

Article

A Two-Stage Framework for Time-Frequency Analysis and Fault Diagnosis of Planetary Gearboxes

Pinyang Zhang  and Changzheng Chen *

School of Mechanical Engineering, Shenyang University of Technology, Shenyang 110870, China; zhagpinyang@163.com

* Correspondence: czchen@sut.edu.cn

Abstract: In the operation and maintenance of planetary gearboxes, the growth of monitoring data is often faster than its analysis and classification. Careful data analysis is generally considered to require more expertise. Rendering the machine learning algorithm able to provide more information, not just the diagnosis conclusion, is promising work. This paper proposes an analysis and diagnosis two-stage framework based on time-frequency information analysis. In the first stage, a U-net model is used for the semantic segmentation of vibration time-frequency spectrum to highlight faulty feature regions. Shape features are then calculated to extract useful information from the segmented image. In the second stage, the decision tree algorithm completes the health state classification of the planetary gearboxes using the input of shape features. The real data of wind turbine planetary gearboxes and augmented data are utilized to verify the proposed framework's effectiveness and superiority. The F1-score of segmentation and the classification accuracy reach 0.942 and 97.4%, respectively, while in the environmental robustness experiment, they reached 0.747 and 83.1%. Equipping the two-stage framework with different analytical methods and diagnostic algorithms can construct flexible diagnostic systems for similar problems in the community.

Keywords: planetary gearbox; time-frequency analysis; fault diagnosis; U-net; vibration



Citation: Zhang, P.; Chen, C. A Two-Stage Framework for Time-Frequency Analysis and Fault Diagnosis of Planetary Gearboxes. *Appl. Sci.* **2023**, *13*, 5202. <https://doi.org/10.3390/app13085202>

Academic Editor: Ki-Yong Oh

Received: 23 March 2023

Revised: 9 April 2023

Accepted: 20 April 2023

Published: 21 April 2023



Copyright: © 2023 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

Planetary gearboxes are widely used in many industrial categories such as wind power generation, mining, metal forming, etc., due to their compact structure, large transmission ratio, and stable operation [1,2]. In these applications, planetary gearboxes are usually installed in the transmission chain transmitting large torque and are subject to harsh working environments such as dynamic load and extreme temperature. Hence, they are prone to various failures [3]. It is important to ensure the safe operation of equipment to detect faults as early as possible through the analysis of sensor data to avoid unplanned shutdowns or catastrophic failures [4–6].

As a well-established field, the vibration analysis-based fault diagnosis method has been widely used in planetary gearboxes. The representation methods of vibration signals can generally be summarized into three categories: time domain, frequency domain, and time-frequency (T-F) domain [7]. T-F domain analysis is a powerful tool for dealing with non-stationary signals. Chen and Feng [8] proposed an iterative generalized T-F reassignment method by exploiting the uniqueness of iterative generalized demodulation to decompose nonstationary, multi-component signals into mono-components of constant frequency, thus, meeting the requirement of mono-component with linear instantaneous frequency by T-F reassignment and improving the T-F readability in planetary gearbox fault diagnosis. With consideration of the time-varying characteristics of planetary gearboxes, Han and Feng [9] used the local maximum mean discrepancy to evaluate the data distributions between relevant subclasses in source and target tasks and proposed a deep residual joint subclass alignment transfer network based on T-F features. Yuan et al. [10]

extracted fault features from wavelet T-F images from the perspective of image texture analysis and proposed a novel fusion fault diagnosis framework of gray level co-occurrence matrix and label consistent K-SVD. Tu et al. [11] developed a new method termed as generalized wavelet-based synchro squeezing transform to deal with a strong, modulated non-stationary signal, which has better performance as compared with traditional approaches in the energy concentration of T-F representations and accuracy of the mode reconstruction. In [12], Dhamande and Chaudharicon considered a more complex but real situation: compound gear-bearing fault and proposed a fault diagnosis method based on T-F statistical features of the discrete wavelet transform and continuous wavelet transform.

T-F segmentation is one of the crucial means of analyzing T-F distribution. Limited by computer calculation levels, early investigations mainly focused on threshold segmentation methods represented by the Otsu method [13], edge detection [14], and gray level histogram [15]. With the development of computer technology and machine learning, researchers have put forward many new ideas and methods for T-F analysis. Medical Informatics is the field where early T-F analysis technologies have achieved fruitful results. For example, Zhang et al. [16] combined the Hilbert transform with the Wigner-Ville distribution to bring about hybrid T-F analysis and used ResNet to analyze cardiac arrhythmia via heartbeat classification. Cheng et al. [17] proposed a novel method based on T-F analysis and the CNN-LSTM cascade model to automatically detect atrial fibrillation, solving the problem of burst atrial fibrillation detection based on electrocardiograms. In the field of mechanical fault diagnosis, Yan et al. [18] proposed a fusion method based on multi-resolution T-F spectrum segmentation and sparse decomposition of vibration signals. Compared with traditional Gabor T-F atoms, the new method has a faster pursuit speed of the best atom and higher approximation precision. Saulig et al. [19] combined K-means clustering with local Rényi entropy to distinguish the basic structural differences between helpful information content and noise component in the T-F plane, developing an unsupervised adaptive T-F analytical algorithm. In fact, the T-F analysis task of the planetary gearbox is particularly challenging. Because of the complex structure and time-varying characteristic of the planetary gearbox, the characteristic regions in the T-F spectrum are often intertwined with the background, or the contrast and sharpness may be very low. Therefore, traditional image processing methods are difficult for achieving satisfactory performance of such tasks.

Previous works have applied deep learning to planetary gearbox fault diagnosis based on T-F analysis. However, these studies usually regard T-F analysis and fault diagnosis as two separate tasks. Other works focus on developing an indivisible intelligent system to accomplish the gearbox fault diagnosis task based on T-F information [20–24]. This paper aims to incorporate T-F spectrum analysis and fault diagnosis based on T-F feature information into a continuous but stage-by-stage diagnosis framework. This is more similar to manual diagnosis processes. The specific indications of fault feature regions and the high-accuracy fault classification diagnosis can improve confidence in the fault diagnosis conclusion for operators because it provides richer and more comprehensive diagnosis information about the health status of equipment, making the diagnosis results more reliable.

Based on the above views, this paper proposes a two-stage diagnosis framework for planetary gearboxes. Because the traditional threshold-based algorithms are error prone for high-resolution T-F spectrum, we implement a fully convolutional network U-net for semantic segmentation; that is, to label the regions of interest (ROI) pixel by pixel. The U-net model highlights the pixels belonging to fault feature regions from the background. Then, the shape features are utilized to extract valuable information from the feature region distribution. Finally, the decision tree algorithm is used to determine the health status of the planetary gearbox.

