

Article

Gearbox Fault Diagnosis Based on ICEEMDAN-MPE-AWT and SE-ResNeXt50 Transfer Learning Model

Hongfeng Gao ¹, Tiexin Xu ², Renlong Li ² and Chaozhi Cai ^{2,*} 

¹ Handan Branch of Hebei Special Equipment Supervision and Inspection Institute, Handan 056000, China; mm02939766@163.com

² School of Mechanical and Equipment Engineering, Hebei University of Engineering, Handan 056038, China; xutixin1999@163.com (T.X.); long_long0321@163.com (R.L.)

* Correspondence: caichaozhi1983@163.com

Abstract: Because the gearbox in transmission systems is prone to failure and the fault signal is not obvious, the fault end cannot be located. In this paper, a gearbox fault diagnosis method grounded on improved complete ensemble empirical mode decomposition with adaptive noise, a multiscale permutation entropy and adaptive wavelet thresholding (ICEEMDAN-MPE-AWT) denoising method and an SE-ResNeXt50 transfer learning model are proposed. Initially, the vibration signal is denoised by ICEEMDAN-MPE-AWT, the denoised vibration signal is then converted into a Gram angle field (GAF) diagram, and then the parameters are transferred by the fine-tuning transfer learning strategy. Finally, a GAF diagram is input into the model for training to achieve fault extraction and classification. In this paper, the open gear dataset of Southeast University is used for experimental research. The experimental results show that when using the ICEEMDAN-MPE-AWT and when the signal-to-noise ratio (SNR) of the experimental data is -4 dB, the average accuracy of the GASF+TSE-ResNeXt50 and the GASF+TSE-ResNeXt18 can reach 98.8% and 97.5%, respectively. When the SNR is 6 dB, the accuracy of the above two models reaches 100% and 99.3%, respectively. Moreover, when compared to alternative approaches, the noise reduction method in this paper can better remove noise interference so that the model can better extract fault features. Therefore, the method proposed in this article shows significant improvement in noise reduction and fault classification accuracy compared to other methods.



Citation: Gao, H.; Xu, T.; Li, R.; Cai, C. Gearbox Fault Diagnosis Based on ICEEMDAN-MPE-AWT and SE-ResNeXt50 Transfer Learning Model. *Appl. Sci.* **2024**, *14*, 2565. <https://doi.org/10.3390/app14062565>

Academic Editor: Tobias Meisen

Received: 31 January 2024

Revised: 8 March 2024

Accepted: 14 March 2024

Published: 19 March 2024



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/>).

Keywords: gearbox; fault diagnosis; transfer learning; ICEEMDAN; MPE; AWT; GASF; transmission system

1. Introduction

During the era of Industry 4.0, artificial intelligence and machine learning are continually developing, and neural networks, as a part of artificial intelligence and machine learning, increasingly attract attention. The employment of neural networks to replace workers for machine fault diagnosis and monitoring has become a mainstream method, and many machines are equipped with gear systems on the transmission system. Due to long, continuous work and a harsh working environment, gear sets are very prone to wear, broken teeth, pitting and other damage. When the gearbox fails, it will have a significant effect on the performance of the machine. When the gearbox fails, the vibration signal will be changed. By analyzing the vibration signal, the characteristic signal related to the fault can be separated to determine the fault category of the mechanical system.

The vibration signal generated by the gear contains important information about the development of the fault. Using signal processing technology, the key features related to faults can be clearly identified. The commonly used signal processing techniques include empirical mode decomposition (EMD) [1], cepstrum analysis, wavelet analysis [2] and the wavelet-based threshold denoising method [3]. Gao et al. [4] put forward a fault diagnosis method for rotating machinery grounded in empirical mode decomposition. They

improved the decomposed intrinsic mode function (IMF) and proposed a combined mode function (CMF) to combine adjacent IMFs together to obtain a more accurate oscillation mode describing signal characteristics. Lei et al. [5] introduced a new adaptive ensemble empirical mode decomposition (EEMD) method to address the issue of mode mixing during signal decomposition by improving the EEMD technique. The method overcame the drawback of mode mixing by adaptively selecting the sifting number during the decomposition process for signal analysis. However, the operation efficiency of EEMD [6] is low, and some noise remains after it is decomposed. To this end, Kuai et al. [7] first decomposed the original signal into six IMF components and residual component by using complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and then expressed the time complicatedness of the IMF through permutation entropy to measure characteristics. Kou et al. [8] introduced an improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) energy entropy combined with a support vector machine (SVM), optimized by an artificial fish swarm algorithm for fault diagnosis. The ICEEMDAN energy entropy of different vibration signals was calculated to recognize the location of bearing faults and the accuracy of the model was improved. Wu et al. [9] utilized multiscale permutation entropy (MPE) technology to extract characteristics from bearings and combined with an SVM to automate the fault diagnosis process. Its performance is superior to that of the approach derived from single-scale permutation entropy. Sacerdoti et al. [10] conducted a comparative study on the fault diagnosis of rolling bearings by using methods such as deterministic/random signal separation, time frequency, and cyclical analysis; the results showed that the combination of cepstrum prewhitening and squared envelope spectrum and improved envelope spectrum had the highest accuracy. Antoni et al. [11] discussed the spectrum analysis of cyclostationary signals, pointing out the similarities, differences, and potential pitfalls between cyclic spectrum analysis and classical spectrum analysis.

While the evolution of deep learning theory is ongoing, various models with strong data mining abilities have been extensively used in the realm of fault diagnosis. Zhang et al. [12] utilized a one-dimensional convolutional neural network (1DCNN) for fault diagnosis of rolling bearings. They straightforwardly input the original one-dimensional vibration signal into the network for feature extraction and classification, and good results were obtained. He et al. [13] put forward a ResNet network model to settle a matter of gradient disappearance and gradient explosion in deep neural network training by introducing residual blocks. Guo et al. [14] proposed an adaptive transfer learning fine-tuning method to find the best fine-tuning strategy for each instance and obtained relatively high accuracy on the vision 10-item dataset. Lately, in order to realize better fault diagnosis results in the field of fault diagnosis, the approach of extracting vibration signal features from high-dimensional scales has gradually been adopted [15]. Among various fault diagnosis methods, many transform one-dimensional signals into two-dimensional feature maps [16]. For example, the Gramian angular field (GAF) [17] was first adopted to transform the original time series into two-dimensional feature maps with unique features. Then, the data containing rich feature information were combined with CNN, ResNet, DenseNet [18] MobileNet [19], VGGNet [20] and other models to achieve ideal diagnosis results.

The signals collected in the actual working conditions are often not ideal, and the mechanical operating environment is often more complex, especially when the passenger flow is concentrated and the use environment is more complex, such as high-speed rail, railway stations, airports, overpasses, tunnels and transportation hubs. The signal collected by the sensor contains much noise and is a nonstationary, nonlinear random signal. In fault diagnosis, noise interferes with fault extraction. Therefore, to extract accurate and comprehensive fault characteristic data and realize accurate fault diagnosis, the interference of noise must be reduced as much as possible. To solve the above matters, this paper put forward a fault diagnosis method of gearbox based on an ICEEMDAN-MPE-adaptive wavelet threshold (ICEEMDAN-MPE-AWT) and SE-ResNeXt50 transfer learning models. The principal innovations of this paper are described below:

- (1) ICEEMDAN was used to self-adaptively decompose noisy signals so as to extract effective information in the signals, which can better deal with aliasing and pseudo-mode problems that may exist in the signal decomposition process. At the same time, by calculating the MPE of each IMF, the noise-dominated and signal-dominated components were effectively screened, improving the accuracy of fault diagnosis.
- (2) The AWT noise reduction method was adopted. On the basis of introducing a correction factor to dynamically adjust the threshold, the convergence factor $\theta = k/L$ was used to further adjust the calculation results of the threshold; thus, the noise reduction effect was improved.
- (3) The Gramian angular sum field (GASF) was used to convert one-dimensional data into unique two-dimensional feature images, and the SE ResNeXt50 transfer learning model was proposed by combining fine-tuned transfer learning, ResNeXt and SEnet, which improved the generalization ability and accuracy of the model and ensured the precision of gearbox fault diagnosis.

