mathematics

MDPI

Article

A New Fault Diagnosis Method for Rolling Bearings with the Basis of Swin Transformer and Generalized S Transform

Jin Yan [1,2], Xu Zhu [1], Xin Wang [1] and Dapeng Zhang [1,*]

[1] Guangdong Provincial Key Laboratory of Intelligent Equipment for South China Sea Marine Ranching, Guangdong Ocean University, Zhanjiang 524088, China
[2] Shenzhen Research Institute, Guangdong Ocean University, Shenzhen 518120, China
* Correspondence: zhangdapeng@gdou.edu.cn

Abstract: In view of the rolling bearing fault signal non-stationarity, strong noise can lead to low fault diagnosis accuracy. A Swin Transformer and generalized S Transform fault diagnosis method is proposed to solve the problems of difficult signal feature extraction and low diagnostic accuracy. Generalized S transform is used to improve the resolution of bearing fault signals, the Swin Transformer model is used to master the shallow weight required for identifying rolling bearing faults for highly fault characteristic expression signals, and the deep weight is obtained by backpropagation training. Finally, the extracted features are input into the improved Softmax classifier for fault classification. The various signal processing methods for the bearing signal processing ability are compared, and this model's diagnosis ability and the ability to resist noise are verified. The experimental results show that the method has a remarkable ability and an accuracy of above 90% in the anti-noise test and also has a good robustness.

Keywords: rolling bearing; vibration signal; fault diagnosis; Swin Transform; generalized S transform

MSC: 37M10

check for updates

Academic Editors: Joaquim Infante Barbosa, José Alberto Rodrigues, Yiu Yin Raymond Lee and Carlos Conceicao Antonio

Received: 6 October 2024
Revised: 6 December 2024
Accepted: 23 December 2024
Published: 26 December 2024

Citation: Yan, J.; Zhu, X.; Wang, X.; Zhang, D. A New Fault Diagnosis Method for Rolling Bearings with the Basis of Swin Transformer and Generalized S Transform. *Mathematics* **2025**, *13*, 45. https://doi.org/10.3390/math13010045

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, in the context of the era of the continuous development and integration of techniques like deep learning, "Internet Plus (+)", Internet of Things (IOT), intelligent detection, etc., the industrial field is experiencing a huge leap from Industry "3.0" to Industry "4.0" [1]. The operational status of bearings, as the core components of modern machinery, which has a growing tendency to be high-speed and high-precision, directly affect the safety and efficiency of the entire production process. It is important to note that at the moment, 30% of all problems in rotating machinery equipment, including centrifugal fans, impellers, conveyors, etc., are caused by bearing failures in the great majority of industrial machinery [2]. Therefore, real-time, rapid, and precise diagnostics of rotating machinery's bearings have significant scientific relevance as well as practical utility. Traditional techniques include vibration signal analysis [3], sound signal analysis [4], lubricating oil analysis [5], acoustic emission detection [6], and so forth. The most popular of them is the vibration signal analysis approach.

Nowadays, with the rise of artificial intelligence technology, data-driven intelligent fault diagnosis methods based on machine learning algorithms have been widely studied [7–11]. Conventional machine learning algorithms for diagnostics mainly consist of

characteristic extraction and pattern classification processes [10,11], and the feature extraction usually uses fast Fourier transform [12], variation mode decomposition [13], statistical features [14], spectral analysis [15], wavelet transform [16], and Hilbert–Huang transform [17], as well as other advanced signal processing methods to extract time domain, frequency domain, and time–frequency domain features from raw fault data. Zhang [18] used a one-dimensional CNN for the fault diagnosis of rolling bearings, eliminating the need for noise reduction preprocessing in traditional fault diagnosis, directly inputting the original one-dimensional vibration signal into the CNN for feature extraction and classification, and introducing a certain degree of interference to enhance the anti-noise ability of the model. Fuan [19] proposed an adaptive one-dimensional CNN rolling bearing fault diagnosis method based on particle swarm optimization for the uncertainty of hyperparameter selection in a one-dimensional CNN, making the algorithm have a higher diagnostic accuracy and robustness. Eren [20] proposed a compact one-dimensional CNN fault diagnosis method, which takes the original vibration signal as the input and performs one-dimensional convolution, which greatly reduces the computing cost. XiaM et al. proposed a CNN-based fault diagnosis method for rotating machinery, which uses the structural characteristics of the CNN network to achieve the fusion of multi-sensing information, which has a higher diagnostic accuracy than traditional methods [21].

For environments with high noise levels and fluctuating workloads, Wei Zhang [22] suggested a rolling bearing problem diagnostic approach based on convolutional neural networks to eliminate the need for human feature extraction. Based on the bearing defect diagnosis framework, Xiaoli Zhao et al. [23] presented a normalized conditional variation auto-encoder with adaptive focal loss (NCVAE-AFL) to improve the dataset's feature learning capacity and achieve a better diagnostic accuracy. By combining supervised learning with episodic metric meta-learning, Duo Wang et al. suggested a meta-learning model based on feature space metrics and demonstrated the viability of the method via tests [24]. Using two phases of data reconstruction and meta-learning, the novel hierarchical recursive technique for data reconstruction suggested by Hao Su et al. is appropriate for small-sample bearing failure detection under various operating situations [25]. Li [26] put out a bearing failure diagnostic technique that can combine data from several sensors. This technique employs a binary tree support vector machine (BT-SVM) for pattern identification and defect diagnostics. The energy values of many sensors are employed as feature vectors. It successfully cuts down on the feature extraction time while increasing diagnostic precision. Ding Xu [27] et al. proposed a time–frequency manifold image synthesis method to realize bearing fault diagnosis. The above study builds images based on time–frequency methods to transform fault diagnosis problems into image classification problems.

Ville et al. [28] proposed the Wigner–Ville distribution (WVD) for secondary time–frequency analysis, which effectively improved the focus of time–frequency analysis. Abboud et al. [29] suggested an enhanced square envelope spectroscopy approach based on the time domain filtering theory. This method had an excellent anti-interference effect and was effective in diagnosing the vibration signal's fault type. Time–frequency analysis provides additional benefits over time domain and frequency domain analysis when processing non-stationary signals. It can also define more fault features that are not accessible in the time domain or frequency domain. However, for non-stationary signals, time domain analysis and frequency domain analysis have several restrictions that might result in the loss of identifying information. In this paper, we combine time–frequency analysis and new deep learning models to use end-to-end methods for rolling bearing fault diagnosis. In order to deal with a large amount of data, there are shortcomings of a slow processing speed, which greatly limits the development of rolling bearing fault diagnosis technology. The one-dimensional bearing signal is first transformed into a two-dimensional image

signal using the generalized S transform, and then this image signal is fed into the Swin Transformer model to identify faults. The Xichu big and small bearing dataset was used to test the model's efficacy, and a rolling bearing fault detection bench was constructed to test the model's generalization capabilities.

Recently, there are still two shortcomings that are difficult to be solved. Firstly, machine learning-based approaches inevitably need to resort to generating discriminative features with the help of signal processing methods, but the manual extraction of features still requires a high level of professional knowledge. Secondly, the features obtained after careful manual extraction and selection according to the requirements of specific tasks are not always effective in the face of unknown working conditions or application scenarios. As the vibration signals in real industry are becoming more and more complicated, and the time–frequency imagery obtained from the existing time–frequency analysis methods is still poorly focused and has serious cross-terms, and its resolution is not enough to fully reflect the frequency and energy distribution information, it is essential to further research the novel time–frequency analysis methods.

In this article, the GST is applied to convert 1D bearing signals into 2D image signals, which are then introduced into a Swin Transformer model for fault identification. Using Case Western Reserve University's large and small bearing datasets to demonstrate the validity of the model, a rolling bearing fault diagnostic bench was built to demonstrate the generalizability of the model.

2. Materials and Methods

2.1. Swin Transformer

Swin Transformer is an upgraded image processing network based on the concept of Vision Transformer with excellent image classification capability. It has four base models, which are Swin_T, Swin_S, Swin_B, and Swin_L. Considering the complexity of the input signals and the calculative cost, Swin_T is chosen as the basic structural model. The architecture of the Swin Transformer (Swin_T) appears in Figure 1.

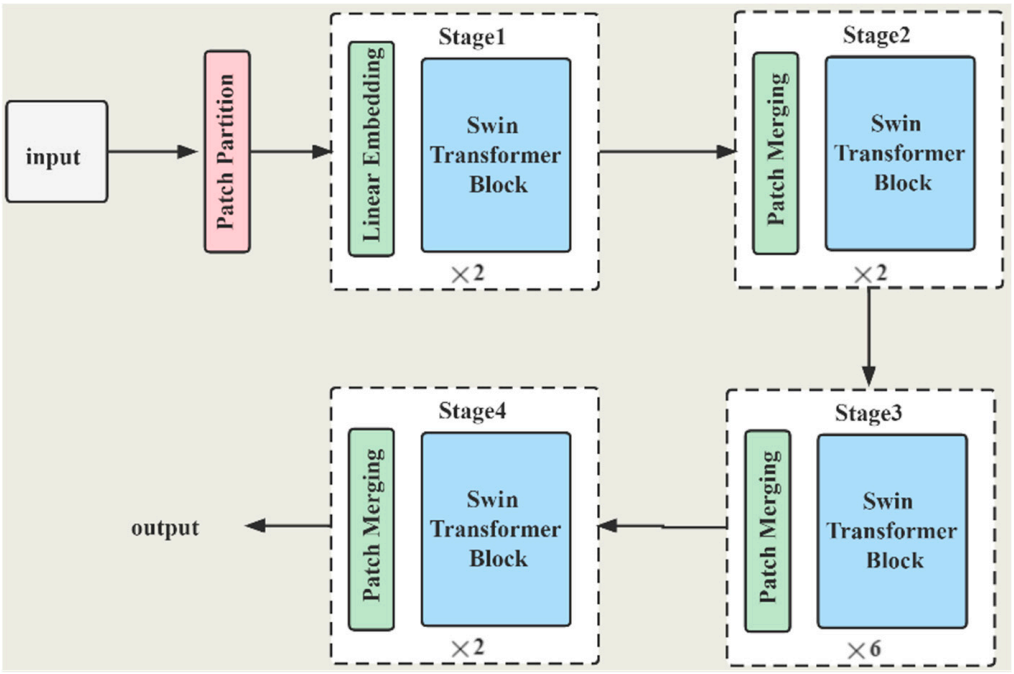


Figure 1. Swin Transformer network structure (Swin-T).

The algorithmic process of the Swin Transformer has the following general steps: take the three-channel RGB image signals as the input, cut the original image by module partition (patch partition) operation, and obtain the non-overlapping image block (patch). After expanding the block, the obtained feature is then input into the linear inlay layer to reduce the dimensionality of the feature. Next, generate a feature block, and input this feature block into the Swin Transformer block to perform feature extraction. After the block splicing (patch merging) operation, the image feature size is cut to half of its original size, and the number of channels is increased to twice their original size. The output after each stage is the input features to the classifier for performing the image classification task.