The main contributions of this paper are summarized as follows:

1. Compared with previous studies that considered T-F analysis or fault diagnosis alone, a more comprehensive two-stage framework is proposed to combine the two tasks. To the best of our knowledge, we are the first paper to discuss planetary gearbox fault diagnosis based on T-F information from the perspective of an analysis and diagnosis two-stage task.
2. The modified U-net model is adopted to process T-F image, which can utilize large-range context information to improve the segmentation accuracy and enhance the robustness against environmental variations.
3. The effectiveness of the proposed method is verified by using the real data of in-service wind turbines. T-F images and ground truth labels are available online [25].

The remainder of this paper is organized as follows. Section 2 briefly introduces the U-net, feature extraction method, and decision tree algorithm used in the research. Section 3 introduces the two-stage gearbox fault diagnosis framework proposed in this paper. Section 4 describes the experimental dataset and analyses the results of running our method on this dataset. Conclusions are outlined in Section 5.

2. Preliminaries

2.1. U-Net

In 2015, Ronneberger et al. [26] first proposed the U-net model for the problem of cell image segmentation. As an important variant of fully convolutional neural network, U-net does not contain fully connected layers but uses symmetrically arranged convolutional layers to compress and reconstruct feature information, thereby exploiting context information at multiple scales and generating pixel-level image segmentation results. So far, U-net and its variants have seen the most applications in the field of medical image processing [27] and are expanding into machine vision [28], image-based fault detection [29], etc. Figure 1 illustrates the modified U-net architecture used in this paper.

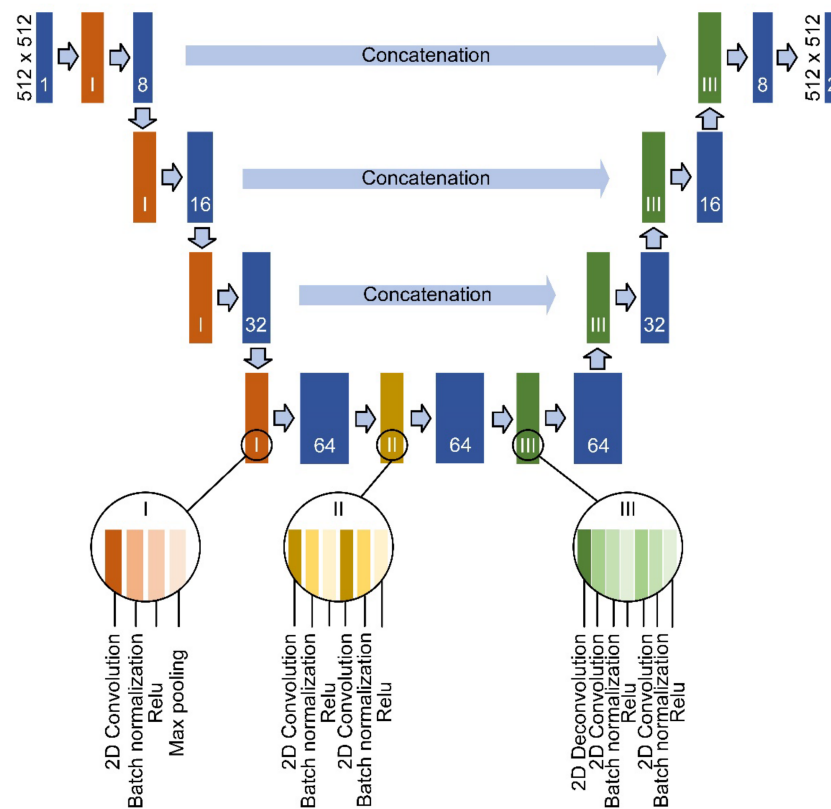


Figure 1. U-net structure.

The left half of the U-net is the contracting path, also known as the encoding path, which consists of four repeated contracting operators. The contracting operator has the typical structure of a convolutional neural network, including a 3×3 convolutional layer, a batch normalization layer, followed by a rectified linear unit (ReLU), and a 2×2 pooling layer. The right half of the U-net is the expanding path, also known as the decoding path. Its overall structure is very similar to the contracting path, except with its more complex expanding operators. Some of the most crucial changes include deconvolution instead of convolution and the addition of concatenation from the contracting path. Based on the above design, U-net presents a simple and elegant U-shaped structure, which is also the origin of its name.

In order to explain the U-net structure in more detail, several critical components in the network are described next:

1. The convolutional layer is composed of a set of convolutional kernels that can be set high and wide. The learnable convolutional kernel enables the convolutional layer to generalize the feature information in the input and map it into a new feature space. In addition, the setting of shared weights enables the convolutional layer to have lower computational complexity than the fully connected layer. For each convolutional kernel, its output can be expressed as:

$$O_i = f\left(\sum_j \omega_{ij} \times X_j + b_i\right) \quad (1)$$

where X_j is the j th input channel, O_i is the i th channel of the feature map, and $f(\bullet)$ is the activation function. ω_{ij} and b_i denote convolutional weight and bias term respectively, both of which are trainable parameters.

2. Deconvolution is also known as up-convolution or transposed convolution. Its function in the U-net model is to increase the resolution of feature maps rather than computing the true inverse of convolution. In order to obtain an appropriate expansive capability matching the contracting path, we use a 2×2 deconvolutional kernel and set the deconvolutional stride to 2.
3. Activation functions enable convolutional neural networks to model nonlinear mapping hidden in data. Common activation functions include Sigmoid, Tanh, ELU, ReLU [30], etc. As a non-saturated activation function, ReLU can alleviate the problems of gradient vanishing and exploding while accelerating model learning. Therefore, ReLU is selected as the activation function in the contracting operator and expansive operator.

$$\text{ReLU}(x) = \max(0, x) \quad (2)$$

4. Batch normalization is an effective tool to deal with the problem of feature distribution drift during batch training. Severe feature distribution drift will reduce the stability of neural network training and aggravate the over-fitting issue, which is particularly obvious in deep neural networks. Batch normalization transforms the distribution of neuron activations into the standard normal distribution. For batch training data $[x_1, x_2, \dots, x_m]$, the normalized result is calculated as follows:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (3)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (4)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (5)$$

$$y_i = \gamma \hat{x}_i + \beta \quad (6)$$

where μ_B and σ_B^2 denote the mean and variance of batch training data, respectively. \hat{x}_i is the normalized result after subtracting the mean and variance of batch training data, and y_i is the normalized result calculated based on the scaling factor γ and translation factor β .