The remainder of this paper is structured as follows. The ICEEMDAN-MPE-AWT noise reduction method is introduced in Section 2. Section 3 introduces the Gramian angular field. The SE-ResNeXt50 transfer learning model is proposed in Section 4. Section 5 carries out the fault diagnosis experiment and analyzes the experimental results. Some conclusions of this paper are drawn in Section 6.

2. ICEEMDAN-MPE-AWT Denoising

2.1. ICEEMDAN

The ICEEMDAN algorithm is a further improvement on CEEMDAN. Traditional EMD decomposition results in mode mixing, while CEEMDAN partially eliminates mode mixing by adding white noise [21]. However, the introduced white noise is difficult to remove, resulting in increased signal reconstruction errors. In order to solve the problem of reconstruction error, ICEEMDAN was proposed. Unlike CEEMDAN, the white noise added by ICEEMDAN is the k th IMF component after EMD decomposition, that is, $E_k(w^{(i)})$. By introducing adaptive noise adjustment, ICEEMDAN can better deal with aliasing and pseudo-signature problems that may exist during signal decomposition and improve the precision and dependability of signal analysis. The ICEEMDAN algorithm can be described as follows.

x is defined as a vibration signal containing noise; $E_j(\cdot)$ is defined as the j -th IMF component obtained after EMD decomposition; $w^{(i)}$ is defined as the i th Gaussian white noise following the $N(0,1)$ distribution ($i = 1, 2, 3, \dots, i$); $M(\cdot)$ is defined as the local mean of the solution signal; and the coefficient β_k is defined as the SNR of the k stage, when $k < 1$, $\beta_0 = \varepsilon_0 \frac{\sigma(x)}{\sigma(E_1(w^{(i)}))}$, $k \geq 1$, and $\beta_k = \varepsilon_0 \sigma(r_k)$, where ε_0 is the amplitude set in advance, $\sigma(\cdot)$ is the mathematical expectation operator and r_k is the residual of the k stage. The specific calculation process of the ICEEMDAN algorithm based on the above definition is as follows.

- (1) Add $E_1(w^{(i)})$ to the original signal after multiplying the coefficient β_0

$$x^{(i)} = x + \beta_0 E_1(w^{(i)}). \quad (1)$$

The residual of the first stage can be obtained by separately calculating the i -th local mean in Equation (1) using the EMD method

$$r_1 = \langle M(x^{(i)}) \rangle \quad (2)$$

where $\langle \cdot \rangle$ is the operator for calculating the average value throughout the process.

- (2) Subtract the residual difference between the original signal and the first stage to obtain the first component of the signal, denoted as IMF_1 :

$$\text{IMF}_1 = x - r_1. \quad (3)$$

- (3) Calculate the local mean of $r_1 + \beta_1 E_2(w^{(i)})$ to obtain the residual of the second stage

$$r_2 = \left\langle M(r_1 + \beta_1 E_2(w^{(i)})) \right\rangle. \quad (4)$$

Subtract Equation (4) from Equation (2) to obtain the second component of the signal, denoted as IMF₂

$$\text{IMF}_2 = r_1 - r_2 = r_1 - \left\langle M(r_1 + \beta_1 E_2(w^{(i)})) \right\rangle. \quad (5)$$

- (4) When $j = 3, \dots, J$, the residual of the j -th stage can be written as

$$r_j = \left\langle M(r_{j-1} + \beta_{j-1} E_j(w^{(i)})) \right\rangle. \quad (6)$$

- (5) The j -th component of IMF can be calculated as follows:

$$\text{IMF}_j = r_{j-1} - r_j. \quad (7)$$

- (6) Return to step (4) to calculate the next r and IMF.

2.2. MPE

Permutation Entropy (PE) is an approach used to quantify the complexity and randomness of a time series or chaotic dynamical system. It was introduced by Bandt et al. [22]. The count of permutation entropy is simple and robust, and the requirement of time series length is not high. Even if the time series is very short, it is possible to compute. However, the fault characteristic information of a gearbox is spread across multiple scales, and the permutation entropy of a single scale will miss the characteristic information of other scales. MPE is a multiscale complexity method that can measure the randomness of signals at different scales. It first coarse granulates the original time series by a multiscale algorithm, then constructs the multiscale time series and calculates the permutation entropy of each scale. MPE is calculated as below.

- (1) For one-dimensional time series $Z = (z_1, z_2, \dots, z_n)$, the coarse-grained time series can be constructed by Equation (8):

$$y_j^{(s)} = \frac{1}{s} \sum_{j_s}^{i=(j-1)s+1} z_i, \quad 1 \leq j \leq \frac{n}{s}, \quad (8)$$

where s is a positive integer called the scale factor.

- (2) The time series reconstruction of $y_j^{(s)}$ can be obtained by Equation (9):

$$Y_l^{(s)} = \{y_l^{(s)}, y_{l+\tau}^{(s)}, \dots, y_{l+(m-1)\tau}^{(s)}\}, \quad (9)$$

where l is the component of the l -th reconstruction, τ is the delay time, and m is the embedding dimension.

- (3) Reordering the results obtained from Equation (9) in ascending order yields

$$S(g) = \{j_1, j_2, \dots, j_m\}, \quad (10)$$

where $g = 1, 2, \dots, k$, $k \leq m!$, and $S(g)$ is one of $m!$ symbol arrangement patterns.

- (4) Define the PE by using the probability $P_g(g = 1, 2, \dots, k)$ of each symbol sequence, then there is

$$H_p(Z, m, t) = - \sum_{k=1}^k P_g \ln P_g. \quad (11)$$

When $P_g = 1/m!$ and $H_p(Z, m, t)$ is the maximum value of $\ln(m!)$, a normalized result can be obtained by dividing $H_p(Z, m, t)$ by $\ln(m!)$

$$\text{PE} = H_p = \frac{H_p(Z, m, t)}{\ln(m!)}. \quad (12)$$

MPE has three parameters: embedding dimension m , scale factor s and delay τ . The delay τ has little influence on the time series calculation, which is generally 1. For the embedding dimension m , a value that is too small will lead to too few permutation states in the phase space reconstruction process, and it is difficult to accurately measure the dynamic mutation of the time series. A value that is too large will lead to too many permutation states and uniform distribution, which is insensitive to small changes in the system and also increases the computational cost. According to the research of Cao et al. [23], the optimal representation effect of MPE could be obtained by selecting $m = 5$, $m = 6$, or $m = 7$. Therefore, $m = 6$ was selected in this study. A standard method for choosing the scale factors is not uniform, and, generally, a value greater than or equal to 10 can be selected. In this study, $s = 11$ was selected.

2.3. AWT Denoising

Wavelet threshold denoising decomposes signals into various spectral components through wavelet transform. Then the noise wavelet coefficients in high-frequency bands are filtered by setting thresholds, and the effective signal wavelet coefficients in low-frequency bands are extracted and retained to accomplish the objective of noise removal. Therefore, the core of wavelet threshold denoising is a process that combines the measured vibration signal filtering with effective feature extraction. In wavelet threshold denoising, threshold estimation and the threshold function play a key role in the wavelet threshold denoising effect. A threshold that is too large results in the loss of the effective signal, while one that is too small fails to achieve the denoising effect. Consequently, it is crucial to determine the appropriate threshold. The heuristic threshold selection method [24] relying on optimal selection can dynamically modify and choose thresholds throughout, considering noise characteristics and wavelet decomposition, which often achieve better practical results.