1. Patch Merging

Patch merging is a down-sampling operation similar to pooling in convolutional neural networks. In each stage block, it is necessary to carry out patch merging, as depicted in the figure; input a $(H \times W \times C)$ feature, the feature points at one position apart are labeled with the same number. After completing the labeling, the labeled feature points are extracted and combined together in the direction of the channel. After splicing, a new feature $((H/2) \times (W/2) \times 4C)$ is formed in the direction of the channel, after the LayerNorm layer to normalize the pixel points, and finally, the feature is linearly processed along the direction of the channel by the all-connected layer, and the feature is transformed into $((H/2) \times (W/2) \times 2C)$ output. The whole patch merging process is to cut the input feature size to half of its original size and increase the number of channels to twice their original size. This operation can save computation and improve network computing efficiency. Moreover, the patch merge does not lose signal and the entire patch merging process is demonstrated in Figure 2.

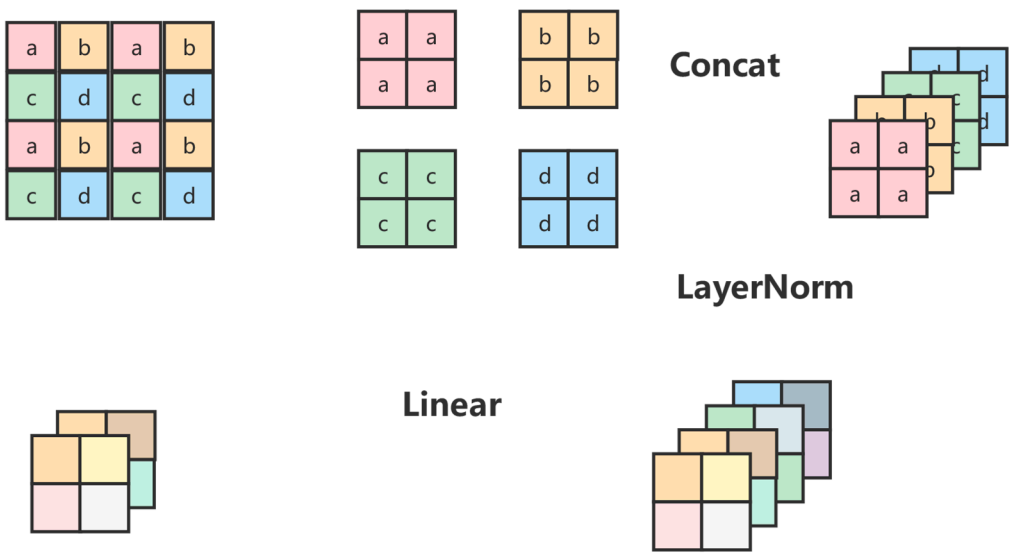


Figure 2. Patch merging process.

2. Swin Transformer Block

As shown in Figure 3, Swin Transformer Block consists of a Windows Multi-head Self-Attention (W-MSA) module and a Shifted Windows Multi-Head Self-Attention (SW-MSA) module. Windows Multi-Head Self-Attention (SW-MSA) module. And two multilayer perceptrons (MLPs) are connected after it, and the nonlinear ability of the whole network is enhanced by (GELU) activation function between them, which in turn improves the fitting ability of the whole network. And there is a normalization layer (LN) before W-MSA, SW-MSA and MLP, and a residual connection is connected after each module.

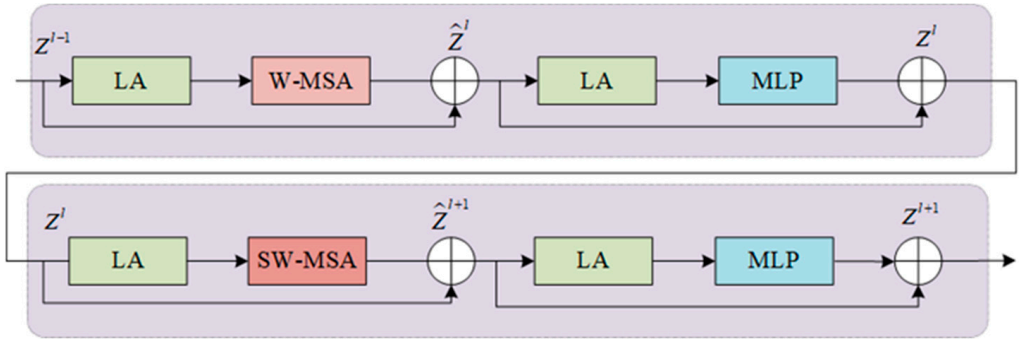
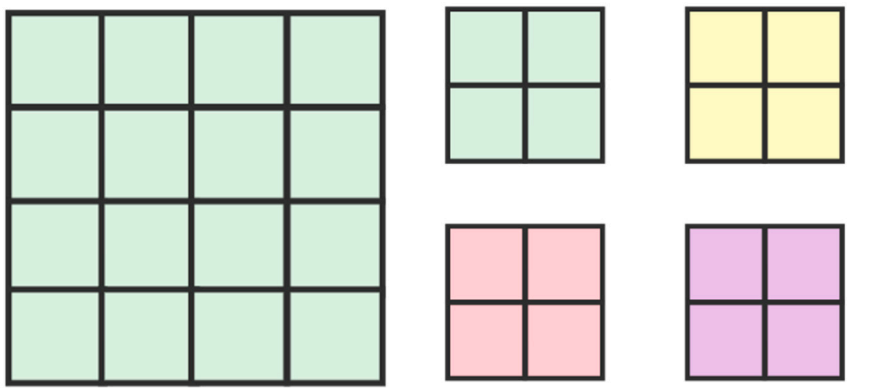


Figure 3. Swin Transformer block module structure.

- Multi-head self-attention mechanism

With the development in neural networks, Ashish Vaswani et al. [30] proposed a network structure based entirely on the attention mechanism in 2017, which significantly reduces the time required for network training. The self-attention layer can compute the weights of local features by localizing them into global features, and then the global features can be obtained from the sum of the local features. Nevertheless, due to the high computational complexity of MSA, it is necessary to introduce a window module in the Swin Transformer to reduce the computational effort of MSA. Figure 4 illustrates the differences between MSA and window-based self-attention approaches (windowed multiple self-attention, W-MSA).



(**a**) Multi-headed self-attention (**b**) Windows multi-head self-attention

Figure 4. Differences between the two attention mechanisms.

As Figure 4a shows, the weights for each pixel point in the feature are computed for the ordinary multi-attention mechanism as follows in Equation (1):

$$Q(MSA) = 4hwC^2 + 2(hw)^2C \tag{1}$$

Figure 4b represents the multi-attention mechanism after the addition of windows, where the features are first divided into window spaces by windows, and then the weights of each window space are calculated in a single degree, as follows in Equation (2):

$$Q(W-MSA) = 4hwC^2 + 2MhwC \tag{2}$$

where Q denotes the computational complexity of the multi-attention mechanism, h denotes the height of the feature, w denotes the width of the feature, C denotes the number of channels of the feature, and M denotes the size of the window. Since the size of the window

is much smaller than the size of the feature, the computational complexity of using the window is much lower than that of the non-windowed computation, which can effectively improve the computational speed of the model.

- Offset window self-attention mechanism (SW-MSA)

Adding windows may greatly speed up the computation of weights in the feature map, but doing so breaks the connections between windows and reduces the correlation between local and global data. To address this problem, the Swin Transformer performs weight adjustment with the SW-MSA module (also known as Shifted W-MSA). Figure 5 depicts the shifted feature block.

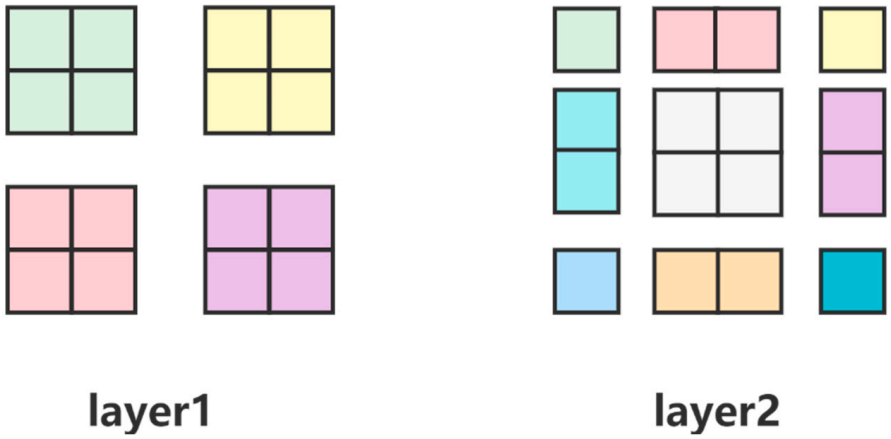


Figure 5. Output module after two attention calculations in Swin Transformer block.

As shown in the figure above, the features output from layer 1 circularly shift the window to the upper left $(\frac{M}{2}, \frac{M}{2})$ pixels to obtain the output from layer 2. When calculating the weights, nine windows of different sizes are spliced into four feature blocks, with the grey feature block in the middle remaining unchanged as a separate window; the feature blocks on the four corners are spliced into a new computational window: the feature blocks on the left, right, top, and bottom are spliced into a single window. At this time, each window contains elements of other windows with interrelated weights, and MSA calculations are performed fast for each feature separately, thus reducing the amount of computation and correlating the local features with the global ones. Figure 6 shows the schematic diagram of pixel shifting in the computation.

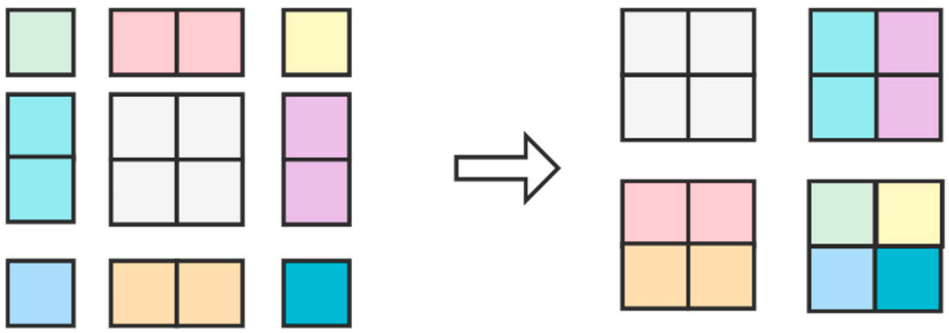


Figure 6. Schematic diagram of SW-MSA window calculation.

2.2. Multi-Classification Algorithm

Softmax classifiers are commonly used to deal with multiple classification problems. In this article, we improve the multi-classification algorithm based on the Softmax function.

The mathematical expression of the Softmax function [31–36] can be expressed as illustrated in Equation (3) below

$$h_\omega(x^{(i)}) = \begin{bmatrix} \varphi(y^{(i)} = 1 \middle| x^{(i)} : \omega \\ \varphi(y^{(i)} = 2 \middle| x^{(i)} : \omega \\ \varphi(y^{(i)} = k \middle| x^{(i)} : \omega \end{bmatrix} = \frac{1}{\sum_{j=1}^{p} e^{\omega_{\mathrm{m}}{}^T x(i)}} \begin{bmatrix} e^{\omega_1^T x^{(i)}} \\ e^{\omega_1^T x^{(i)}} \\ : \\ e^{\omega_p^T x^{(i)}} \end{bmatrix} \tag{3}$$

Among them, $h_\omega(x^{(i)})$ is the output value after being processed and normalized by the Softmax classifier, $\{\omega_1, \omega_2, \cdots\cdots, \omega_p\}$ is the model parameter of the Softmax classifier, and $\frac{1}{\sum_{j=1}^{p} e^{\omega_{\mathrm{m}}{}^T x(i)}}$ is to normalize the output value and reduce the output value of the model to between 0 and 1. The ith sample in the input signal is denoted by $x^{(i)}$, and the probability that this sample's classification prediction belongs to k is $\varphi\left(y^{(i)} = 1 \middle| x^{(i)} : \omega\right)$.