2.2. Shape Feature

The segmented T-F image can be regarded as a binary image. It has the same resolution as the input image, so there is a data redundancy for the subsequent fault classification algorithm. Therefore, this paper innovatively uses a region-based shape feature extraction method to reduce the dimensionality of the data.

Geometric moment [31] is a concise and effective region-based shape feature. For a general function $f(x,y)$, its (p,q) -order geometric moment is defined as follow:

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy, \quad p, q = 0, 1, 2, \dots \quad (7)$$

Some important properties of shapes can be derived from geometric moments:

M_{00} defines the mass of a shape.

$(M_{10}/M_{00}, M_{01}/M_{00})$ defines the centroid of a shape.

(M_{20}, M_{02}) defines the moments of inertia of a shape, which describes the mass distribution of the shape relative to the coordinate axes.

The above properties are useful for representing the distribution of characteristic patches in the segmented T-F image.

2.3. Decision Tree

The decision tree method introduced by Quinlan [32] is a powerful machine learning algorithm that constructs a knowledge-based leaf-branch system by inductive inference from historical data. So far, researchers have proposed various algorithms to induct the decision tree-based diagnosis model, such as CART [33], BOAT [34], and SPRINT [35]. This paper selected CART because of its popularity and simplicity. Generally, developing a fault classifier with CART can be summarized into two phases: the building phase and the pruning phase.

1. Building phase: The CART-based decision tree is a binary tree. Namely, each split generates exactly two branches. A test attribute x and a test threshold t_x can divide the training set into two subsets. In the CART model, the attribute-threshold domain (X, T) is searched to obtain the combination that produces the purest subset. This process is repeated many times to segment subsets and subsets of subsets until the algorithm cannot put forward a new segmentation to obtain higher purity subsets or the preset maximum depth is reached. A fully grown, binary, tree-like structure makes the identification of crucial variables quite easy.
2. Pruning phase: If the decision tree grows to the maximum size without restriction, developing as a nonparametric model, there will usually be an overfitting problem which will reduce the accuracy of the decision tree in the whole instance space. Therefore, it is necessary to delete unreliable branches or limit the model's degrees of freedom. Setting the maximum depth of the tree structure is the most common pruning method. The other choices of hyperparameters include the minimum number of samples per leaf node, the maximum number of leaf nodes and the complexity parameter. For more information about the CART algorithm, see [36–38].

3. The Proposed Method

3.1. Architecture of the Two-Stage Framework

This paper aims to develop a general tool for automatic analysis and diagnosis of planetary gearbox failures from T-F data. For this purpose, a two-stage framework for T-F spectrum analysis and fault diagnosis is proposed in this paper. The first stage uses U-net to label the characteristic patches representing faults in the T-F spectrum, also known as semantic segmentation. Then, the well-trained U-net is used as a segmentation tool, and geometric moment features in the segmented T-F image are extracted to facilitate the second stage calculation. In the second stage, a decision tree is trained to analyze the shape features and determine the health status of the gearbox. The overall architecture of the two-stage framework is shown in Figure 2.

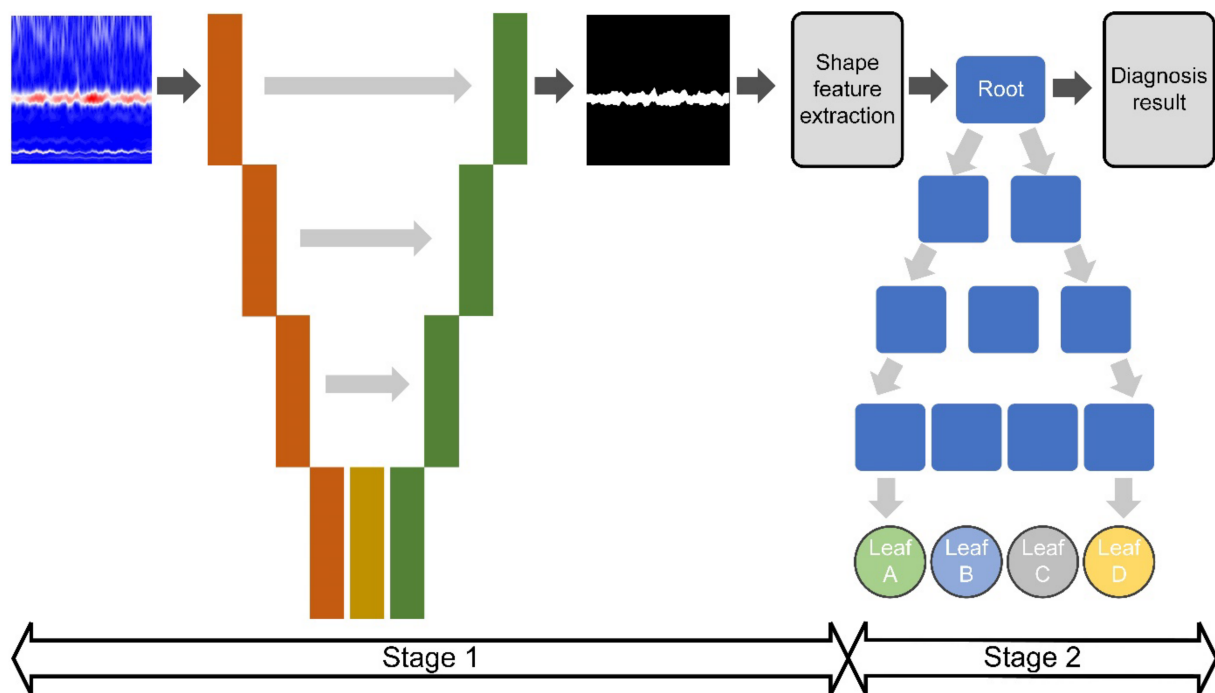


Figure 2. Architecture of the proposed two-stage framework.

The proposed framework is built using Python machine learning library Pytorch [39]. The structure of U-net mainly refers to Ronneberger's paper and some modifications have been made to adapt to the T-F analysis task, including the redesign of the network structure and the removal of the overlap-tile input strategy, etc. The decision tree algorithm is mainly based on the DecisionTreeClassifier model in Scikit-Learn package [40]. Because of its outstanding computational efficiency, Adam is used as the optimization algorithm for the U-net model. The learning rate is set to 1×10^{-3} , and the weight decay rate is set to 1×10^{-5} . The model uses categorical cross-entropy for the loss function. The stopping criterion of training is set such that the loss on the validation set has not improved significantly for five successive epochs or reached the epoch maximum (50 epochs).