This article improved upon the heuristic threshold method and proposed an adaptive threshold choice method. Since the wavelet transform divided the subbands into high and low frequencies based on thresholds at each wavelet decomposition level, using a globally consistent fixed threshold could potentially result in the original signal departing from its original shape when removing noise signals in different signal subspaces across different decomposition levels. By introducing a correction factor to dynamically adjust the thresholds, a threshold choice method based on the correction factor was proposed as shown in Equation (13) [25]:

$$\lambda_{L,k} = \sigma' \sqrt{2\lg N} / (S_{L,k} + b), \quad (13)$$

where, $\lambda_{L,k}$ is the adaptive threshold, σ' is the estimate of the noise standard deviation σ , $S_{L,k}$ are sub-band level parameters, which are used to indicate the threshold modification across various layers of decomposition, and can be introduced as below.

$$S_{L,k} = 2^{(L-k/L)}, \quad (14)$$

where L is the total number of layers of wavelet decomposition; k is the current number of decomposition layers; and b is the regulatory factor.

This paper proposed an adaptive threshold calculation method based on the number of wavelet decomposition layers, as shown in Equation (15).

$$\lambda_{L,k} = (k/L) \times \sigma' \sqrt{2\lg N} / (S_{L,k} + b). \quad (15)$$

During the procedure for threshold selection, the convergence factor $\theta = k/L$ is used to further adjust the calculation result of the threshold, and the threshold amplitude is gradually decreased with the addition of the number of decomposition layers in order to upgrade the resolution of the low-frequency signal and improve the noise reduction effect. The selection of small- and medium-wave bases of wavelet transform should try to select the wavelet basis function with tight support for orthogonality and symmetry [26], and it is

similar to the waveform of the signal to be processed so as to reduce the calculation amount and avoid waveform distortion in the course of signal decomposition and reconstruction. The dbN series wavelet bases show good performance in this respect, and the wavelet threshold denoising based on db4 is especially remarkable. Therefore, in this paper, the db4 wavelet basis was selected for signal denoising. Another important factor that affected the effectiveness of the wavelet transform is the number of wavelet decomposition levels. Too many decomposition layers will distort the reconstructed signal, and too few decomposition layers will not reduce the noise. Therefore, it is also very important to choose the appropriate number of wavelet decomposition layers. In the current paper, the number of wavelet decomposition layers can be obtained by Equation (16):

$$J = \text{fix}(\log_2(\frac{f_s}{f_0})\sqrt{\frac{1}{8}} + 0.5) - 1, \quad (16)$$

where a fix is rounded in the direction of zero.

2.4. Denoising Process of the ICEEMDAN-MPE-AWT

To address the issue of reduced fault recognition accuracy due to the difficulty in extracting fault characteristics caused by a large amount of noise in the signals captured by sensors in complex work environments; first, a denoising method called ICEEMDAN-MPE-AWT was proposed. Then, the noise-dominated IMF components C_i were denoised by using an adaptive wavelet thresholding method. Finally, the denoised noise-dominated components and the undenoised signal-dominated components were reconstructed to obtain the denoised signal. The process of fault diagnosis based on ICEEMDAN-MPE-AWT and SE-ResNeXt50 transfer learning model is illustrated in Figure 1.

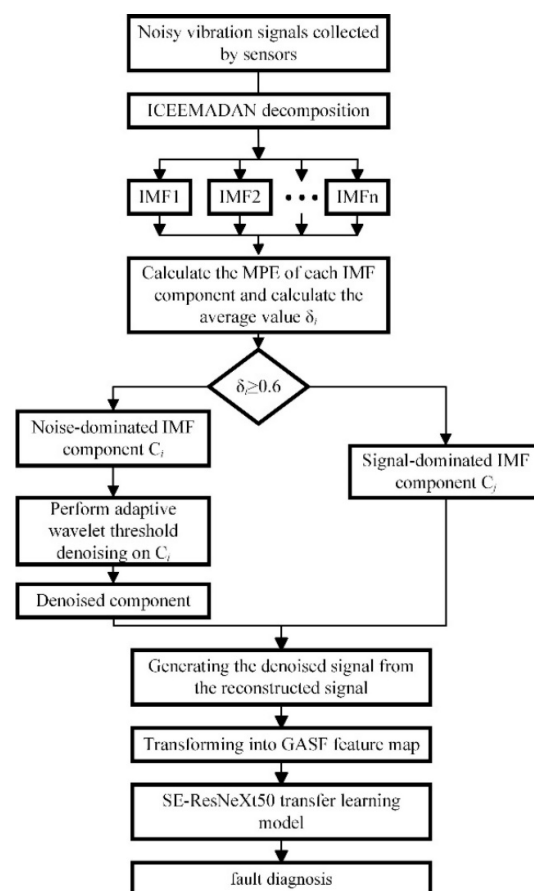


Figure 1. The process of fault diagnosis based on ICEEMDAN-MPE-AWT and SE-ResNeXt50 transfer learning model.

3. Gramian Angular Field

Gramian angular field can transform a one-dimensional time series to two-dimensional image encoding. In Gramian angular field, the elements on the main diagonal correspond to different timestamp values. The article encoded time series into images by using GASF based on the polar coordinate matrix, so as to preserve the correlation between one-dimensional information and the time series. The process of transforming a one-dimensional time series into a two-dimensional image based on GASF is as follows.

- (1) For a given original time series $x = \{x_1, x_2, \dots, x_n\}$, it can be normalized to between 0 and 1 through Equation (17):

$$\tilde{x}_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}}, \quad (17)$$

where, $\tilde{x}_t$ is the scaled signal at time t , and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values in the sequence, respectively. Then, the normalized time series can be mapped to the polar coordinate system according to Equation (18):

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), 0 \leq \tilde{x}_i \leq 1 \\ r_i = \frac{t_i}{N}, t_i \in N \end{cases}, \quad (18)$$

where ϕ_i is the angular value of the polar coordinate corresponding to $\tilde{x}_i$, r_i is the polar coordinate radius corresponding to $\tilde{x}_i$, t_i is the timestamp corresponding to $\tilde{x}_i$, and N is the length of the time series.

- (2) After the one-dimensional signal is mapped to the polar coordinate system, the value of the angular and difference between each point is used to measure the time dependence of the different time intervals. Finally, the GASF can be obtained according to Equation (19):

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cos(\phi_n + \phi_2) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix}, \quad (19)$$

4. SE-ResNeXt50 Transfer Learning Model

4.1. ResNeXt

The ResNeXt [27] network block structure uses the features of VGG and inception networks for reference. The same topology modules are stacked to reduce the excessive free selection of hyperparameters in the training process, and the gradient dispersion caused by a network model that is too deep is avoided. ResNeXt improves the scalability of the model by stacking residuals of the same topology in parallel, reduces hyperparameters and changes the number of branches of control group convolution by cardinality. Figure 2 shows the network structure block of ResNeXt.

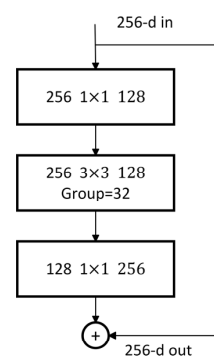


Figure 2. ResNeXt with cardinality = 32.

4.2. SE-ResNeXt50

Hu et al. [28], proposed the SEnet structure from the channel dimension and obtained the weight of each feature channel through the SEnet structure, with the goal of improving the representation capability of the network by convolution of the interdependence between feature channels. The attention module focuses attention on the region of interest, inhibits unnecessary features, assigns different weights to different features and improves the training efficiency and accuracy of the model.