The Softmax classifier optimizes the model by using the cross-entropy loss function [25], which is computed as follows:

$$Loss = \frac{1}{\mathrm{n}} \sum_{i=1}^{m} \left(-y^{(i)} \log\left(h_\omega\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right) \log\left(1 - h_\omega\left(x^{(i)}\right)\right)\right) \tag{4}$$

where n is the number of input samples; m is the number of categories; $y^{(i)}$ is the label corresponding to the sample; when using the Softmax classifier, the parameter is usually expressed in a matrix, and the expression of the matrix is as follows:

$$\omega = \begin{bmatrix} \omega_1^T \\ \omega_2^T \\ : \\ \omega_{\mathrm{p}}^T \end{bmatrix} \tag{5}$$

Substituting the parameters into the loss function, the loss function expression is as follows:

$$Loss = -\frac{1}{m}\left(\sum_{i=1}^{n} \sum_{i=1}^{k} \psi[y^{(i)} = j] \log \frac{e^{\omega_j^T x^{(i)}}}{\sum_{j=1}^{k} e^{\omega j^T x(i)}}\right. \tag{6}$$

where $\psi(\cdot)$ is 1 when the predicted result is the same as the real value, and 0 when it is different.

Due to the existence of countless vectors in the vector space, when any two of them with an appropriately large span are given a category meaning, the sample points will be clustered towards the corresponding category vectors, and the difference in the distance values will be greater when the distance calculation is performed on the samples using the standard base coordinates; thus, the category differences will be more obvious. Therefore, the aggregation of similar features in the mapping space can be enhanced by increasing the distance between different feature clusters, which can successfully improve the accuracy of classification. The similarity between samples and parameters is represented by the cosine distance, which is used together with the feature amplitude to determine the weight of the feature vector that belongs to the final class. The modified weight formula is as follows:

$$W_c x = ||W_c|| \cdot \|x\| \cos(\omega_c) \tag{7}$$

where c is the classification category, and W_C is the weight index that divides features into categories in the classifier. For a feature, the cosine distance between the value of the feature vector and the angle between the weight vectors will determine the final classification result.

The Softmax classification impact is the best after updating the Softmax settings to obtain the ideal loss amount. The following formula is obtained by substituting the modified weight parameters into the loss function.

$$Loss_i = -\log\left(\frac{e^{\|W_{y_i}\|\cdot\|x_i\|\cos(\omega_{y_i})}}{\sum_{j=1}^{k} e^{\|W_j\|\cdot\|x_i\|\cos(\omega_j)}}\right) \tag{8}$$

Among them, W_{y_i} represents the column of the W weight matrix, x_i, y_i are the feature and label of the instance, ω_{y_i} represents the angle between W_{y_i} and x_i, and W_j refers to all weights of class j.

2.3. Troubleshooting Process

In this research, a fault classification model based on time–frequency pictures is developed as a way to effectively raise the accuracy of fault detection. The time–frequency characteristics of the vibration signal will be converted to a time–frequency picture, and then the time–frequency image signals will be input into the Swin Transformer model to carry out the automatic extraction of the features and select the Swin-T model with a smaller specification according to the characteristics of the vibration signals. Then, the extracted features are processed by linear full connectivity, and finally, the software classifier of the reformer is used for the multi-classification prediction of faults. The construction of the classification model built using the time–frequency diagram and Swin Transformer is depicted in Figure 7. Figure 8 depicts the process for diagnosing a rolling bearing failure using a time–frequency graph and a Swin Transformer.

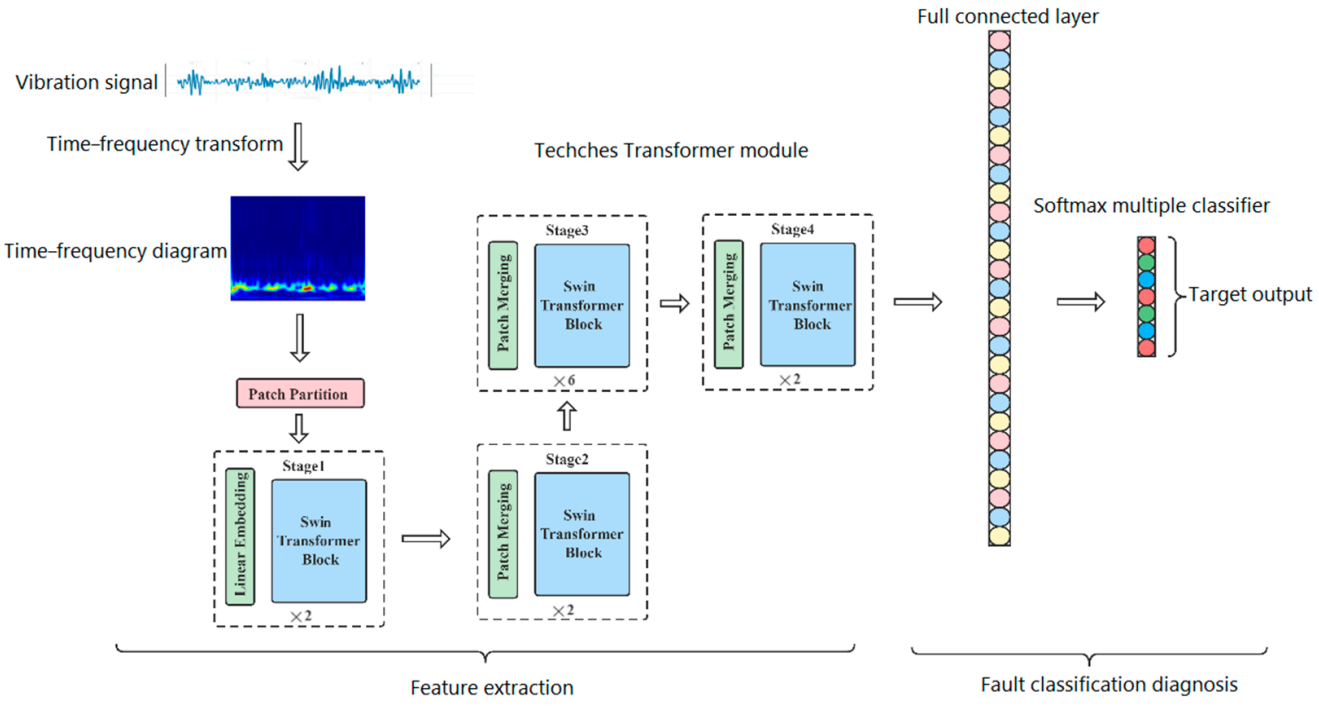


Figure 7. Classification model structure based on time–frequency diagram and Swin Transformer.

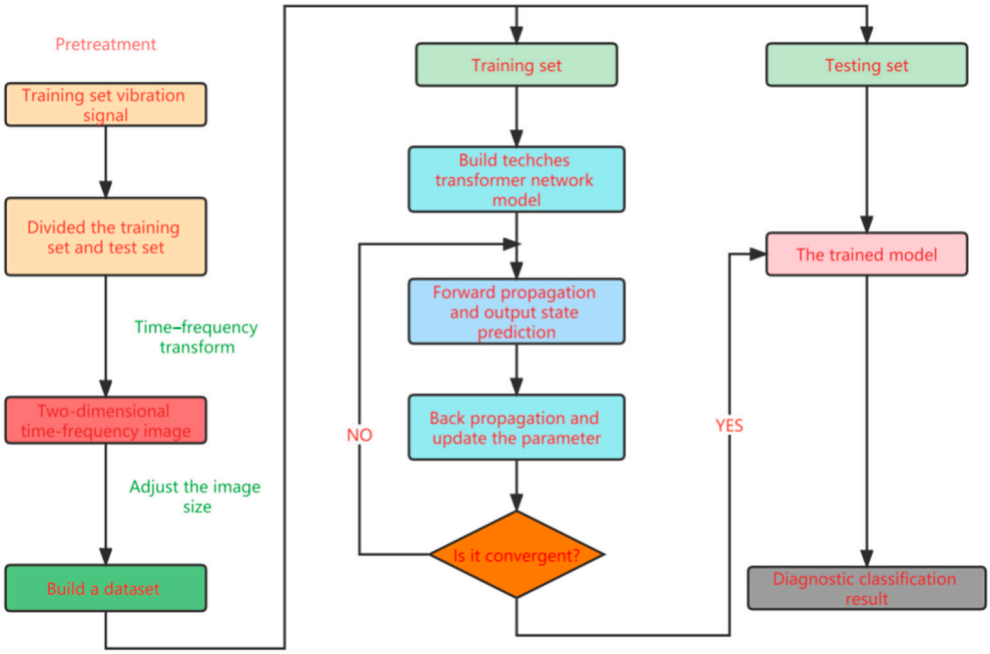


Figure 8. Experimental flow chart.

This experiment uses the bearing dataset from the experimental dataset of Case Western Reserve University, and the bearing failure data under different working conditions at 48 kHz are selected. The experiment includes three steps: data pretreatment, model training, and its prediction. The following are the details regarding the experiment:

(1) The acquired vibration signal data are divided into the training set and test set in the ratio of 8:2, the 1D vibration signal is transformed into a 2D time–frequency picture, the size of the image output is modified, and the image is labeled with fault categories for subsequent training and testing

(2) The model of the Swin Transformer network is constructed, its hyperparameters (such as the number of iterations, learning rate, decay rate, batch size, etc.) are configured, its weights and biases are initialized at random, and the training data are fed into it. Using block splicing (patch merging), the window self-attention mechanism (W-MSA), the offset window self-attention (W-MSA), the Shift Window Self-Attention (SW-MSA), and multilayer perceptron (MLP) are used to extract image features, the predicted value of the output state by forward propagation is calculated, and the parameters of the network are updated by back propagation to make the error between the labeled value and the predicted value smaller until the loss function reaches convergence, and then the parameters are saved to complete the training.

(3) The test set is imported into the trained model, the classification diagnostics are run, and the effectiveness of its classification recognition is checked.

