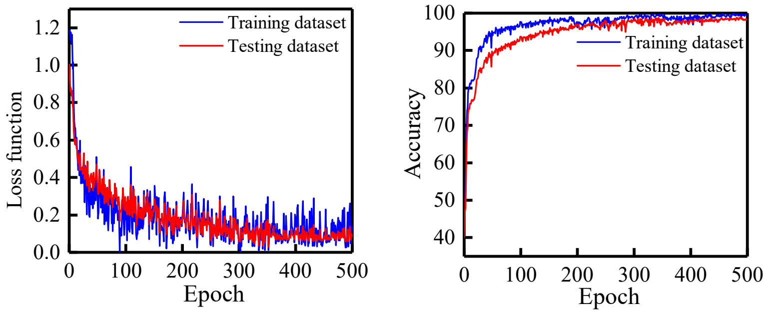
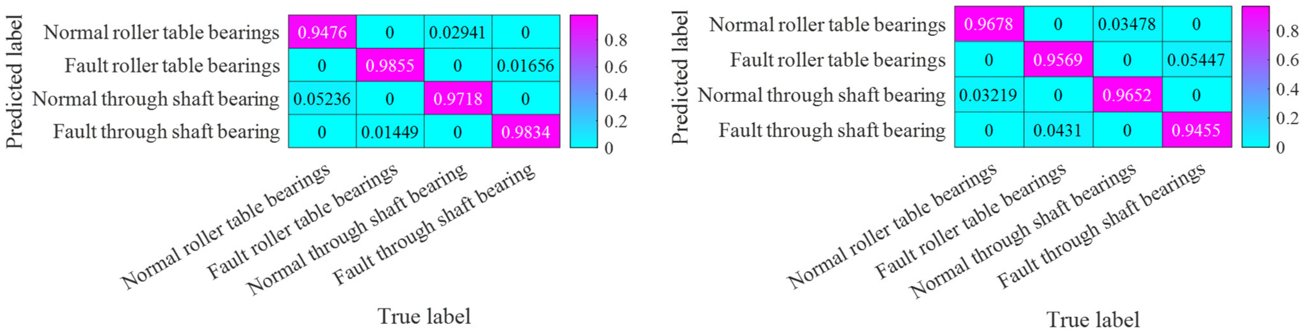


Figure 18. Loss function and accuracy curve of training process.



(**a**) Loss function (**b**) Accuracy curve

Figure 19. Loss function and accuracy curve of transfer learning training process.



(**a**) (**b**)

Figure 20. Identification confusion matrix of the bearings for the intelligent transfer diagnostic model. (**a**) Confusion matrix of training data of through shaft bearing. (**b**) Confusion matrix of the testing data of through shaft bearing.

Furthermore, the collected monitoring data from the bearing system of the annealing kiln were put into CNN-LSTM, DCORAL, and DDA neural network to identify the working condition. The results in Table 4 show that identification accuracy of the proposed method for the intelligent fault transfer diagnosis of the bearing system in the annealing kiln is 99.85%, and the mean error is 0.08%. Compared to other method, the result is better.

Table 4. Comparison of intelligent fault transfer diagnosis for bearing system in annealing kiln.

Method	CNN-LSTM	DCORAL	DDA	The Proposed Method
Accuracy/%	86.30769 ± 0.05015	91.84615 ± 0.02794	93.84615 ± 0.01738	99.84615 ± 0.08406
Time/min	45.53	37.23	48.61	72.53

5. Conclusions

This work proposes an intelligent transfer diagnosis method for bearing faults in the annealing kiln roller system on the basis of the Inception and LSTM modules to monitor the condition and diagnose the faults of the roller table and through shaft bearings under different working conditions. This method can overcome the influence of strong mechanical background noise interference and inconspicuous sample features and efficiently and accurately determine the health condition of bearings. The main conclusions of the paper are as follows:

(1) Considering the sparsity and kurtosis of vibration signals, the decomposed components using the ITD method can select and denoise strong mechanical noise, and CWT and histogram equalization can obviously enhance the time-frequency representations, clearly displaying the differentiated feature information on the time-frequency maps and promoting intelligent identification.