Based on SEnet and ResNeXt, this paper constructed SE-ResNeXt50 and adopted a fine-tuning transfer learning strategy. Firstly, the source SE-ResNeXt50 model was trained on the ImageNet dataset [29] to obtain fault feature information. Then, a new SE-ResNeXt50 target model was created by transferring all the designs and parameters of the source model to the target model and adding an output layer. Eventually, the target model was trained on the target dataset.

The overall structure of the transfer learning model based on GASF and SE-ResNeXt50 is presented in Figure 3. First, the signal was intercepted by overlapping sampling, and then converted into a unique 2D feature map by GASF transformation. Then the feature map was split up into a training set and a test set, and the data of the training set were enhanced. Finally, the training set and test set were introduced into the SE-ResNeXt50 transfer learning model. The parameters of the SE-ResNeXt50 model used in this paper are demonstrated in Table 1.

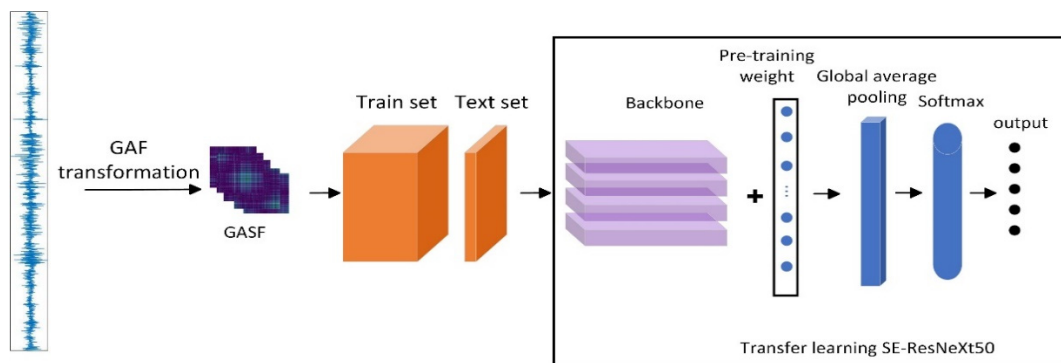


Figure 3. Overall structure.

Table 1. Structure parameters of SE-ResNeXt50.

Stage	Output Size	SE-ResNeXt-50 ($32 \times 4d$)
conv1	112×112	Conv, 7×7 , 64, stride 2
conv2	56×56	max pool, 3×3 , stride 2
		$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \\ fc, [16, 256] \end{bmatrix} \times 3$
conv3	28×28	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 512 \\ fc, [32, 512] \end{bmatrix} \times 4$
conv4	14×14	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 1024 \\ fc, [64, 1024] \end{bmatrix} \times 6$
conv5	7×7	$\begin{bmatrix} 1 \times 1, 1024 \\ 3 \times 3, 1024 \\ 1 \times 1, 2048 \\ fc, [128, 2048] \end{bmatrix} \times 3$
global average pool	1×1	5-d fc, softmax

5. Experiment and Analysis

5.1. Data Sources

The experimental data chosen in this paper are the gearbox fault diagnosis dataset from Southeast University of China [30]. The gearbox fault dataset of Southeast University is the vibration signal collected by drivetrain dynamic simulator (DDS). DDS is manufactured by Changxing Shengyang Technology Co., Ltd., Huzhou, China. The transmission system dynamics simulator mainly encompasses a motor, a motor controller, a planetary gearbox, a parallel gearbox, a brake and a brake controller. The picture of the test rig can be found in the reference [30]. In the test rig, the gears are straight cylindrical gears, the number of gear teeth in the reduction gearbox are 32, 80, 72 and 40, the module is 2, the transmission ratio of the reduction gearbox is 4.5, the brake used for gearbox loading is the magnetic powder brake, whose power is 3 KW, the data acquisition card used in measurement equipment is VQ-8, the sampling frequency is 51.2 kHz, the vibration sensor used in the measurement equipment is TES001V, and the sensitivity is 100 mv/g. The experimental platform can simulate five gear states under two working conditions. The two working conditions are (1) motor speed of 1200 rpm, brake load of 0 Nm; and (2) motor speed of 1800 rpm, brake load of 7.32 Nm. In the experiment, the sampling frequency of the data acquisition system is 51.2 kHz, and the duration of each fault sampling is 20 s. Table 2 lists the specific fault types.

Table 2. Fault Types.

Fault Label	Fault Types
0	normal
1	Tooth defect
2	Tooth broken
3	Tooth root crack
4	Tooth surface wear

5.2. Data Processing

Due to the limited fault vibration data collected, it may be difficult to retrieve fault features. Therefore, the acquired vibration signals were enhanced in this paper. Data enhancement is a way to increase the value of limited data without substantially increasing it. In this paper, the original one-dimensional vibration signal was augmented using the overlap sampling method. The original data were first segmented using sliding windows of size 1024 and a sliding stride of 300. The segmented signals were then transformed into 224×224 feature images using GASF transformation. Figure 4 depicts the schematic diagram of the overlap sampling. The gearbox dataset consisted of five classifications for both 0 Nm and 7.32 Nm loads. For each classification, 3500 feature images were generated. In each iteration, 3000 feature images were randomly selected and split into a training set and a testing set in an 8:2 ratio. The distribution of specific fault samples is shown in Table 3.

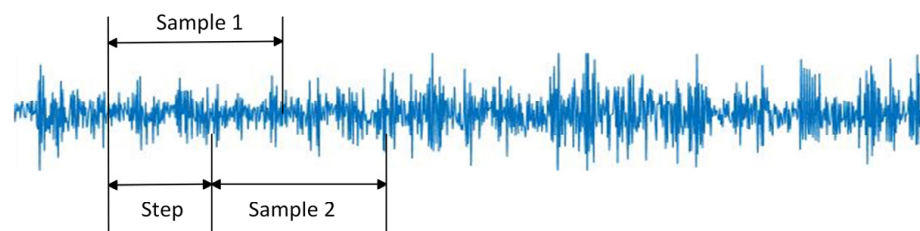


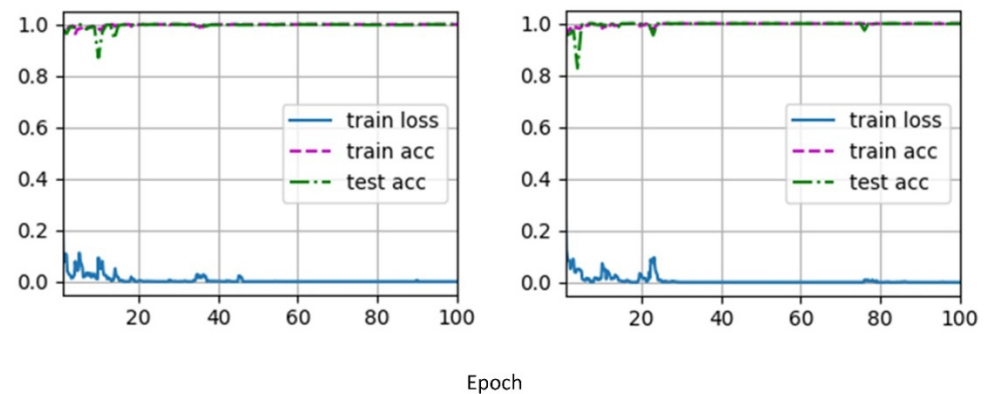
Figure 4. Overlap sampling.

Table 3. Fault samples distribution of gearbox.