3.2. Evaluation Metrics of Performance

In this paper, three evaluation metrics are used to quantitatively evaluate the ability of different algorithms to label ROI in T-F spectrum, including precision (P), recall (R), and F1-score (also known as Dice coefficient). The joint use of three indicators helps us to analyze the algorithm performance from a more comprehensive perspective: precision reflects the ability of the model to accurately select valuable information, recall reflects

the ability of the model to avoid missing valuable information, and F1-score reflects the coordination between the above two abilities, obviously a more rigorous metric.

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (10)$$

where TP represents the number of pixels belonging to true positive, FP represents false positive and FN represents false negative.

4. Case Study

4.1. Data Collection and Augmentation

The vibration signals used in the case study were collected from 1.5 MW pitch-controlled wind turbines located on a wind farm in northeast China, as shown in Figure 3. The transmission chain mainly includes a bladed rotor, a planetary gearbox with a ratio of 100.48:1, and a doubly-fed induction generator. The planetary gearbox contains two planetary transmission mechanisms and one parallel transmission mechanism. The accelerometer was attached to the ring of the second-stage planetary mechanism by magnetic. Figure 4 illustrates the internal structure of the gearbox and the location of the sensors. Table 1 lists teeth number information of the second-stage planetary mechanism.

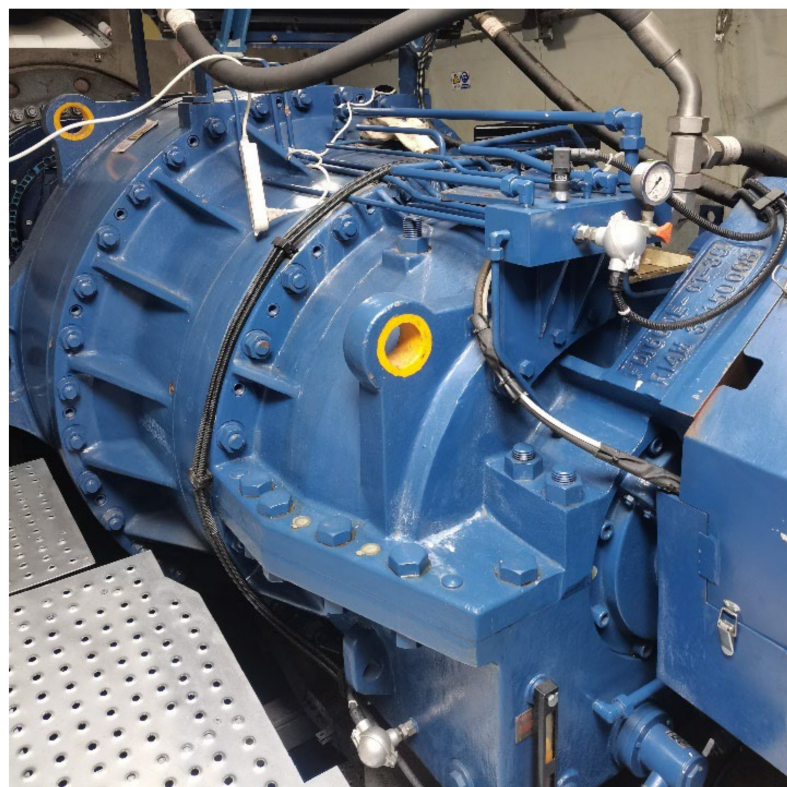


Figure 3. Planetary gearbox in the case study.

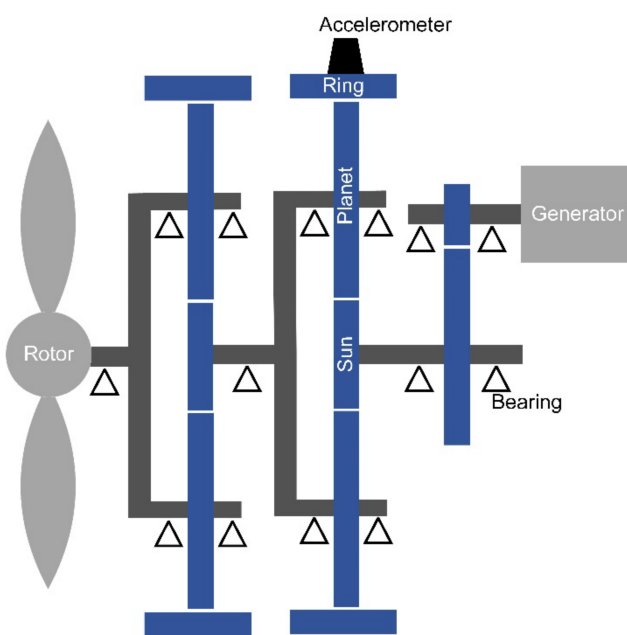


Figure 4. Internal structure of the planetary gearbox and the location of the accelerometer.

Table 1. Teeth number information of the second-stage planetary mechanism.

| Component | Planet | Ring Gear | Sun Gear |
|-----------|--------|-----------|----------|
| Teeth | 39 | 101 | 22 |

The case study involves four health conditions: normal (NS), ring gear tooth pitting (RP), planetary gear meshing misalignment (MA), and pitting-misalignment concurrence (CO). All the above failures occur in the second-stage planetary mechanism. Each health condition corresponds to four planetary gearboxes and, thus, there are 16 gearboxes used in this experiment. During the data collection, the speed of the gearboxes continuously varied. The characteristics of varying speed operation of wind turbines and changeable operating environment bring more challenges to fault analysis and diagnosis.

The piezoelectric accelerometer samples at a frequency of 16,384 Hz and each signal lasts for 10 s. Ten segments are randomly intercepted from each signal to calculate the T-F spectrum. In this paper, generalized S-transform [41] (generalized factor = 2) is used to generate 512×512 T-F images from signal segments. Figure 5 shows the vibration signal and T-F spectrums corresponding to the four health conditions.

The shortfall in dataset scale, accompanied by the possibility of overfitting, is one of the common problems in model training. $10 \times 4 \times 4 = 160$ samples are insufficient to train deep learning models in this case study. Therefore, the Augmentor toolkit [42] is utilized to mirror or distort the original image to obtain 800 additional training samples. After data augmentation, the total number of experimental samples reaches 960. Referring to similar studies [43], datasets of this scale can effectively inhibit over-fitting.

4.2. Performance Validation

In semantic segmentation methods, each pixel in the sample is given a label. The characteristic patches or bands in the T-F image are highlighted with a grayscale of 255. The grayscale of other parts is set to 0. The experimental data is divided into two non-overlapping parts in a ratio of 4:1, which are used as training and test datasets, respectively. The size of minibatch data in training is set at 24 to determine how many training samples jointly calculate a parameter update in the model. A computer (Intel Core i5-10400 CPU with 16 GB of RAM) and an NVIDIA GTX 1660 with 6 GB of GPU memory are used to conduct all experiments in this section.