(2) By combining Inception's multi-scale feature extraction and the LSTM's temporal relationship learning ability, an intelligent diagnosis framework for the roller table bearing is established to learn fusion features from the time-frequency representation. The experimental results of the measured data indicate that the proposed framework has excellent feature-extraction ability and can effectively and accurately identify the health condition of roller bearings.
(3) By combining the MK-MMD method, an intelligent transfer diagnostic framework is established to identify the health condition using cross-domain condition samples. This method can reduce the distribution differences in vibration data and achieve the intelligent transfer diagnosis of roller table and through shaft bearings under different speed conditions in the annealing kiln roller system. Considering the plant scene and operation condition of the glass product line, the verification results from the experimental and measured data show that the intelligent transfer diagnostic model established in this work has high recognition accuracy for the health condition and fault characteristics of roller and through shaft bearings.

Author Contributions: Data curation, S.L. and T.L.; Methodology, K.Z.; Writing—original draft, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the natural science foundation of Chongqing (cstc2020jcyj-msxmX0334), Science and Technology Research Project of Chongqing Education Commission (KJQN202101119) Chongqing Doctor's Research Project (CSTB2022BSXM-JCX0163).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article.

Acknowledgments: We sincerely thank the reviewers for their constructive comments and the editors for their patient replies.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Abid, A.; Khan, M.T.; Iqbal, J. A review on fault detection and diagnosis techniques: Basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639–3664. [CrossRef]
2. Niu, Y.; Fei, J.; Li, Y.; Wu, D. A novel fault diagnosis method based on EMD, cyclostationary, SK and TPTSR. *J. Mech. Sci. Technol.* **2020**, *34*, 1925–1935. [CrossRef]
3. Randall, R.B.; Antoni, J. Why EMD and similar decompositions are of little benefit for bearing diagnostics. *Mech. Syst. Signal Process.* **2023**, *192*, 110207. [CrossRef]
4. Zhang, C.; Li, Z.; Hu, C.; Chen, S.; Wang, J.; Zhang, X. An optimized ensemble local mean decomposition method for fault detection of mechanical components. *Meas. Sci. Technol.* **2017**, *28*, 035102. [CrossRef]
5. Li, F.; Li, R.; Tian, L.; Chen, L.; Liu, J. Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions. *Mech. Syst. Signal Process.* **2019**, *116*, 462–479. [CrossRef]
6. Imaouchen, Y.; Kedadouche, M.; Alkama, R. A frequency-weighted energy operator and complementary ensemble empirical mode decomposition for bearing fault detection. *Mech. Syst. Signal Process.* **2017**, *82*, 103–116. [CrossRef]
7. Cheng, Y.; Wang, Z.; Chen, B.; Zhang, W.; Huang, G. An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis. *ISA Trans.* **2019**, *91*, 218–234. [CrossRef]
8. Cheng, J.; Yang, Y.; Hu, N.; Cheng, Z.; Cheng, J. A noise reduction method based on adaptive weighted symplectic geometry decomposition and its application in early gear fault diagnosis. *Mech. Syst. Signal Process.* **2021**, *149*, 107351. [CrossRef]
9. Feng, Z.; Lin, X.; Zuo, M.J. Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *72*, 223–240. [CrossRef]
10. Geng, Y.; Pu, W. Quantitative identification of pulse-like ground motions based on intrinsic time-scale decomposition technique. *Soil Dyn. Earthq. Eng.* **2023**, *171*, 107977. [CrossRef]
11. Hu, Y.; Bao, W.; Tu, X.; Li, F.; Li, K. An adaptive spectral kurtosis method and its application to fault detection of rolling element bearings. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 739–750. [CrossRef]
12. Hou, B.; Wang, D.; Yan, T.; Wang, Y.; Peng, Z.; Tsui, K.L. Gini indices II and III two new sparsity measures and their applications to machine condition monitoring. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 1211–1222. [CrossRef]