Fault Table	Normal	Tooth Defect	Tooth Broken	Tooth Root Crack	Tooth Surface Wear	Load
Train	120	120	120	120	120	0 Nm
Test	480	480	480	480	480	
Train	120	120	120	120	120	7.32 Nm
Test	480	480	480	480	480	

5.3. Model Training and Comparative Analysis

All experiments in this paper were implemented using the TensorFlow2.7.0 deep learning framework and ran on a computer with an Intel core i5-12490F processor and a RTX4060 graphics card. In addition, the computer's operating system was Windows11. During model training, the mini-batch size was set to 20. The Adam optimization algorithm and LambdaLR custom learning rate scheduling strategy were employed, and the initial learning rate was set to 0.001. The generated GASF image data were shuffled randomly and fed into the SE-ResNeXt50 model with transfer learning (TSE-ResNeXt50) in batches for training. Figure 5 shows the training accuracy, test accuracy, and loss curves of GASF+TSE-ResNeXt50 method under 0 Nm and 0.72 Nm loads.



(a) Training results when the load is 0 Nm (b) Training results when the load is 7.32 Nm

Figure 5. Training accuracy, testing accuracy and loss curves of GASF+TSE-ResNeXt50.

As can be seen from the figure, TSE-ResNeXt50 trained on GASF datasets generated under different loads, and all achieved high diagnosis accuracy from the beginning. When the load is 0 Nm, the accuracy and loss curves do not fluctuate greatly, and the convergence speed is fast. When the load is 7.32 Nm, the accuracy and loss curves show small fluctuations at the beginning of training, but all converge quickly after 20 epochs and tend to stabilize, and the accuracy reaches 100%. The training results show that using GASF to convert one-dimensional vibration signals into two-dimensional feature maps yields more obvious fault features, making it easier to extract fault features. Therefore, the method proposed in this article has some ability to extract abstract features at the beginning of training and has achieved good results.

For the sake of additionally checking the reliability of the suggested approach, the GASF+TSE-ResNeXt50 method was experimentally compared to other fault diagnosis methods by using the GASF gear dataset under loads of 7.32 Nm and 0 Nm, respectively, as training samples, and all experimental results were the average of 10 test results. In this paper, a comparison and analysis were conducted between five models: TICNN [12], GASF+SE-ResNeXt18, GASF+MobileNetV3, MTF+TSE-ResNeXt50, and GASF+TSE-ResNeXt50. Among the above methods, the MTF (Markov transition field) in MTF+TSE-ResNeXt50 is a method of transforming one-dimensional time series data into two-dimensional feature images; it adopts a quantile partitioning strategy, which refers to the division of each

sample into different regions that contain the same number of sampling points. Table 4 demonstrates the experimental results of the five methods written above.

Table 4. Accuracy of Different Models on GASF Gearbox-Based Datasets.

Model	0.72 Nm	0 Nm
TICNN	99%	99.2%
GASF+SE-ResNeXt18	99.5%	99.6%
GASF+MobileNetV3	99.7%	99.8%
MTF+TSE-ResNeXt50	99.8%	100%
GASF+TSE-ResNeXt50	100%	100%

It can be found from Table 4 that in the gearbox fault diagnosis, the TSE-ResNeXt50 model based on MTF and GASF dataset has good generalization performance. When using the GASF dataset, the accuracy of the TSE-ResNeXt50 reached 100% under two different loads. When using the MTF dataset, the accuracy of the TSE-ResNeXt50 reached 99.8% and 100% under two different loads, respectively. The accuracy of GASF+SE-ResNeXt18 and GASF+TSE-ResNeXt50 is very close, which confirms the advantages of the SE-ResNeXt module. Although the accuracy of GASF+MobileNetV3 and TICNN has also reached over 99%, overall, GASF+TSE-ResNeXt50 performs better in gear fault classification, with higher accuracy and stronger robustness.

5.4. Noise Reduction Performance Analysis

With the aim of conveniently measuring the efficiency improvement of the model after noise diminishment, Gaussian white noise with a certain signal-to-noise ratio (SNR) is added to the original vibration signal. The calculation formula of SNR is as follows

$$SNR_{dB} = 10 \log_{10}(P_{signal}/P_{noise}) \quad (20)$$

where P_{signal} is the power of the original vibration signal, and P_{noise} is the power of the added Gaussian white noise.

Figure 6 shows the comparison result between the original signal of the gearbox tooth broken under 7.32 Nm load and the noised signal with SNR = 0. As seen in the figure, after increasing the noise, the fault characteristics in the signal are interfered with to different degrees, and the fault features are not apparent. Hence, noise lessening is needed for the noisy signals to upgrade the discernment precision of the model for the fault signals containing a lot of noise.

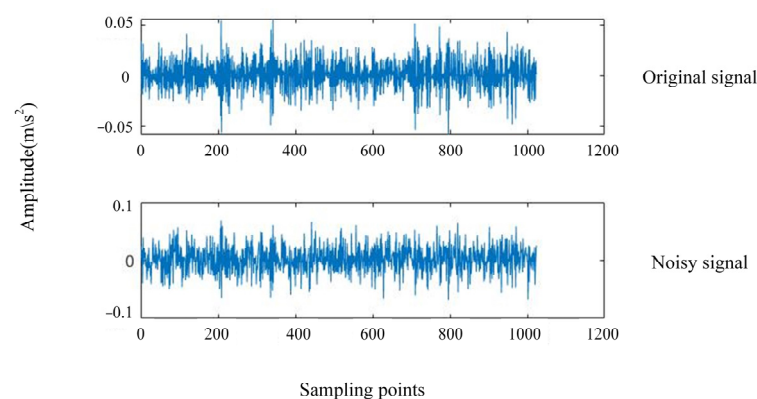


Figure 6. Comparison result between the original signal and the noised signal.

The noise reduction approach presented in this paper was adopted for noise reduction. Firstly, the ICEEMDAN algorithm was used to decompose the noised signal, and the result is presented in Figure 7. As can be found from the figure, after ICEEMDAN decomposition,

eight IMF components and one residual component are obtained, respectively. ICEEMDAN successfully decomposed the noisy fault signal into high-, medium- and low-frequency components, and each component has different time scales, without obvious mode aliasing. Then, the average MPE of each IMF component was calculated, which is presented in Table 5. As depicted in the table, the average value of IMF1–IMF4 is greater than 0.6, so the adaptive wavelet threshold noise reduction is needed.

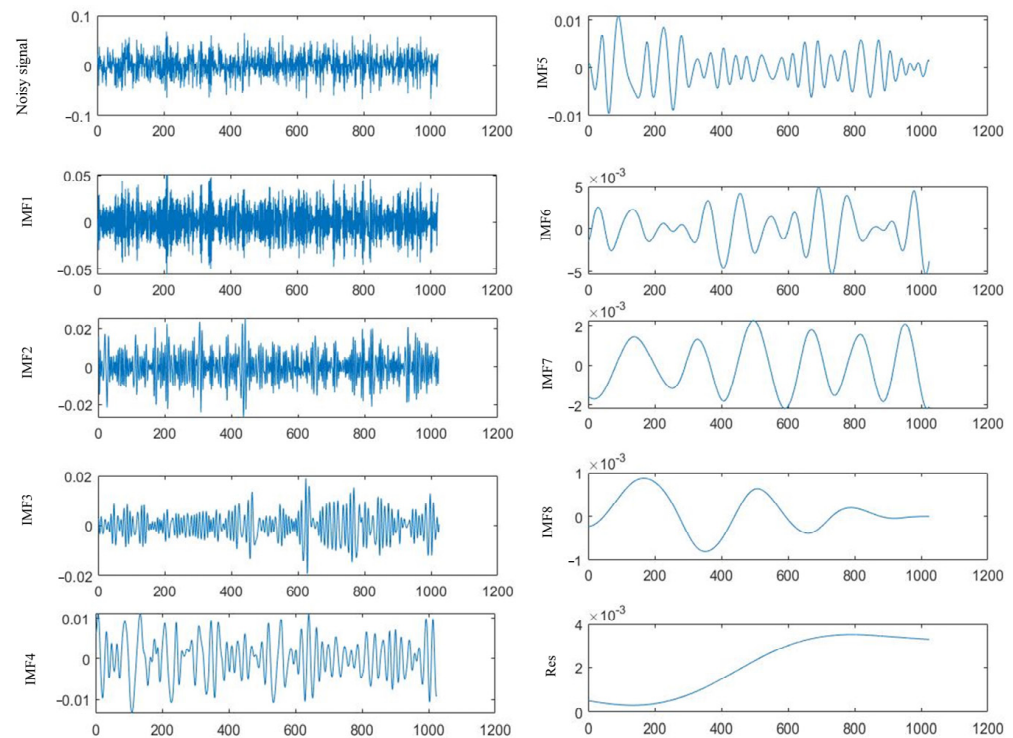


Figure 7. The result of ICEEMDAN decomposition.