3. Results

3.1. Experimental Data

As Figure 9 shows, this experimental bench is mainly composed of the following parts: on the left side is a motor with a power of 1.5 KW, in the middle is a torque transducer, and on the right side is a power tester. In the data collection experiment, a single-point fault bearing is used, and EDM is used to manufacture faulty bearings with fault diameters of 0.007 inches, 0.014 inches, 0.021 inches, 0.028 inches, and 0.040 inches. On the other hand, the datasets used for model training in this article are from Case Western Reserve University's bearing dataset, which is commonly used to measure the performance of

rolling bearing defect detection models. The bearings on the bearing test rig were replaced with bearings of different failure sizes, and the vibration signal data were captured by using accelerometers mounted on the test rig and stored at both 12 kHz and 48 kHz sampling frequencies throughout the data acquisition experiment. Datasets under four distinct operating situations, namely 0 hp, 1 hp, 2 hp, and 3 hp loads, of which 0.007 inch, 0.014 inch, and 0.021 inch are the same bearing type, are collected for this experiment. This experiment selects the above three kinds of fault size of the bearing, which includes the inner ring failure, outer ring failure, and rolling body failure; together with the normal operation data of the bearing, the whole dataset can be divided into 10 kinds of operating states, and according to the different loads are divided into three datasets, A, B, and C. Each dataset is taken as 100 samples for each kind of operating state and randomly disrupted according to 9:1. The training set and test set are divided, and dataset D is a multi-operating condition dataset containing three kinds of loads, and the specific dataset division is depicted in Table 1 below:

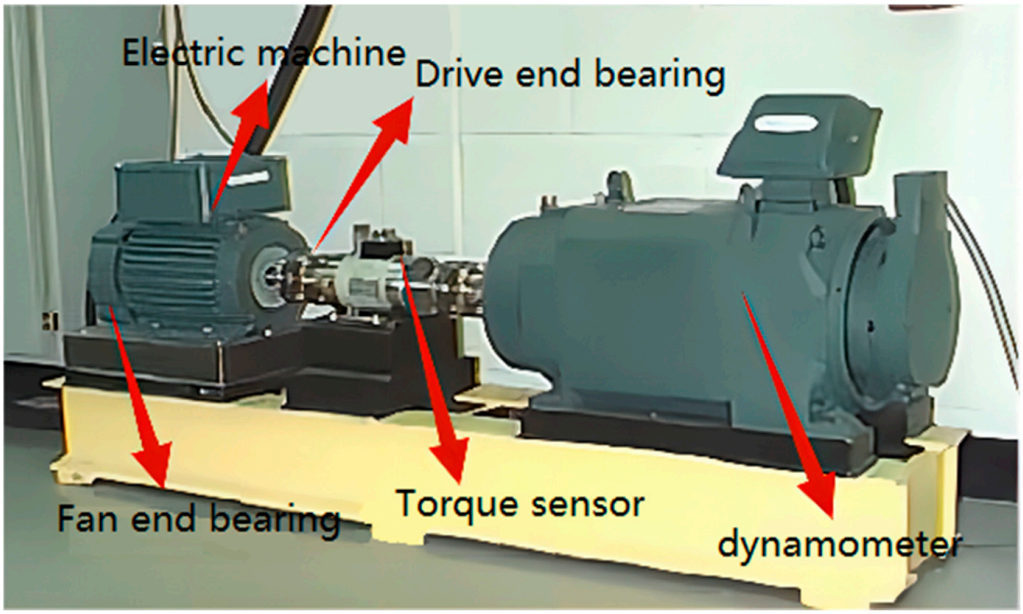


Figure 9. Case Western Reserve University bearing test bench.

Table 1. Dataset division.

Damage Position		Normal	Inner Ring			Outer Ring			Rolling Element			Load
Label		0	1	2	3	4	5	6	7	8	9	
Damage diameter		0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021	
A	Training	900	900	900	900	900	900	900	900	900	900	0
	Testing	100	100	100	100	100	100	100	100	100	100	
B	Training	900	900	900	900	900	900	900	900	900	900	1
	Testing	100	100	100	100	100	100	100	100	100	100	
C	Training	900	900	900	900	900	900	900	900	900	900	2
	Testing	100	100	100	100	100	100	100	100	100	100	
D	Training	2700	2700	2700	2700	2700	2700	2700	2700	400	2700	0~2
	Testing	300	300	300	300	300	300	300	300	300	300	

3.2. Comparison and Analysis of Time–Frequency Analysis Approaches

In an effort to acquire models with higher diagnostic accuracy, we employed a comparative validation approach utilizing the Case Western Reserve University bearing dataset to achieve the best-performing time–frequency analysis approach. In the following, the vibration signals will be analyzed and corresponding time–frequency images will be generated using the linear STFT, WT, and GST, and the nonlinear WVD, respectively.

(i) Short-time Fourier analysis

In short-time Fourier analysis, the choice of window length will directly affect the resolution of the time–frequency diagram; when the window exceeds the appropriate size, the time resolution will deteriorate, or even become a Fourier transform, which will result in a loss of time scale information; when the window is smaller than the appropriate size, the frequency resolution will also deteriorate, which will cause the loss of part of the frequency information. Here, we choose 64, 128, 256, and 512 window size lengths for the STFT, and the obtained time–frequency diagram is depicted in Figure 10 below.

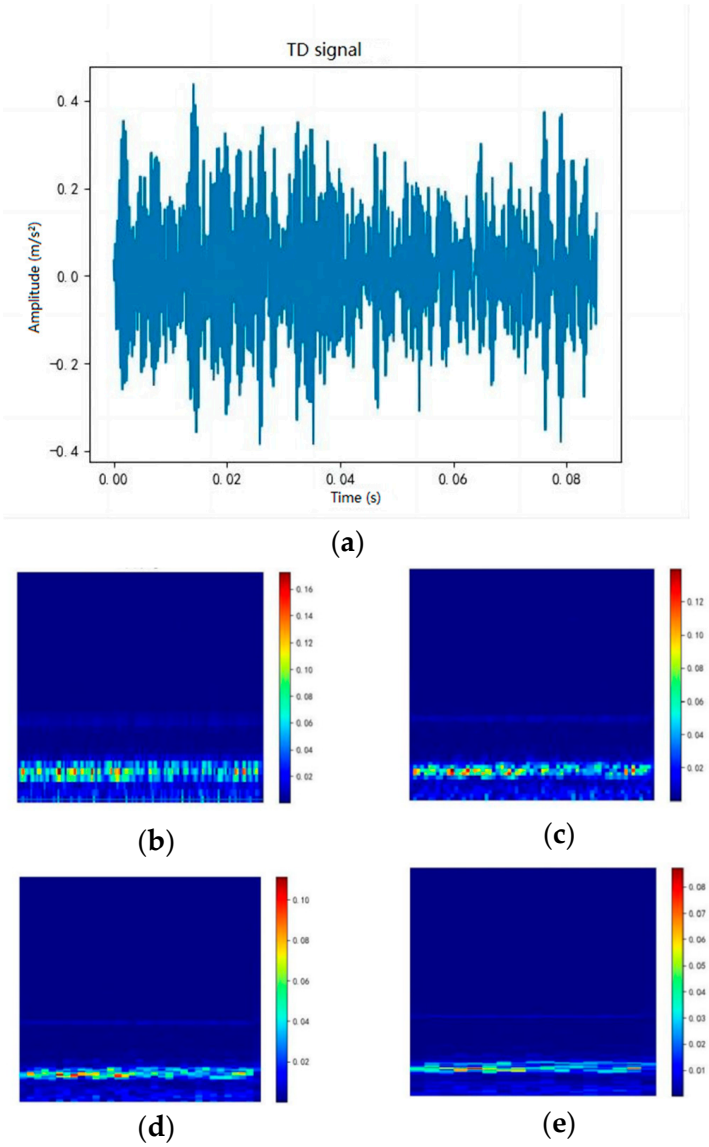


Figure 10. Short-time Fourier transform time–frequency diagram. (**a**) Time domain waveform of bearing rolling element failure. (**b**) Window size 64. (**c**) Window size 128. (**d**) Window size 256. (**e**)Window size 512.

From the above four time–frequency diagrams, it has been found that the frequency resolution of the time–frequency diagrams increases gradually when the window size increases, and when the window size reaches 512, the frequency resolution is the best, and the time resolution is also more obvious.

(ii) Continuous wavelet transform

In wavelet analysis, due to the choice of wavelet basis having a greater impact on the effect of the time–frequency diagram, different wavelet bases are chosen to deal with different signals, and so the effect is different; in order to obtain the best wavelet analysis effect, we intercepted the Case Western Reserve University bearing failure of the subsequent analysis of the dataset, the Grid wavelet basis, Hill's wavelet basis, and the complex Mossy's wavelet basis for the signal to carry out a continuous wavelet transform, and the time–frequency diagrams are shown in Figure 11 below.

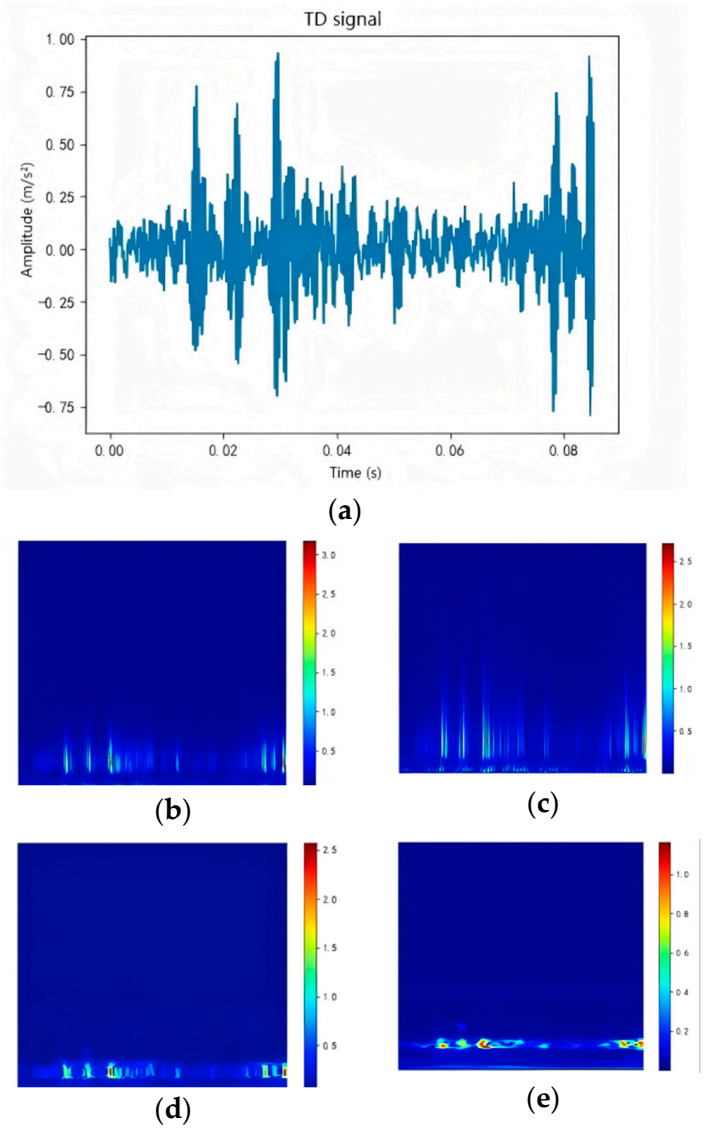


Figure 11. Continuous wavelet transform time–frequency diagram. (**a**) Time domain waveform of bearing rolling element failure. (**b**) Haar wavelet basis. (**c**) Mesh wavelet basis. (**d**) Shan wavelet basis. (**e**) Complex Morlet wavelet basis.

(iii) Generalized S transform

The generalized S transform is developed from the S transform, which is very similar to the fundamental wavelet transform of the Morlet wavelet. However, only the Gaussian

function part of the S transform changes the size of the window function with the frequency and performs translational changes in the time dimension, and the chi-square harmonic waveform part only performs telescopic changes, which makes the S transform defective for the feature representation of non-smooth signals. So, the parameters a and b are added to adjust the shape of the Gaussian window to optimize its resolution in the time and frequency domains, as shown in the following equation:

$$GST(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{a|f|^b}{\sqrt{2\pi}} e^{[-\frac{1}{2}a^2 f^{2b}(t-\tau)^2]} e^{(-i2\pi ft)} dt \quad (9)$$

The generalized S transform combines the strengths of the STFT and the WT, effectively avoiding the limitation of the short-time Fourier window size and the problem of selecting the wavelet basis function, which is very advantageous in the domain of time–frequency analysis. The time–frequency diagram obtained from the GST is illustrated in Figure 12.