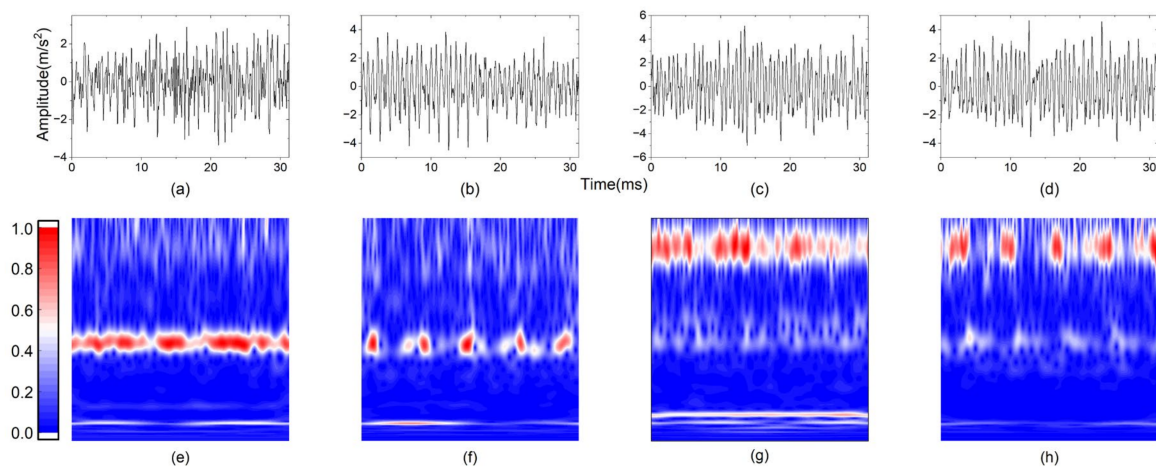


Figure 5. Vibration signals and T-F spectrums in the four health conditions: (a,e) NS marked by characteristic band in meshing frequency bin; (b,f) RP marked by characteristic patches occurring with meshing frequency in meshing frequency bin; (c,g) MA marked by high energy in $2 \times$ meshing frequency bin; (d,h) CO marked by characteristic patches occurring with meshing frequency in $2 \times$ meshing frequency bin.

Figure 6 illustrates the results of identifying fault feature regions using the U-net method in test dataset. We can find that, in general, the U-net model can identify the characteristic patches or characteristic bands well for the test samples in four health states, revealing the gear meshing phenomenon or impact phenomenon contained in the T-F spectrum. Tooth surface pitting faults and pitting misalignment concurrent faults seem to be two more difficult cases. There are a small amount of noise pixels and error boundaries in the network output. One possible reason is that patch-like areas are easily confused with intense background noise. In order to illustrate the performance of U-net method and analyze the differences between four types of test samples, Table 2 lists the detailed evaluation metrics more specifically. The Precision, Recall, and F1-score of U-net are all satisfactory. The F1-score in the whole test dataset reaches 0.942. By comparison, the tooth surface pitting fault and pitting misalignment concurrent fault have poor performance, with F1-score of 0.891 and 0.87, respectively. These are consistent with our observations in Figure 6.

Table 2. Performance metrics of the U-net.

| Data | Precision | Recall | F1-Score |
|--------------|-----------|--------|----------|
| Test Dataset | 0.933 | 0.951 | 0.942 |
| NS | 0.943 | 0.953 | 0.948 |
| RP | 0.886 | 0.897 | 0.891 |
| MA | 0.955 | 0.955 | 0.955 |
| CO | 0.801 | 0.952 | 0.87 |

Furthermore, the Otsu method and Fourier filtering are used to process the samples in the test dataset to conduct the performance comparison between the proposed method and traditional methods, as shown in Figure 7. The time consumption to process each T-F image is also shown in the figure. We find that the U-net method is more successful than the two classical algorithms thanks to its strong feature learning ability and context information processing ability. Although the Otsu method gives the highest computational efficiency with a time consumption per sample of approximately 6 ms, there is an apparent imbalance between its Precision and Recall metrics, suggesting too many false positive pixels in the results. The Fourier filtering method shows better performance than the Otsu method, with F1-score reaching 0.858. However, because the size of the mask in the filter needs to

be optimized during image processing, the computational efficiency of Fourier filtering is degraded. U-net, as a deep learning method, can automatically capture multi-scale features without manual tuning of model parameters. A trained model, in addition, can process T-F samples quickly, which meets the real-time requirements in fault analysis. In conclusion, as the first stage of the proposed framework, the U-net method is competent for the T-F analysis task among planetary gearbox fault diagnosis.