Table 5. MPE of each IMF component.

s	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	Res
1	0.834	0.734	0.497	0.338	0.245	0.189	0.158	0.127	0.108
2	0.865	0.849	0.722	0.480	0.340	0.249	0.185	0.141	0.114
3	0.837	0.787	0.789	0.591	0.415	0.304	0.213	0.152	0.120
4	0.787	0.795	0.768	0.682	0.470	0.348	0.241	0.169	0.126
5	0.777	0.779	0.697	0.725	0.518	0.387	0.263	0.186	0.131
6	0.738	0.763	0.716	0.732	0.553	0.434	0.287	0.192	0.137
7	0.728	0.736	0.697	0.732	0.572	0.463	0.310	0.211	0.142
8	0.709	0.707	0.711	0.698	0.598	0.471	0.335	0.217	0.146
9	0.702	0.687	0.692	0.689	0.627	0.504	0.331	0.227	0.151
10	0.674	0.680	0.677	0.669	0.643	0.500	0.368	0.246	0.156
11	0.651	0.658	0.658	0.640	0.652	0.503	0.369	0.254	0.161
AVG	0.755	0.743	0.693	0.634	0.512	0.396	0.278	0.193	0.136

IMF components with an average MPE greater than 0.6 were first denoised by using AWT, and then the denoised IMF components were reconstructed with the undenoised IMF components. Specifically, the denoised IMF1–IMF5 components were reconstructed with IMF6–IMF8 and the residual component, resulting in a denoised signal. Figure 8 illustrates the comparison result compared to the denoised signal and the noisy signal. It can be seen that the amplitude of the noisy signal is $-0.1 \sim 0.1$; after the noise reduction approach raised in this paper, the amplitude is reduced to $-0.05 \sim 0.05$. Moreover, it can be found from the figure that the signal after noise reduction has more significant fault

characteristics, more significant vibration and shock, and a higher coincidence degree with the original signal compared with the signal without noise reduction.

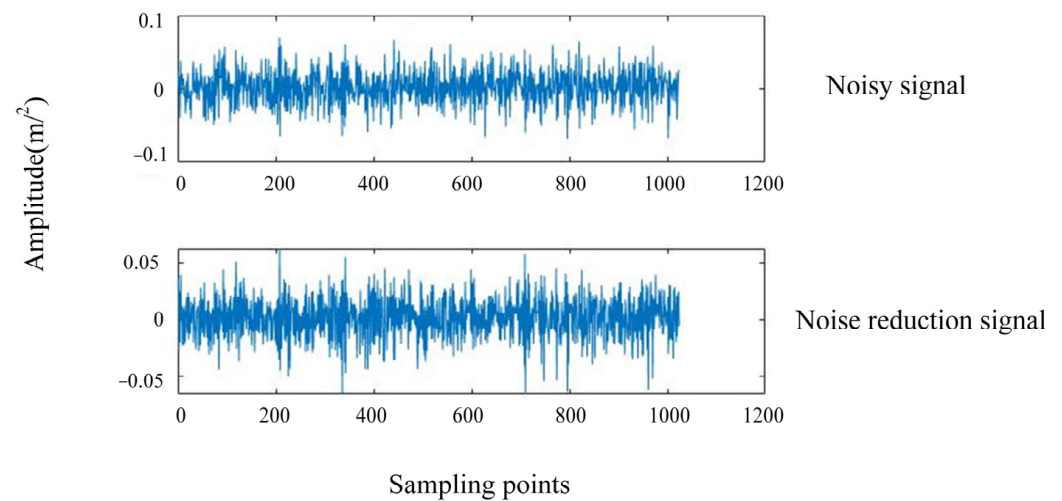


Figure 8. The comparison result between the denoised signal and the noised signal.

In order to verify the efficacy of the ICEEMDAN-MPE-AWT noise reduction method in improving the fault diagnosis of the model in noisy environments, the experimental dataset with a load of 7.32 Nm was selected, and noise with SNR ranging from -4 dB to 6 dB was introduced to the test set. Then, the precision of the raised method before and after noise reduction was measured and compared to prove the effectiveness of the raised approach. For the sake of checking the generalization of the noise reduction method raised in this paper, the GASF+TSE-ResNeXt50 model and GASF+TSE-ResNeXt18 were selected as the test model for fault diagnosis. The recognition accuracy of the model on the test set was taken as the evaluation index, 10 repeated tests were carried out for each model, and the average value of the 10 tests was seen as the final results. The results are illustrated in Table 6. It can be found from Table 6 that after noise reduction by adopting the approach in this paper, the accuracy of GASF+TSE-ResNeXt50 can still reach 98.8% when the SNR is -4 dB. Therefore, the approach introduced in this paper has better noise reduction efficacy in noisy environments and can ensure the precision of the model. Figure 9 shows the precision curves of the two models before and after noise reduction, which can more intuitively see the effectiveness of ICEEMDAN-MPE-AWT noise reduction method. Figure 10 shows the confusion matrices of GASF+TSE-ResNeXt50 before and after denoising when SNR is -4 dB and the load is 7.32 Nm. The horizontal coordinate represents the true label of the failure, the vertical coordinate represents the predicted category of the failure, and the numbers on the main diagonal of the matrix represent the number of samples correctly classified for each type of failure.

Table 6. Accuracy under different SNR.

Model	SNR (dB)	−4	−2	0	2	4	6
GASF+TSE-ResNeXt50	Before noise reduction	95.5%	97.5%	98.5%	99.0%	99.7%	100%
	After noise reduction	98.8%	99.2%	99.7%	99.8%	100%	100%
GASF+TSE-ResNeXt18	Before noise reduction	92.5%	94.2%	96.3%	97.7%	98.2%	99.2%
	After noise reduction	97.5%	98.2%	98.8%	99%	99.2%	99.3%

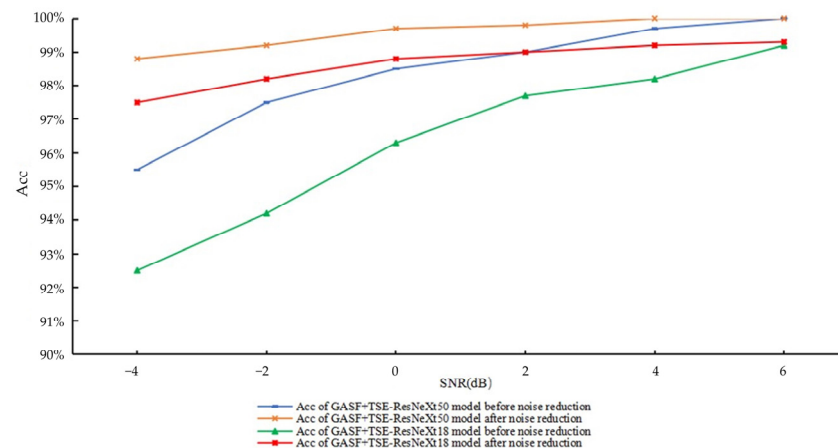


Figure 9. GASF+TSE-ResNeXt50 and GASF+TSE-ResNeXt18 models before and after noise reduction.