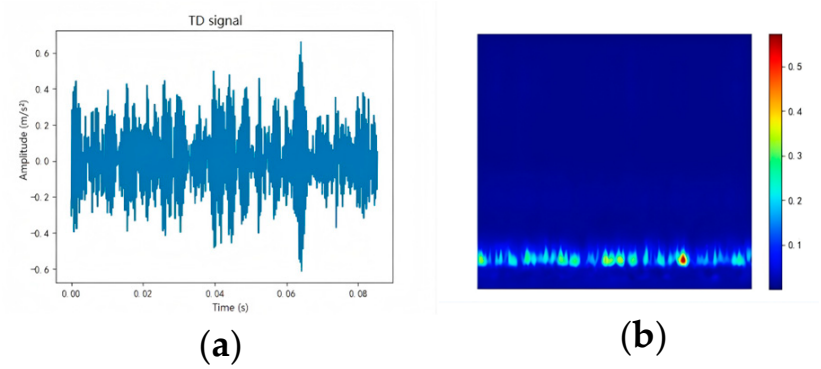


Figure 12. Generalized S transform. (**a**) Time domain waveform of bearing rolling element failure. (**b**) Generalized S-transform time-frequency diagram.

The time–frequency analysis of the Case Western Reserve University data by the GST gives time–frequency maps with clear time–frequency resolution, and the GST can be better characterized by the characteristic information of vibration and its energy characteristics are clearer.

(iv) Wigner–Ville distribution

From the time–frequency plot of the WVD in Figure 13 above, it can be noticed that the WVD is less efficient in analyzing the bearing vibration signals, with a lower time–frequency resolution and pseudo-distribution of superimposed interference.

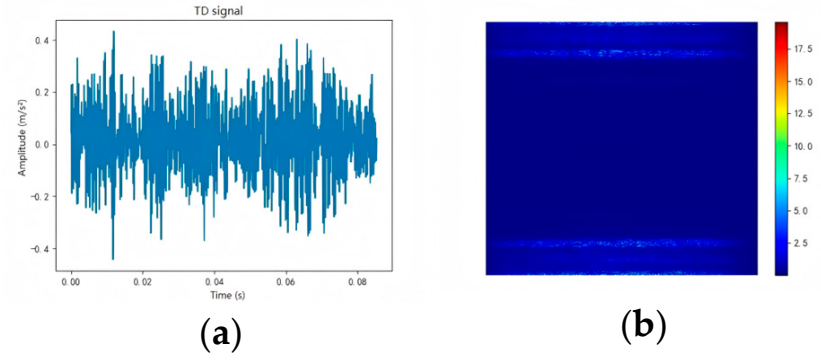


Figure 13. Wigner–Ville Distribution.(**a**) Time domain waveform of bearing rolling element failure. (**b**)Wigner-Ville Distribution.

3.3. Data Preprocessing

In this experiment, firstly, the STFT, the CWT, the GST, and the WVD are compared. Figure 14 demonstrates the acquisition capability of the time–frequency resolution.

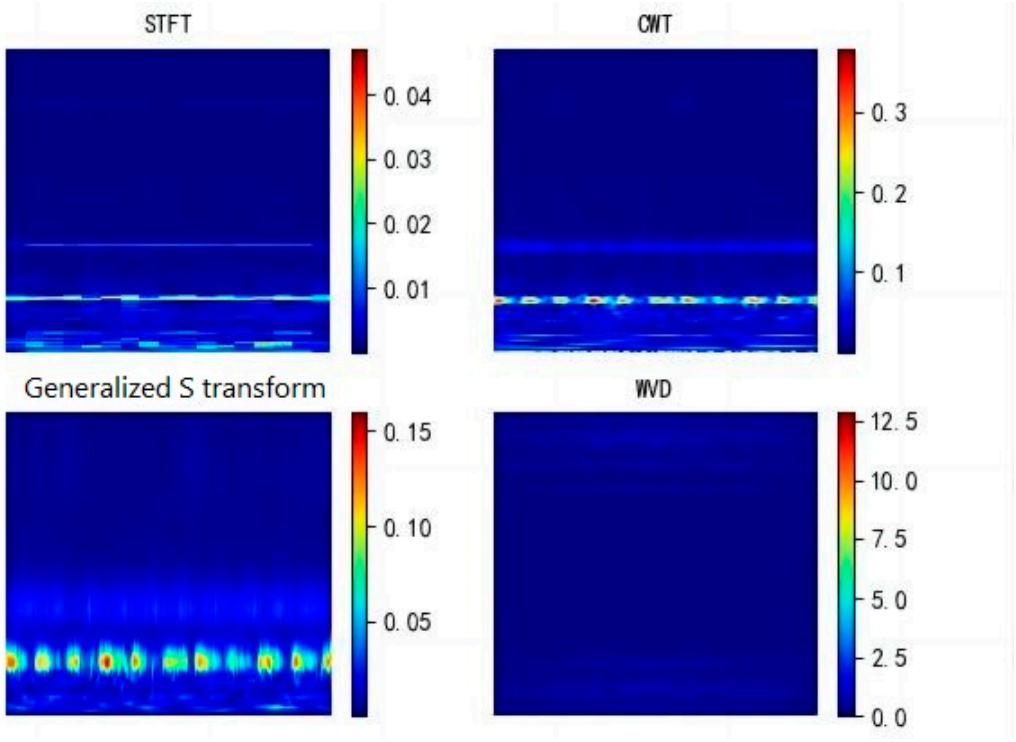


Figure 14. Comparison of four kinds of time–frequency diagrams.

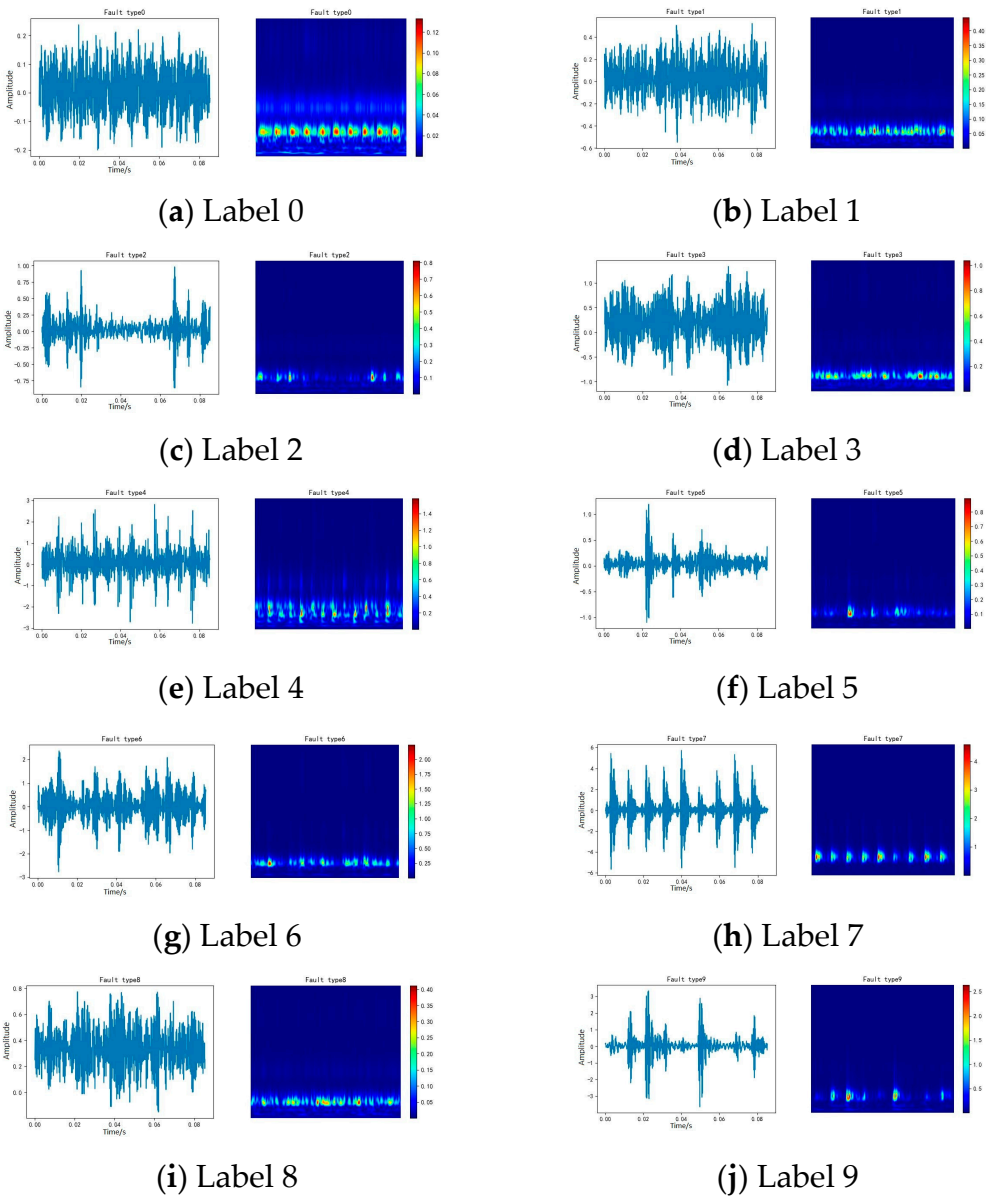
Figure 14 compares the four time–frequency analysis approaches, from which a conclusion can be drawn that the time–frequency resolution of the STFT is low, while the wavelet analysis approach can more obviously be a time–frequency resolution, which can be seen in the ability of the generalized S transform characteristics. Compared with the STFT and WT, the GST has more obvious energy characteristics and more prominent signal characteristics, while the WVD has a lower time–frequency resolution and still has pseudo-distributed signal interference.

In the final analysis, 4096 consecutive sample points were randomly intercepted from the original vibration data to obtain a new sample. Under the same conditions, 100 samples were selected for each running state and the remaining 1000 samples were randomly partitioned into a training set and test set in the ratio of 9:1. The time domain signals of the training and test sets were converted into 2D image signals utilizing the continuous wavelet transform approach, and the output size of the selected images was 64×64, as shown in Figure 15.

3.4. Experimental Verification

The experiment is based on a Python-based PyTorch deep learning platform. The computer operating system is Windows 10, the graphics card is NVIDIA Quadro RTX5000, the edition of PyTorch is 1.11.0, and the running memory is 128 G. The following hyperparameter values are input to the network (Table 2).

The model has been trained and validated utilizing the above four datasets (A, B, C, and D), and its loss functions and accuracies are depicted in Figures 16 and 17, respectively.