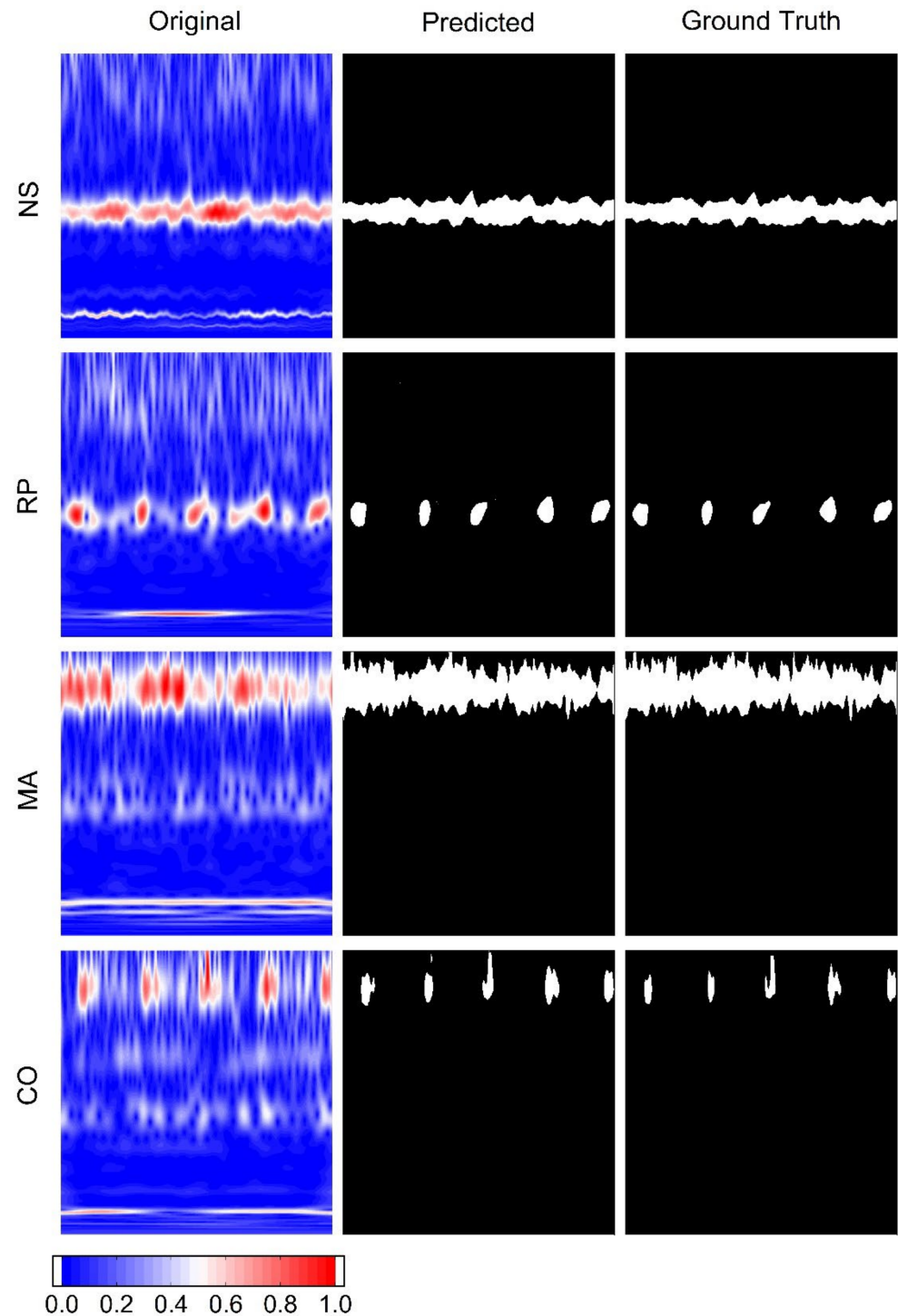


Figure 6. The results of identifying fault feature regions using U-net. White represents the characteristic area, and black represents the background.

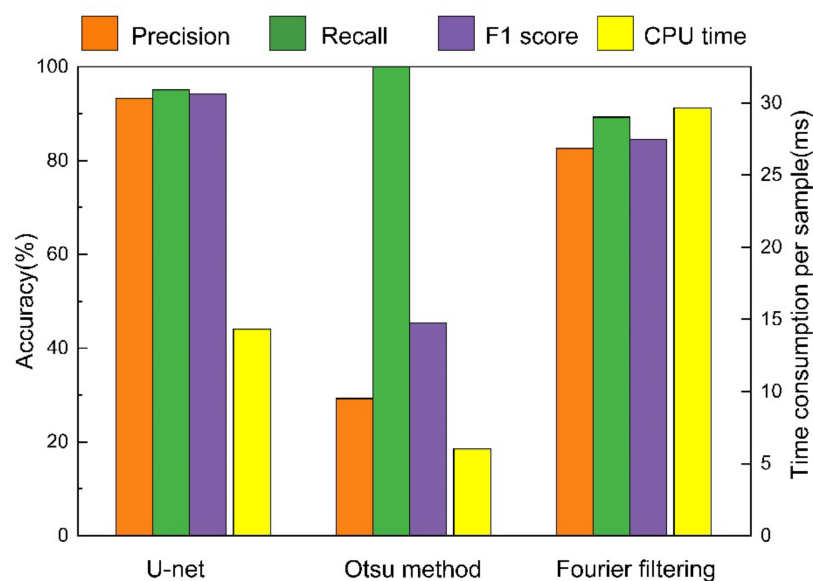


Figure 7. Comparison of performance and time consumption between different methods.

Besides analyzing the T-F information, the ultimate goal of presenting a high-accuracy analytical method is to realize the fast classification of planetary gearbox faults. The labeled fault feature regions are used in the second stage of the proposed framework. This paper uses three geometric moments to extract valuable information from segmented images, including image mass M_{00} and moment of inertia in two directions (M_{20} , M_{02}). In our conception, the former is defined to distinguish between characteristic patches and characteristic bands, while the latter two judge the relative position in the T-F spectrum.

The significantly compressed feature space makes it possible to use a simple classifier. In this experiment, we choose the decision tree algorithm to classify the four health states of planetary gearboxes quickly. The simplicity of the model, convenience of training, and high interpretability of the model are several factors that we focus on. Obviously, the decision tree algorithm is an excellent choice to reveal the faulty information in T-F images.

Figure 8 shows the classification confusion matrix of the decision tree algorithm on the test dataset. Overall, the average accuracy of the diagnosis reaches 97.4%. The limited amount of real data and small feature space have no adverse impact on the diagnosis results. Specifically, meshing misalignment fault (MA) and pitting-misalignment concurrent fault (CO) are easy to confuse because they both have the characteristics of high energy of double meshing frequency bins. Ring gear tooth pitting (RP) fault is a health condition with the lowest diagnosis accuracy. Referring to the observation in the previous subsection, it is more polluted by noise. Figure 9 shows the trained decision tree in the experiment. In summary, the decision tree algorithm combined with the shape feature extraction method can effectively extract the status features in T-F images and accurately classify the health states of planetary gearboxes.

4.3. Robustness against Variable Operating Environment

In the engineering application of fault diagnosis, the variable operating environment greatly influences the system's performance. For example, wind turbines operate under continuous variable speed and variable loads. In the same wind farm, the operating conditions of wind turbines differ by local geographical environment. In addition, there are more difficult diagnostic scenarios, such as the same type of planetary gearboxes mounting in different wind farms. In these cases, diagnostic systems require high robustness against environmental changes. In this sub section, we sequentially take a subset belonging to one wind turbine from the original dataset as the robustness test data and use the remaining data to train the two-stage framework as described in Section 4.1. The performance metrics

and fault classification accuracy of the trained model on the robustness test dataset are shown in Figure 10.

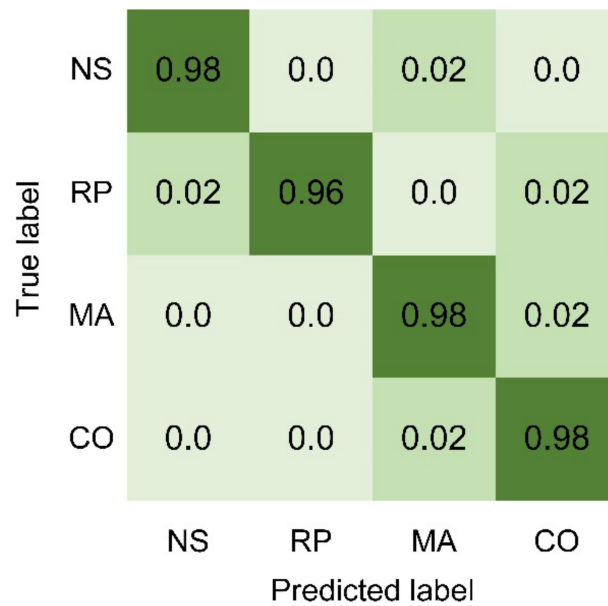


Figure 8. Classification confusion matrix.