13. Dong, H.; Yu, G.; Lin, T.; Li, Y. An energy-concentrated wavelet transform for time-frequency analysis of transient signal. *Signal Process.* **2023**, *206*, 108934. [CrossRef]
14. Zhang, Y.; Xing, K.; Bai, R.; Sun, D.; Meng, Z. An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image. *Measurement* **2020**, *157*, 107667. [CrossRef]
15. Li, X.; Zheng, J.; Li, M.; Ma, W.; Hu, Y. One-shot neural architecture search for fault diagnosis using vibration signals. *Expert Syst. Appl.* **2022**, *190*, 116027. [CrossRef]
16. Qiao, M.; Yan, S.; Tang, X.; Xu, C. Deep convolutional and LSTM recurrent neural networks for Rolling bearing fault diagnosis under strong noises and variable loads. *IEEE Access* **2020**, *8*, 66257–66269. [CrossRef]
17. Shi, J.; Peng, D.; Peng, Z.; Zhang, Z.; Goebel, K.; Wu, D. Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks. *Mech. Syst. Signal Process.* **2022**, *162*, 107996. [CrossRef]
18. Guo, L.; Yu, Y.; Liu, Y.; Gao, H.; Chen, T. Reconstruction domain adaptation transfer network for partial transfer learning of machinery fault diagnostics. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2502710. [CrossRef]
19. Zou, Y.; Liu, Y.; Deng, J.; Jiang, Y.; Zhang, W. A novel transfer learning method for bearing fault diagnosis under different working conditions. *Measurement* **2021**, *171*, 108767. [CrossRef]
20. Cao, X.; Wang, Y.; Chen, B.; Zeng, N. Domain-adaptive intelligence for fault diagnosis based on deep transfer learning from scientific test rigs to industrial applications. *Neural Comput. Appl.* **2021**, *33*, 4483–4499. [CrossRef]
21. Che, C.; Wang, H.; Ni, X.; Fu, Q. Domain adaptive deep belief network for rolling bearing fault diagnosis. *Comput. Ind. Eng.* **2020**, *143*, 106427. [CrossRef]
22. Szegedy, C.; Liu, W.; Jia, Y. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: New York, NY, USA, 2015.
23. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: New York, NY, USA, 2016; pp. 2818–2826.
24. Serin, G.; Sener, B.; Ozbayoglu, A.M.; Unver, H.O. Review of tool condition monitoring in machining and opportunities for deep learning. *Int. J. Adv. Manuf. Technol.* **2020**, *109*, 953–974. [CrossRef]
25. Tian, X.; Li, Y.; Liu, T.; Wang, X.; Tao, D. Eigenfunction-based multitask learning in a reproducing kernel hilbert Space. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 1818–1830. [CrossRef]
26. Du, H.; Zhou, Y. Nostradamus: A novel event propagation prediction approach with spatio-temporal characteristics in non-Euclidean space. *Neural Netw.* **2022**, *145*, 386–394. [CrossRef]
27. Qin, Y.; Yao, Q.; Wang, Y.; Mao, Y. Parameter sharing adversarial domain adaptation networks for fault transfer diagnosis of planetary gearboxes. *Mech. Syst. Signal Process.* **2021**, *160*, 107936. [CrossRef]
28. Wang, H.; Liu, Z.; Peng, D.; Cheng, Z. Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. *ISA Trans.* **2021**, *128*, 470–484. [CrossRef]
29. Chen, X.; Zhang, B.; Gao, D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. *J. Intell. Manuf.* **2021**, *32*, 971–987. [CrossRef]
30. Xiao, R.; Zhang, Z.; Wu, Y.; Jiang, P.; Deng, J. Multi-scale information fusion model for feature extraction of converter transformer vibration signal. *Measurement* **2021**, *180*, 109555. [CrossRef]
31. Shi, H.; Shang, Y. Initial fault diagnosis of rolling bearing based on second-order cyclic autocorrelation and DCAE combined with transfer learning. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3503818. [CrossRef]