(**a**) Label 0 (**b**) Label 1

(**c**) Label 2 (**d**) Label 3

(**e**) Label 4 (**f**) Label 5

(**g**) Label 6 (**h**) Label 7

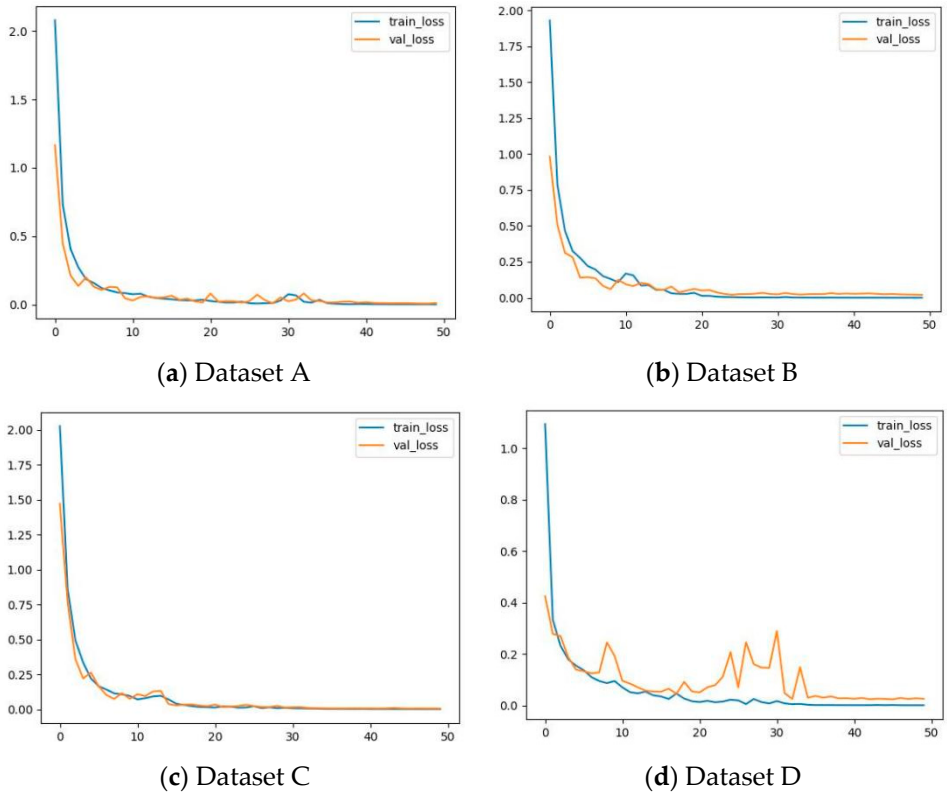
(**i**) Label 8 (**j**) Label 9

Figure 15. Graph of generalized S transform results.

Table 2. Hyperparameter settings.

Image Size	**64 × 64 × 3**
Batch_size	8
Learning_rate	10^{-3}
Weight_decay	10^{-5}
Epoch	50
Optimizer	SGD

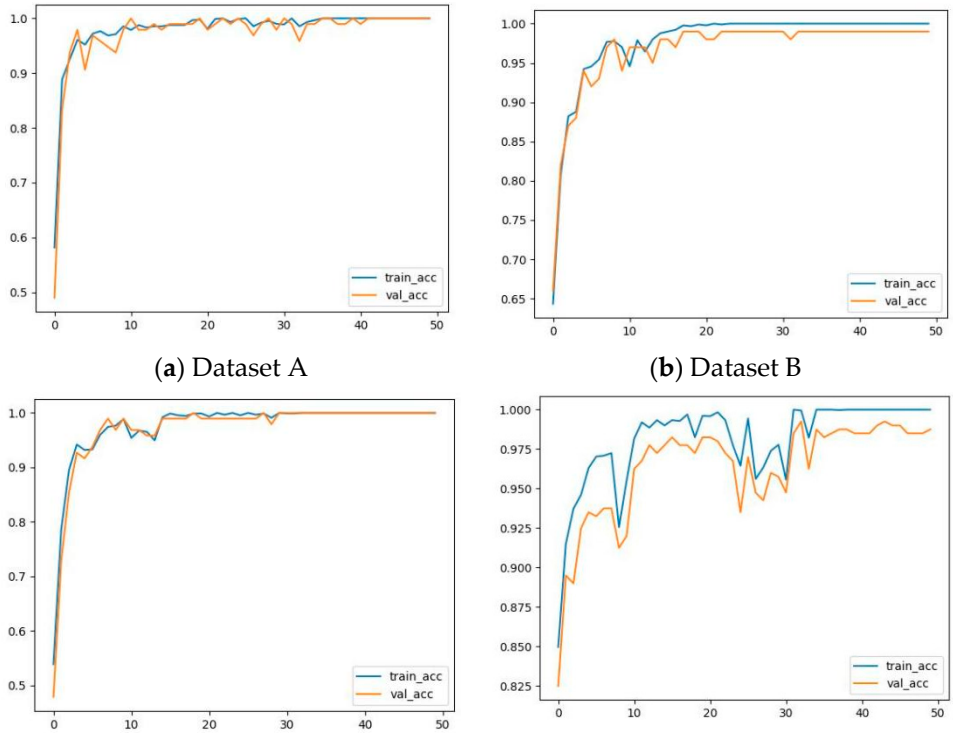
The loss function curve shows that after 50 iterations in the single condition, the loss value of the training set and the loss value of the test set are basically equal to 0, and the loss value of the training set and the test set in the multi-condition is close to 0. From the accuracy curve, it can be seen that the final accuracy of the dataset in the single condition is stable at 1, and that the accuracy of the dataset in the multi-condition is close to 1. The model achieves the accurate prediction of the fault type and performs slightly better than the single-condition dataset.



(**a**) Dataset A (**b**) Dataset B

(**c**) Dataset C (**d**) Dataset D

Figure 16. Transformation curve of loss value with the number of iterations.



(**a**) Dataset A (**b**) Dataset B

(**c**) Dataset C (**d**) Dataset D

Figure 17. Accuracy variation curve with the number of iterations.

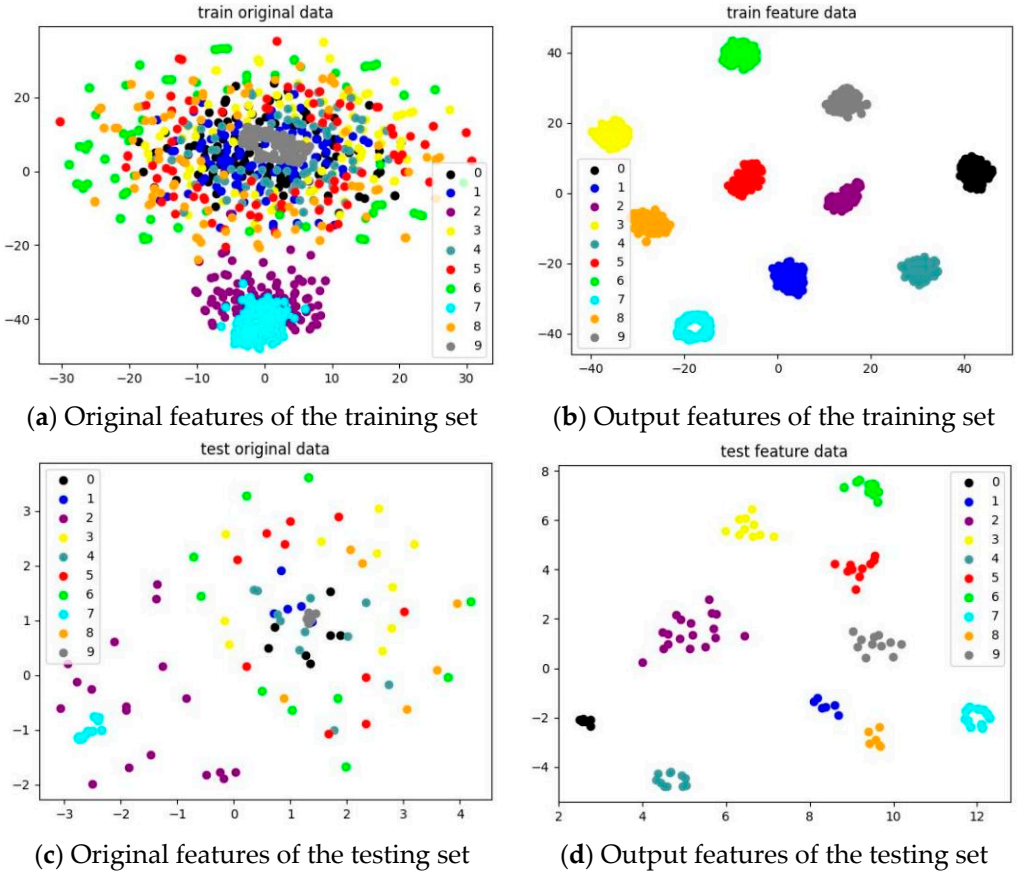
In order to specify the accuracy of the method proposed in this article, the performance of the model is illustrated using Accuracy, Precision, Recall, and F1 Score, and then

10 repetitions of the experiment are performed for each dataset and the average is taken and the results are shown in Table 3 below.

Table 3. Model performance evaluation table.

	Accuracy	Precision	Recall Rate	F1 Score
Data A	100	100	100	100
Data B	99.25	99.48	99.32	99.82
Data C	100	100	100	100
Data D	99.37	99.45	99.26	99.28

Table 3 shows that the model can successfully recognize both single-condition data and multiple-condition data with a ten-time average accuracy of more than 99%. The prediction process of the Swin Transformer model can be visualized and analyzed by the T-SNE downscaling technique, which is a common downscaling method for mapping high-dimensional data into 2D or 3D data, and then visualized and analyzed, as shown in Figure 18, which displays the visualization result graph of dataset A.



(**a**) Original features of the training set (**b**) Output features of the training set

(**c**) Original features of the testing set (**d**) Output features of the testing set

Figure 18. Feature visualization.

The initial features of the training set are shown in Figure 18a. The distribution of the original traits is chaotic, as can be seen in the image, making it difficult to categorize them precisely. Figure 18b displays the characteristics after 50 training iterations. It is evident that the characteristics of different types of bearing failures are effectively differentiated. The model can also successfully delineate distinct borders between various characteristics in the test set. These all demonstrate that the time–frequency graph and Swin Transformer-based

rolling bearing fault detection model has outstanding feature learning capacity and can recognize various bearing failure kinds.

In this article, we choose to utilize the more popular machine learning techniques and deep learning models for comparative validation and analysis to further validate the effectiveness of the rolling bearing defect detection model based on time–frequency diagrams and the Swin Transformer suggested in this article. The dataset used in the follow-up approach is consistent with the dataset provided in this study. Table 4 below presents the classification accuracy rate of the various diagnostic techniques for different datasets. The method of this paper is denoted as S+ST, which can show the capability of each model more intuitively. The consequences of the research are depicted in Figure 19. As can be viewed from the figure, the approach suggested in this article outperforms several common diagnostic approaches on all four datasets. This comparative experiment further demonstrates that the Swin Transformer model has a powerful feature learning capability and is able to accurately perform the task of fault classification and identification in the domain of fault diagnosis.

Table 4. Accuracy rates of different experimental methods.