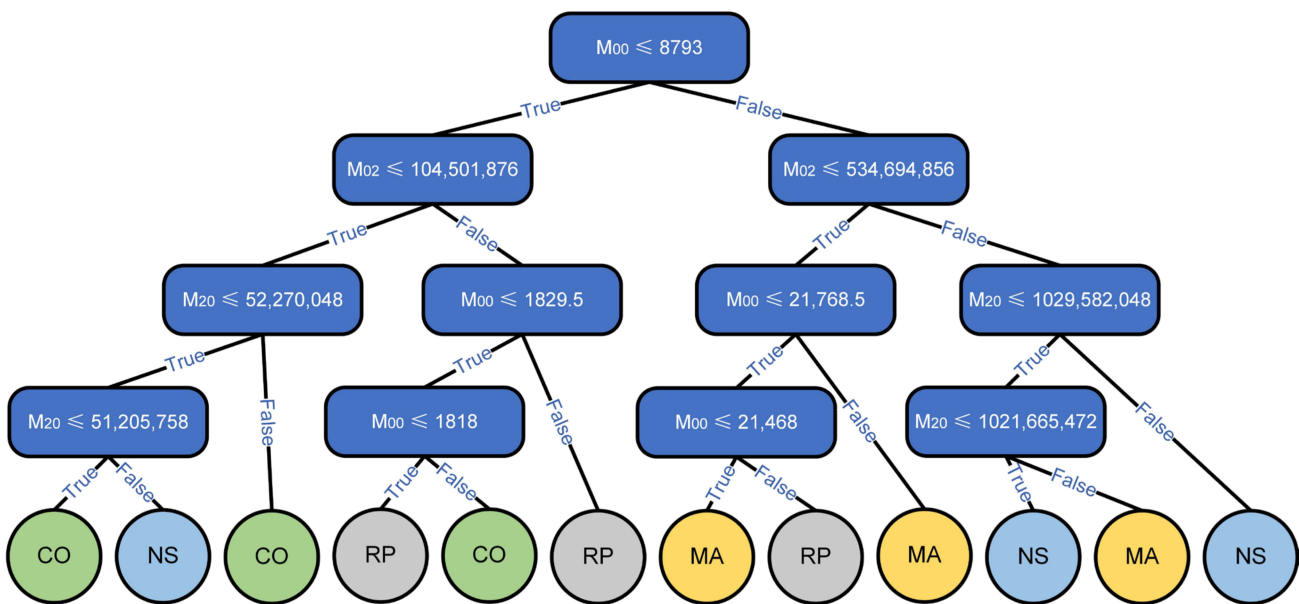


Figure 9. Trained decision tree in the experiment.

We can observe that the segmentation and classification performance in each healthy state is affected by the variable operating environment in some way. The proposed two-stage framework achieves an average F1 score of 0.747 and a classification accuracy of 83.1% on the robustness test dataset. The normal state gives the best segmentation and classification performance on the robustness test dataset, reaching an average F1 score of 0.91 and a classification accuracy of 99.18%. Pitting-misalignment concurrent faults gives the worst robustness. A possible reason is that more complex fault causes make less similarity between different machines. In addition, the experimental results also show that there is no dependence of the classification accuracy on a certain segmentation performance metric when surveying four health states, which indicates that the influence of errors in the analytical stage on the geometric moments is uncertain. To sum up, the pro-

posed framework presents satisfactory robustness against variable operating environments. Therefore, it is suitable for application in the actual industrial environment with variable operating environments.