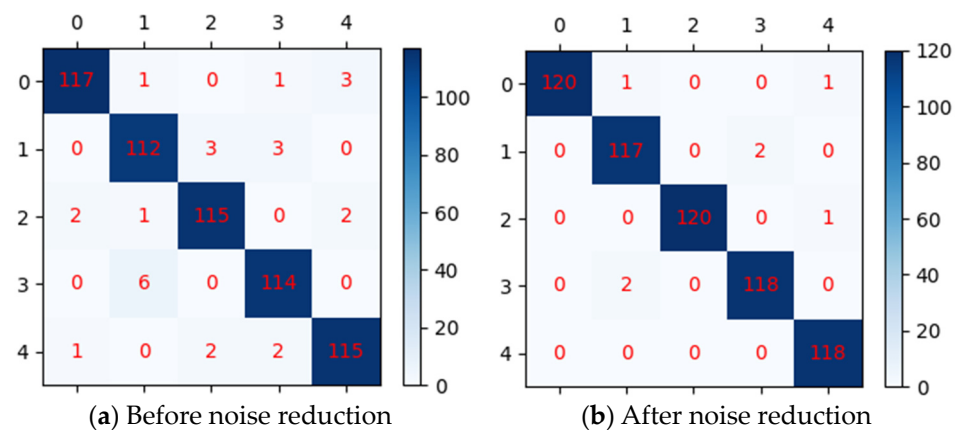


Figure 10. The confusion matrices of GASF+TSE-ResNeXt50 before and after denoising when SNR is -4 dB and load is 7.32 Nm.

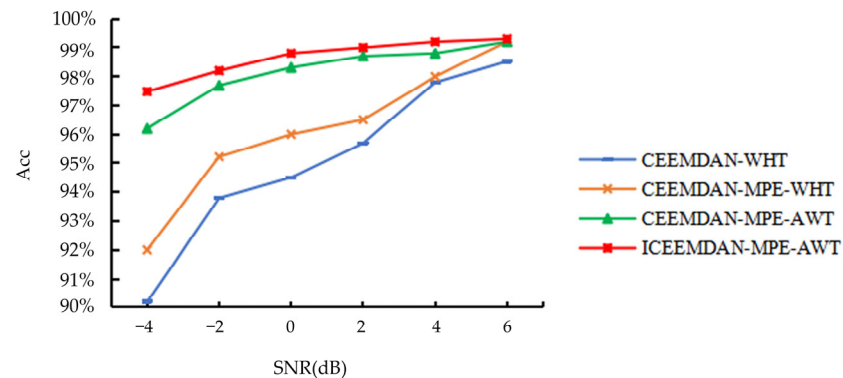
5.5. Performance Analysis of Different Noise Reduction Methods

For the sake of validating the superiority of the noise reduction approach suggested in the current paper, a comparative experimental study was conducted by using ICEEMDAN-MPE-AWT, CEEMDAN-MPE-AWT, ICEEMDAN-WHT (ICEEMDAN with Wavelet Hard Thresholding) and ICEEMDAN-MPE-WHT. Among the above denoising methods, CEEMDAN-WHT used the cross-correlation coefficient between intrinsic mode functions to distinguish whether noise reduction is required for this component. The dataset with a load of 7.32 Nm was selected as the experimental dataset. Since GASF+TSE-ResNeXt50 has achieved a good effect before and after noise reduction, to reflect the effectiveness of the noise reduction approach introduced in this paper, GASF+TSE-ResNeXt18 was selected as the test model. The recognition accuracy of the model on the test set was taken as the appraisal index, 10 repeated tests were carried out for each model, and the average value of the 10 tests was taken as the ultimate result.

The experimental outcomes are demonstrated in Table 7 and Figure 11. It can be found that ICEEMDAN-MPE-AWT and CEEMDAN-MPE-AWT denoising methods achieved high accuracy; when SNR is 6 dB, both denoising methods can ensure an accuracy of over 99%, and as the SNR decreases, the accuracy decreases less. In addition, ICEEMDAN-MPE-AWT still has an accuracy of 99% when SNR is 2 dB, and the accuracy is higher than that of CEEMDAN-MPE-AWT. This shows that ICEEMDAN can better extract the local features and periodic components in the signal than that of CEEMDAN, it acquires the effective information of the signal and obtains more stable and reliable IMF components. Therefore, the combination of ICEMDAN with MPE and AWT has achieved better noise reduction effects and improved the accuracy of the model.

Table 7. Accuracy of each denoising method on GASF+ TSE-ResNeXt18.

SNR (dB)	−4	−2	0	2	4	6
CEEMDAN-WHT	90.2%	93.8%	94.5%	95.7%	97.8%	98.5%
CEEMDAN-MPE-WHT	92.0%	95.2%	96.0%	96.5%	98.0%	99.2%
CEEMDAN-MPE-AWT	96.2%	97.7%	98.3%	98.7%	98.8%	99.2%
ICEEMDAN-MPE-AWT	97.5%	98.2%	98.8%	99.0%	99.2%	99.3%

**Figure 11.** Each denoising method on GASF+SE ResNeXt18.

Compared with CEEMDAN-MPE-AWT, the accuracy of CEEMDAN-MPE-WHT decreased significantly; this indicates that setting wavelet coefficients below the threshold to zero in WHT may give rise to the loss of elaborate information in the signal. Especially for some weak detail signals or low amplitude variation signals, WHT may not be able to effectively retain this detailed information. In addition, WHT needs to manually select the appropriate threshold for signal denoising. However, in practical applications, especially when the statistical characteristics of the signal are not clear or the noise level changes greatly, it is often very difficult to determine a suitable threshold value. Therefore, the accuracy of CEEMDAN-MPE-WHT has decreased to some extent. The AWT used in this paper can adjust the threshold dynamically by introducing a correction factor, and the calculation results of the threshold can be further adjusted by introducing a convergence factor $\theta = k/L$, thereby improving the resolution of low-frequency signals and enhancing the denoising effect. This method can dynamically adjust the threshold value according to the local characteristics of the signal, which can better keep the details and edge information of the signal, has strong adaptability, and improves the stability and accuracy of denoising.

Compared with CEEMDAN-MPE-WHT, CEEMDAN-WHT has similar accuracy when the SNR is high. However, when the SNR decreases, CEEMDAN-WHT fluctuates significantly. This indicates that the cross-correlation coefficient may not be able to accurately distinguish the noise component from the signal component, and it may be easy to divide the low-frequency component into noise, resulting in certain fluctuations in accuracy. In contrast, MPE can reveal the multiscale structure of a signal by calculating the arrangement entropy at different scales, which allows it to consider both local and global characteristics of the signal, thus, providing more comprehensive signal analysis results. At the same time, MPE is robust to noise and interference. It can measure the complexity and randomness of the signal by calculating the arrangement entropy of the signal so that the interference of noise can be reduced to a certain extent. Therefore, CEEMDAN-MPE-WHT has relatively good fault diagnosis accuracy. In summary, ICEEMDAN-MPE-AWT proposed in this paper has better noise reduction performance than other methods in noisy circumstances and can improve the accuracy of the model for noisy signals.