Diagnostic Methods	Accuracy			
	Data A	Data B	Data C	Data D
SVM	82.56	84.58	86.93	80.29
CNN	88.27	84.35	86.95	81.26
LSTM	87.53	88.40	84.58	86.53
WT + CNN	93.65	92.14	95.58	85.27
STFT + SVM	94.47	92.33	93.08	89.23
CNN + LSTM	96.59	98.24	97.28	92.05
S + ST	100	99.25	100	99.37

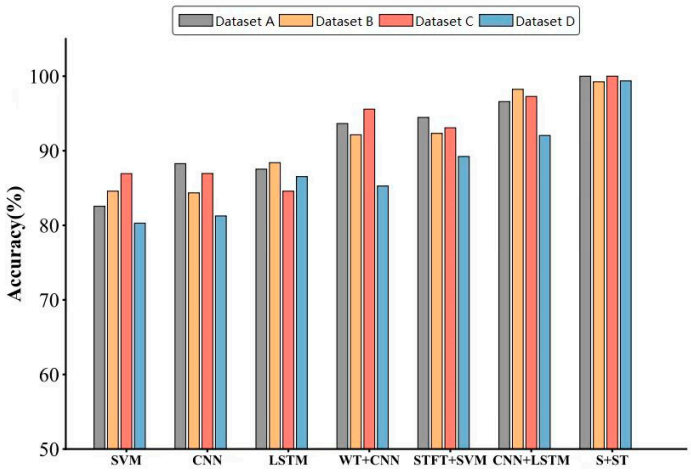
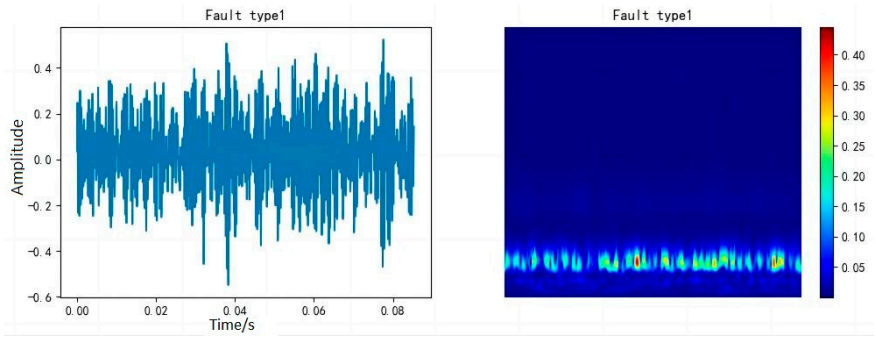


Figure 19. Diagnostic accuracy of different models.

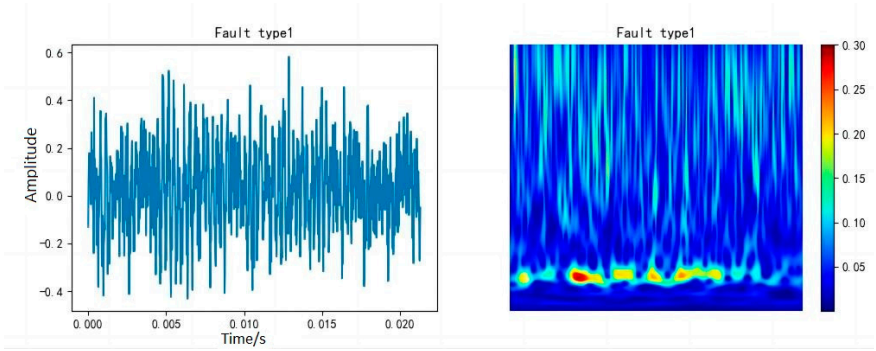
4. Analysis of the Impact of Noise on Model Performance

In practical industrial applications, the presence of noise often influences the diagnostic accuracy rate of the model. In this research, experimental validation will be carried out by introducing white noise into the vibration signal to test the model's ability to resist noise. As the capability of white noise to hide signal features varies with its intensity, therefore, this paper will test the model's immunity to noise by using white noise with diverse

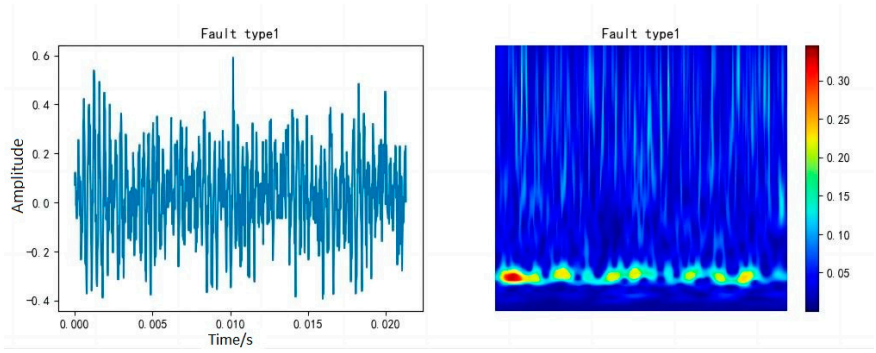
signal-to-noise rates to approximate the true noise intensity in various cases. Choosing the added noise signal-to-noise ratios of 2 db, 4 db, and 8 db, respectively, and taking the 0-loaded dataset A as the original vibration signal dataset, the time domain signals have been analyzed by time–frequency analysis after the addition of the noise, as illustrated in Figure 20 below.



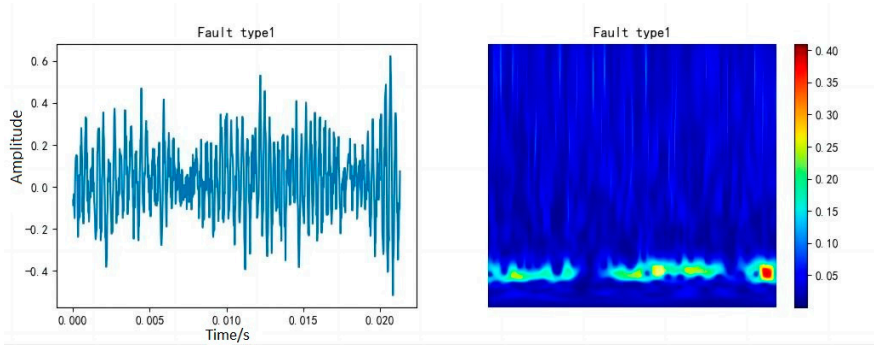
(**a**) 0 db



(**b**) 2 db



(**c**) 4 db



(**d**) 8 db

Figure 20. Time–frequency diagram after adding white noise.

Figure 20 shows the time–frequency pattern obtained by the continuous wavelet transform of the original vibration signal labeled 1 in dataset A, where the signal-to-noise ratios of the white noise interference signals are 2 dB, 4 dB, and 8 dB, respectively. Signals from the three noise environments are input into the network model for training and validating the trained model by using the test set, and the confusion matrix is used to evaluate the performance of the model in the different environments, and the diagnostic performance is evaluated using the confusion matrices of the multiple noise environments. The test set confusion matrices are shown in Figure 21 with noise additions of 0 dB, 2 dB, 4 dB, and 8 dB. The test set with no noise addition and a noise addition of 8 dB has the highest accuracy of 100%. The accuracy rate for the test set with 4 dB of noise added is 98.95%, and the test set with 2 dB of added noise has an accuracy of 96%. It is clear that the ability to adapt to noise disturbances has been improved.

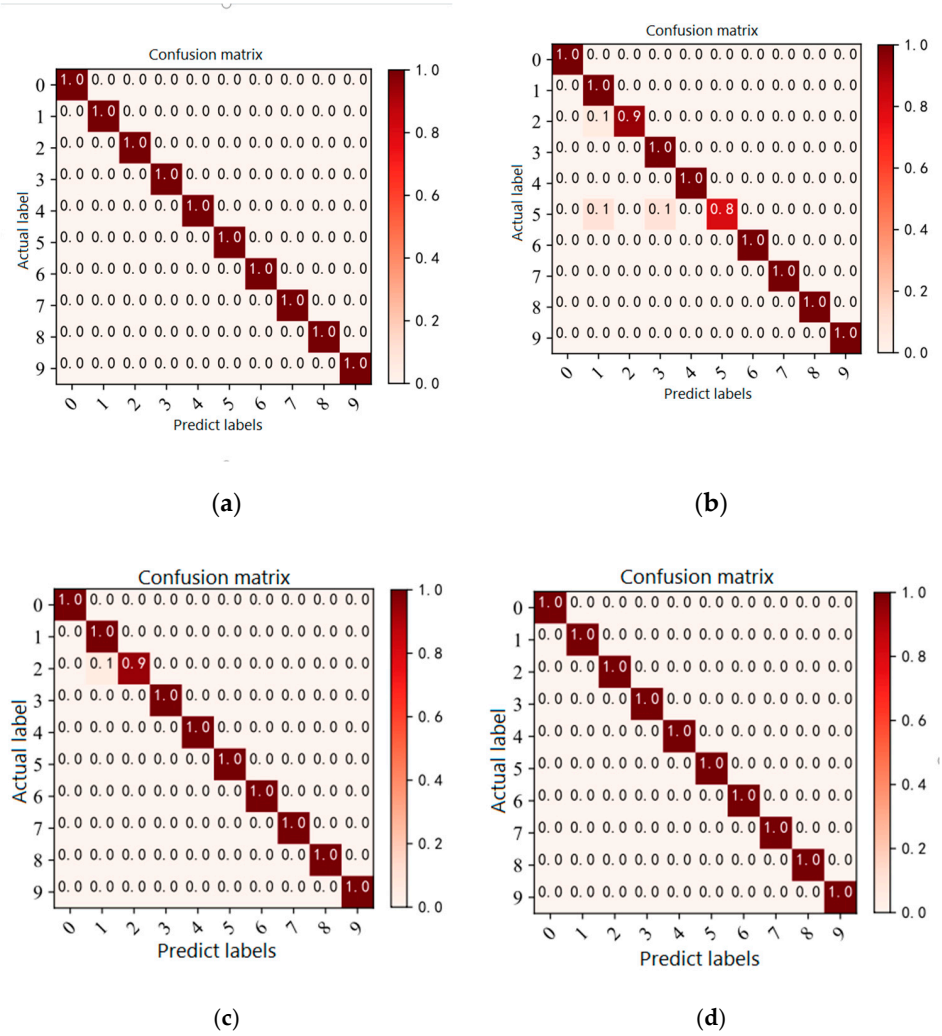


Figure 21. Confusion matrix under different white noises: (**a**) 0 db, (**b**) 2 db, (**c**) 4 db, and (**d**) 8 db.

The experimental results show that the rolling bearing problem detection technique based on the time–frequency diagram and Swin Transformer suggested in this article has good noise robustness. The accuracy is most stable under the influence of various disturbances compared to other methods, as shown in Table 5 below. The single recognition accuracy under various noise disturbances is higher than the other three widely used techniques, which indicates the strong noise robustness of the method. As shown in Figure 22, we depict the comparative results of several methods in a histogram to visualize the noise immunity of this method.

Table 5. Diagnostic accuracy of different models under noise interference.

	CNN	MLP	LSTM	S + ST
0 db	88.43	65.59	87.82	100
2 db	92.01	74.06	89.28	96.00
4 db	93.07	79.25	93.06	98.95
8 db	95.02	85.06	95.35	100
Average	92.1325	75.99	91.3775	98.7375

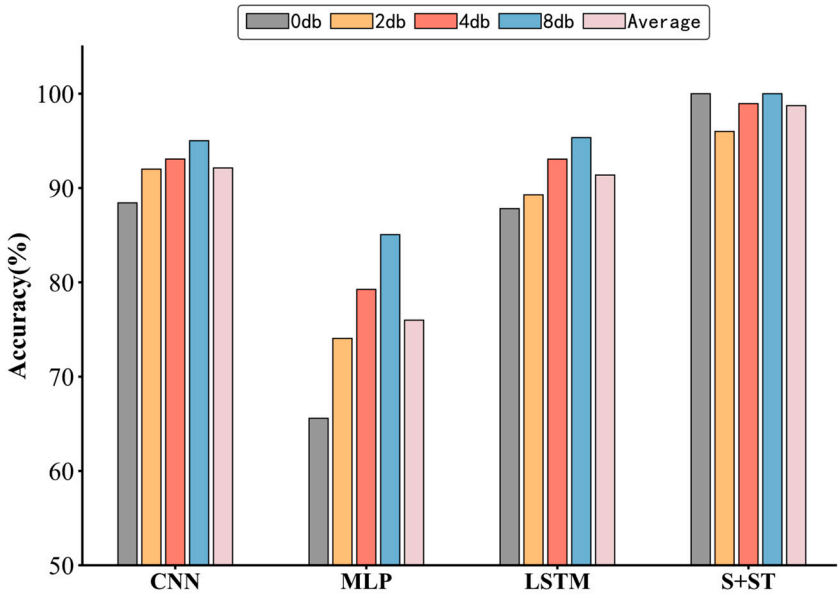


Figure 22. Diagnostic accuracy of different models under noise interference.

5. Conclusions

Aiming at the non-smooth characteristics of bearing signals, this paper compares and discusses the advantages and disadvantages of several commonly used time–frequency analysis methods, including linear STFT, WT, GST, and nonlinear WVD. The time–frequency images obtained by these four different methods are compared, analyzed, and validated using publicly available datasets. The experimental results show that the generalized S transform converts the 1D vibration signals into 2D image signals and retains the time–frequency feature information in the signals more effectively. On this basis, the GST is combined with the Swin Transformer algorithm in the image field, and the experimental results show that, compared with the traditional methods, the method proposed in this paper obtains more accurate and complete time–frequency information in feature extraction, has a higher accuracy in fault classification and identification, and still has a more accurate diagnosis and better noise robustness under noisy conditions.

Author Contributions: Conceptualization, J.Y. and D.Z.; methodology, J.Y. and D.Z.; software, D.Z.; validation, D.Z. and X.Z.; formal analysis, X.Z. and D.Z.; investigation, X.W. and X.Z.; resources, J.Y.; data curation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, D.Z.; visualization, J.Y.; supervision, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the support provided for this research by the Natural Science Foundation of Guangdong Province (2022A1515011562) and National Natural Science Foundation of China (52201355), by Guangdong Provincial Special Fund for promoting high quality

economic development (Yuerong Office Letter [2020]161, GDNRC [2021]56), and Development of intelligent early warning system for regional equipment failure (CY-ZJ-19-ZC-005).

Data Availability Statement: The Case Western Reserve University dataset used in this paper is available at this URL: https://gitcode.com/open-source-toolkit/78d4f/overview?utm_source=tools_gitcode&index=top&type=card&&isLogin=1 (accessed on 6 October 2024).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Coutinho, R.W.; Boukerche, A. Modeling and Analysis of a Shared Edge Caching System for Connected Cars and Industrial IoT-Based Applications. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2003–2012. [CrossRef]
2. Li, C.; De Oliveira, J.V.; Cerrada, M.; Cabrera, D.; Sánchez, R.V.; Zurita, G. A Systematic Review of Fuzzy Formalisms for Bearing Fault Diagnosis. *IEEE Trans. Fuzzy Syst.* **2018**, *27*, 1362–1382. [CrossRef]
3. Pacheco-Chérrez, J.; Fortoul-Díaz, J.A.; Cortés-Santacruz, F.; Aloso-Valerdi, L.M.; Ibarra-Zarate, D.I. Bearing Fault Detection with Vibration and Acoustic Signals: Comparison among Different Machine Leaning Classification Methods. *Eng. Fail. Anal.* **2022**, *139*, 106515. [CrossRef]
4. Mian, T.; Choudhary, A.; Fatima, S. An Efficient Diagnosis Approach for Bearing Faults Using Sound Quality Metrics. *Appl. Acoust.* **2022**, *195*, 108839. [CrossRef]
5. Sadegh, H.; Mehdi, A.N.; Mehdi, A. Classification of Acoustic Emission Signals Generated from Journal Bearing at Different Lubrication Conditions Based on Wavelet Analysis in Combination with Artificial Neural Network and Genetic Algorithm. *Tribol. Int.* **2016**, *95*, 426–434. [CrossRef]
6. Motahari-Nezhad, M.; Jafari, S.M. Experimental and Data Driven Measurement of Engine Dynamometer Bearing Lifespan Using Acoustic Emission. *Appl. Acoust.* **2023**, *210*, 109460. [CrossRef]
7. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [CrossRef]
8. Gong, W.; Wang, Y.; Zhang, M.; Mihankhah, E.; Chen, H.; Wang, D. A Fast Anomaly Diagnosis Approach Based on Modified CNN and Multisensor Data Fusion. *IEEE Trans. Ind. Electron.* **2021**, *69*, 13636–13646. [CrossRef]
9. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
10. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2446–2455. [CrossRef]
11. Li-Hua, W.; Xiao-Ping, Z.; Jia-Xin, W.; Yang-Yang, X.; Yong-Hong, Z. Motor Fault Diagnosis Based on Short-Time Fourier Transform and Convolutional Neural Network. *Chin. J. Mech. Eng. Ji Xie Gong Cheng Xue Bao* **2017**, *30*, 1357–1368.
12. Jalayer, M.; Orsenigo, C.; Vercellis, C. Fault Detection and Diagnosis for Rotating Machinery: A Model Based on Convolutional LSTM, Fast Fourier and Continuous Wavelet Transforms. *Comput. Ind.* **2021**, *125*, 103378. [CrossRef]
13. Xu, C.; Yang, J.; Zhang, T.; Li, K.; Zhang, K. Adaptive Parameter Selection Variational Mode Decomposition Based on a Novel Hybrid Entropy and Its Applications in Locomotive Bearing Diagnosis. *Measurement* **2023**, *217*, 113110. [CrossRef]
14. Jegadeeshwaran, R.; Sugumaran, V. Fault Diagnosis of Automobile Hydraulic Brake System Using Statistical Features and Support Vector Machines. *Mech. Syst. Signal Process.* **2015**, *52*, 436–446. [CrossRef]
15. Muruganatham, B.; Sanjith, M.A.; Krishnakumar, B.; Murty, S.S. Roller Element Bearing Fault Diagnosis Using Singular Spectrum Analysis. *Mech. Syst. Signal Process.* **2013**, *35*, 150–166. [CrossRef]
16. Zhou, K.; Tang, J. A Wavelet Neural Network Informed by Time-Domain Signal Preprocessing for Bearing Remaining Useful Life Prediction. *Appl. Math. Model.* **2023**, *122*, 220–241. [CrossRef]
17. Wei, B.; Xie, B.; Li, H.; Zhong, Z.; You, Y. An Improved Hilbert–Huang Transform Method for Modal Parameter Identification of a High Arch Dam. *Appl. Math. Model.* **2021**, *91*, 297–310. [CrossRef]
18. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A Deep Convolutional Neural Network with New Training Methods for Bearing Fault Diagnosis under Noisy Environment and Different Working Load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [CrossRef]
19. Fuan, W.; Hongkai, J.; Haidong, S.; Wenjing, D.; Shuaipeng, W. An Adaptive Deep Convolutional Neural Network for Rolling Bearing Fault Diagnosis. *Meas. Sci. Technol.* **2017**, *28*, 095005. [CrossRef]
20. Eren, L.; Ince, T.; Kiranyaz, S. A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *J. Signal Process. Syst.* **2019**, *91*, 179–189. [CrossRef]
21. Xia, M.; Li, T.; Xu, L.; Liu, L.; De Silva, C.W. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE/ASME Trans. Mechatron.* **2017**, *23*, 101–110. [CrossRef]
22. Zhao, X.; Yao, J.; Deng, W.; Jia, M.; Liu, Z. Normalized Conditional Variational Auto-Encoder with Adaptive Focal Loss for Imbalanced Fault Diagnosis of Bearing-Rotor System. *Mech. Syst. Signal Process.* **2022**, *170*, 108826. [CrossRef]

23. Wang, D.; Zhang, M.; Xu, Y.; Lu, W.; Yang, J.; Zhang, T. Metric-Based Meta-Learning Model for Few-Shot Fault Diagnosis under Multiple Limited Data Conditions. *Mech. Syst. Signal Process.* **2021**, *155*, 107510. [CrossRef]
24. Su, H.; Xiang, L.; Hu, A.; Xu, Y.; Yang, X. A Novel Method Based on Meta-Learning for Bearing Fault Diagnosis with Small Sample Learning under Different Working Conditions. *Mech. Syst. Signal Process.* **2022**, *169*, 108765. [CrossRef]
25. Li, X.J.; Yang, D.L.; Jiang, L.L. Bearing Fault Diagnosis Based on Multi-Sensor Information Fusion with SVM. *Appl. Mech. Mater.* **2010**, *34*, 995–999. [CrossRef]
26. Ding, X.; He, Q.; Shao, Y.; Huang, W. Transient Feature Extraction Based on Time–Frequency Manifold Image Synthesis for Machinery Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 4242–4252. [CrossRef]
27. Ville, J. Theorie et Application Dela Notion de Signal Analysis. *Câbles Transm.* **1948**, *2*, 61–74.
28. Abboud, D.; Antoni, J.; Sieg-Zieba, S.; Eltabach, M. Envelope Analysis of Rotating Machine Vibrations in Variable Speed Conditions: A Comprehensive Treatment. *Mech. Syst. Signal Process.* **2017**, *84*, 200–226. [CrossRef]
29. Vin Koay, H.; Huang Chuah, J.; Chow, C.-O. Shifted-Window Hierarchical Vision Transformer for Distracted Driver Detection. In Proceedings of the 2021 IEEE Region 10 Symposium (TENSYMP), Jeju, Republic of Korea, 23–25 August 2021; pp. 1–7.
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
31. Yao, Y.; Zou, J.; Wang, H. Model Constraints Independent Optimal Subsampling Probabilities for Softmax Regression. *J. Stat. Plan. Inference* **2023**, *225*, 188–201. [CrossRef]
32. Kumar, H.S.; Gururaj, U. Fault Diagnosis of Rolling Element Bearing Using Continuous Wavelet Transform and K- Nearest Neighbour. *Mater. Today Proc.* **2023**, *92*, 56–60. [CrossRef]
33. Bai, R.; Meng, Z.; Xu, Q.; Fan, F. Fractional Fourier and Time Domain Recurrence Plot Fusion Combining Convolutional Neural Network for Bearing Fault Diagnosis under Variable Working Conditions. *Reliab. Eng. Syst. Saf.* **2023**, *232*, 109076. [CrossRef]
34. Yuan, P.-P.; Zhang, J.; Feng, J.-Q.; Wang, H.-H.; Ren, W.-X.; Wang, C. An Improved Time-Frequency Analysis Method for Structural Instantaneous Frequency Identification Based on Generalized S-Transform and Synchroextracting Transform. *Eng. Struct.* **2022**, *252*, 113657. [CrossRef]
35. Mirzaeian, R.; Ghaderyan, P. Gray-Level Co-Occurrence Matrix of Smooth Pseudo Wigner-Ville Distribution for Cognitive Workload Estimation. *Biocybern. Biomed. Eng.* **2023**, *43*, 261–278. [CrossRef]
36. Talebitooti, R.; Zarastvand, M.R.; Gheibi, M.R. Acoustic transmission through laminated composite cylindrical shell employing Third order Shear Deformation Theory in the presence of subsonic flow. *Compos. Struct.* **2016**, *157*, 95–110. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.