6. Conclusions

This paper focuses on the problem of the signal collected by sensors including a great deal of noise in complex working environments, making the model unable to effectively

obtain the fault characteristics from the signal. The ICEEMDAN-MPE-AWT method for noise reduction was proposed. Firstly, the principle of ICEEMDAN was introduced, and the problems existing in the CEEMDAN algorithm were solved. The principle of MPE and AWT denoising was then introduced, followed by the implementation of the ICEEMDAN-MPE-AWT denoising method proposed in this paper. In addition, a transfer learning model based on GASF and SE-ResNeXt was proposed to diagnose gearbox faults, and the experimental comparison with TICNN, GASF+TSE-ResNeXt18 and other models verified that the raised model has higher precision in gearbox fault diagnosis.

In addition, experimental analysis and research were conducted on the performance of the noise reduction approach raised in this paper. Firstly, the gearbox fault dataset with and without noise reduction was fed into the model for training. The experimental outcomes showed that the training effect of the dataset after noise reduction is significantly better than that of the dataset without noise reduction. Experimental results also showed that the raised approach has good noise reduction performance and can ensure the accuracy of the model. Compared with CEEMDAN-MPE-AWT, ICEEMDAN-WHT and ICEEMDAN-MPE-WHT, the proposed ICEEMDAN-MPE-AWT in this paper has better noise reduction performance.

The method proposed in this paper has a great noise reduction ability, so that the neural network model can still achieve a high fault diagnosis precision in noisy environments. In short, the approach in this paper is expected to play a role in the fault diagnosis and fault monitoring of different machines. Just like wind turbines, which produce noise during work, if the gearbox fails, it cannot be maintained in time solely based on worker experience. In the area of fault diagnosis in the future, we will continue to explore how to conduct fault monitoring and diagnosis of machinery in loud noise conditions so that deep learning can be better applied to the field of machinery.

Author Contributions: H.G.: Formal analysis, Investigation, Methodology, Writing—original draft and Writing—review and editing. T.X.: Formal analysis, Software, Validation and Writing—review and editing. R.L.: Formal analysis, Investigation, and Writing—original draft. C.C.: Funding acquisition, Methodology, Project administration, Formal analysis, and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Market Supervision and Administration Bureau Research Plan Project of Hebei Province under Grant 2023ZD11 and in part by the Key Laboratory of Intelligent Industrial Equipment Technology of Hebei Province (Hebei University of Engineering) under Grant 202206.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in [github] at [URL: https://gitcode.com/cathysiyu/mechanical-datasets/tree/master?utm_source=csdn_github_accelerator&isLogin=1]. reference number [30]. These data were derived from the following resources available in the public domain: [URL: https://gitcode.com/cathysiyu/mechanical-datasets/tree/master?utm_source=csdn_github_accelerator&isLogin=1].

Conflicts of Interest: The authors declare no conflicts of interests.

References

1. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
2. Zhong, X.; Huang, T.; Mei, Q.; Gao, X.; Zhao, X. A gearbox fault diagnosis method based on MKurt spectrum and CYCBD. *Insight-Non-Destr. Test. Cond. Monit.* **2021**, *63*, 472–478. [CrossRef]
3. Chang, S.G.; Yu, B.; Vetterli, M. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.* **2000**, *9*, 1532–1546. [CrossRef]
4. Gao, Q.; Duan, C.; Fan, H.; Meng, Q. Rotating machine fault diagnosis using empirical mode decomposition. *Mech. Syst. Signal Process.* **2008**, *22*, 1072–1081. [CrossRef]

5. Lei, Y.; Li, N.; Lin, J.; Wang, S. Fault Diagnosis of Rotating Machinery Based on an Adaptive Ensemble Empirical Mode Decomposition. *Sensors* **2013**, *13*, 16950–16964. [CrossRef]
6. Lei, Y.; He, Z.; Zi, Y. Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2009**, *23*, 1327–1338. [CrossRef]
7. Kuai, M.; Cheng, G.; Pang, Y.; Li, Y. Research of Planetary Gear Fault Diagnosis Based on Permutation Entropy of CEEMDAN and ANFIS. *Sensors* **2018**, *18*, 782. [CrossRef] [PubMed]
8. Kou, Z.; Yang, F.; Wu, J.; Li, T. Application of ICEEMDAN energy entropy and AFSA-SVM for fault diagnosis of hoist sheave bearing. *Entropy* **2020**, *22*, 1347. [CrossRef] [PubMed]
9. Wu, S.D.; Wu, P.H.; Wu, C.W.; Ding, J.J.; Wang, C.C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **2012**, *14*, 1343–1356. [CrossRef]
10. Sacerdoti, D.; Strozzi, M.; Secchi, C. A Comparison of Signal Analysis Techniques for the Diagnostics of the IMS Rolling Element Bearing Dataset. *Appl. Sci.* **2023**, *13*, 5977. [CrossRef]
11. Antoni, J. Cyclic spectral analysis in practice. *Mech. Syst. Signal Process.* **2007**, *21*, 597–630. [CrossRef]
12. 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]
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE 2016 Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
14. Guo, Y.; Shi, H.; Kumar, A.; Grauman, K.; Rosing, T.; Feris, R. Spottune: Transfer learning through adaptive fine-tuning. In Proceedings of the IEEE/CVF 2019 Conference on Computer Vision and Pattern Recognition (CVPR), Los Angeles, CA, USA, 19–20 June 2019; pp. 4805–4814.
15. Medina, R.; Macancela, J.C.; Lucero, P.; Cabrera, D.; Sánchez, R.V.; Cerrada, M. Gear and bearing fault classification under different load and speed by using Poincaré plot features and SVM. *J. Intell. Manuf.* **2022**, *33*, 1031–1055. [CrossRef]
16. Kuncan, M.; Kaplan, K.; Minaz, M.R.; Kaya, Y.; Ertunç, H.M. A novel feature extraction method for bearing fault classification with one dimensional ternary patterns. *ISA Trans.* **2020**, *100*, 346–357. [CrossRef] [PubMed]
17. Sun, S.; Ren, J. GASF-MSNN: A new fault diagnosis model for spatiotemporal information extraction. *Ind. Eng. Chem. Res.* **2021**, *60*, 6235–6248. [CrossRef]
18. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE 2017 Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
19. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
21. Xu, Y.; Luo, M.; Li, T.; Song, G. ECG signal de-noising and baseline wander correction based on CEEMDAN and wavelet threshold. *Sensors* **2017**, *17*, 2754. [CrossRef] [PubMed]
22. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef]
23. Cao, Y.; Tung, W.W.; Gao, J.B.; Protopopescu, V.A.; Hively, L.M. Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E* **2004**, *70*, 046217. [CrossRef] [PubMed]
24. Misiti, M.; Misiti, Y.; Oppenheim, G.; Poggi, J.M. *Wavelet Toolbox*; The MathWorks Inc.: Natick, MA, USA, 1996; Volume 15, p. 21.
25. Poornachandra, S. Wavelet-based denoising using subband dependent threshold for ECG signals. *Digit. Signal Process.* **2008**, *18*, 49–55. [CrossRef]
26. Yi, T.H.; Li, H.N.; Zhao, X.Y. Noise smoothing for structural vibration test signals using an improved wavelet thresholding technique. *Sensors* **2012**, *12*, 11205–11220. [CrossRef] [PubMed]
27. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE 2017 Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE 2018 Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
29. Huh, M.; Agrawal, P.; Efros, A.A. What makes ImageNet good for transfer learning? *arXiv* **2016**, arXiv:1608.08614.
30. 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. Available online: https://gitcode.com/cathysiyu/mechanical-datasets/tree/master?utm_source=csdn_github_accelerator&isLogin=1 (accessed on 17 January 2018). [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.