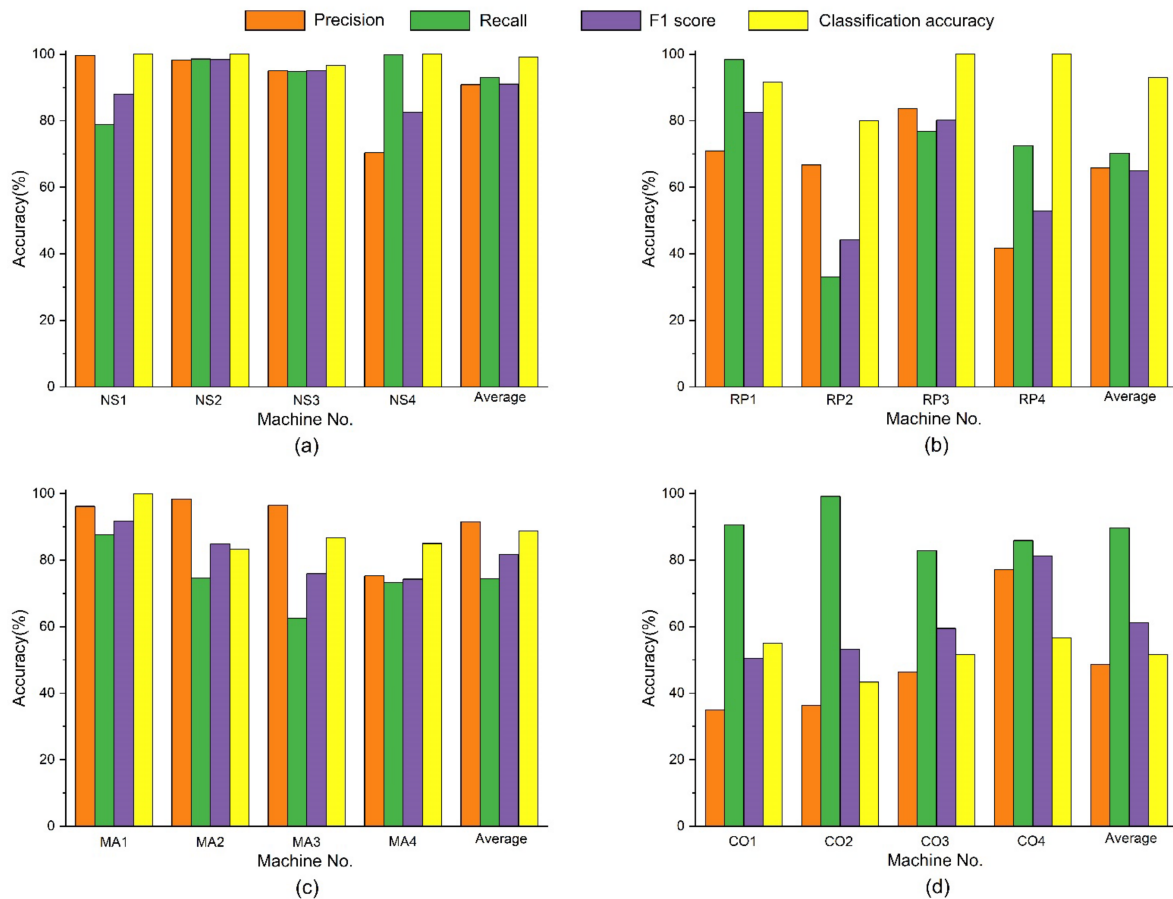


Figure 10. The segmentation performance metrics and fault classification accuracy of the trained model on the robustness test dataset: (a) NS; (b) RP; (c) MA; (d) CO.

5. Conclusions

This paper presents a two-stage fault diagnosis framework for planetary gearboxes, including a U-net based T-F spectrum analytical method in the first stage and a decision tree-based health state classification method in the second stage. Both the specific indications of fault feature regions and the white-box property of the decision tree model can improve confidence in the fault diagnosis conclusion for operators, which is lacking in conventional deep learning models. Using the real data collected from wind turbines, we prove that the U-net model is better than the traditional image processing methods in the T-F spectrum analytical task of planetary gearboxes with a F1-score of 0.942. With a decision tree classifier, the final accuracy of planetary gearbox fault diagnosis reaches 97.4%. Further experiments show that the proposed framework presents high robustness against variable operating environment. Moreover, it is worth noting that our method provides a flexible paradigm for solving similar problems with high openness.

In future work, the authors plan to improve the proposed method from two aspects. First, knowledge generation based on interpretable machine learning is a promising direction for improvement. Second, the authors will follow up on advanced semantic segmentation methods to enhance the robustness of time-frequency analysis.

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

Funding: This research was supported by the National Natural Science Foundation of China (No. 51675350) and the Natural Science Foundation of Liaoning Province (No. 2019MS245).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available online at: <https://iee-dataport.org/documents/vibration-time-frequency-images-planetary-gearboxes> (accessed on 2 May 2022).

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

References

1. Zhao, M.; Kang, M.; Tang, B.; Pecht, M. Deep Residual Networks with Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4290–4300. [CrossRef]
2. Wang, T.; Han, Q.; Chu, F.; Feng, Z. Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review. *Mech. Syst. Signal Process.* **2019**, *126*, 662–685. [CrossRef]
3. Feng, Z.; Gao, A.; Li, K.; Ma, H. Planetary gearbox fault diagnosis via rotary encoder signal analysis. *Mech. Syst. Signal Process.* **2021**, *149*, 107325. [CrossRef]
4. Fu, Y.; Liu, Y.; Yang, Y. Multi-Sensor GA-BP Algorithm Based Gearbox Fault Diagnosis. *Appl. Sci.* **2022**, *12*, 3106. [CrossRef]
5. Hameed, S.S.; Muralidharan, V.; Ane, B.K. Comparative analysis of fuzzy classifier and ANN with histogram features for defect detection and classification in planetary gearbox. *Appl. Soft Comput.* **2021**, *106*, 107306. [CrossRef]
6. Liu, T.; Cui, L.; Zhang, J.; Zhang, C. Research on fault diagnosis of planetary gearbox based on variable multi-scale morphological filtering and improved symbol dynamic entropy. *Int. J. Adv. Manuf. Technol.* **2022**, *124*, 3947–3961. [CrossRef]
7. Malla, C.; Panigrahi, I. Review of Condition Monitoring of Rolling Element Bearing Using Vibration Analysis and Other Techniques. *J. Vib. Eng. Technol.* **2019**, *7*, 407–414. [CrossRef]
8. Chen, X.; Feng, Z. Iterative generalized time-frequency reassignment for planetary gearbox fault diagnosis under nonstationary conditions. *Mech. Syst. Signal Process.* **2016**, *80*, 429–444. [CrossRef]
9. Han, S.; Feng, Z. Deep subclass alignment transfer network based on time-frequency features for intelligent fault diagnosis of planetary gearboxes under time-varying speeds. *Meas. Sci. Technol.* **2022**, *33*, 105010. [CrossRef]
10. Yuan, H.; Chen, J.; Dong, G. Machinery fault diagnosis based on time-frequency images and label consistent K-SVD. *Proc. Inst. Mech. Eng. Part C J. Eng. Mech. Eng. Sci.* **2018**, *232*, 1317–1330. [CrossRef]
11. Tu, X.; Hu, Y.; Abbas, S.; Li, F. Generalized wavelet-based synchrosqueezing transform: Algorithm and applications. *Struct. Health Monit.* **2020**, *19*, 2051–2062. [CrossRef]
12. Dhamande, L.S.; Chaudhari, M.B. Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. *Measurement* **2018**, *125*, 63–77. [CrossRef]
13. Sharma, A.; Kumar, S.; Singh, S.N. Brain tumor segmentation using DE embedded OTSU method and neural network. *Multidimens. Syst. Signal Process.* **2019**, *30*, 1263–1291. [CrossRef]
14. Martinez-Aroza, J.; Gomez-Lopera, J.F.; Blanco-Navarro, D.; Rodriguez-Camacho, J. Clustered entropy for edge detection. *Math. Comput. Simul.* **2021**, *182*, 620–645. [CrossRef]
15. Yang, W.; Cai, L.; Wu, F. Image segmentation based on gray level and local relative entropy two dimensional histogram. *PLoS ONE* **2020**, *15*, e0229651. [CrossRef]
16. Zhang, Y.; Li, J.; Wei, S.; Zhou, F.; Li, D. Heartbeats classification using hybrid time-frequency analysis and transfer learning based on ResNet. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 4175–4184. [CrossRef]
17. Cheng, P.; Chen, Z.; Li, Q.; Gong, Q.; Zhu, J.; Liang, Y. Atrial fibrillation identification with PPG signals using a combination of time-frequency analysis and deep learning. *IEEE Access* **2020**, *8*, 172692–172706. [CrossRef]
18. Yan, B.; Wang, B.; Zhou, F.; Li, W.; Xu, B. Sparse decomposition method based on time-frequency spectrum segmentation for fault signals in rotating machinery. *ISA Trans.* **2018**, *83*, 142–153. [CrossRef] [PubMed]
19. Saulig, N.; Milanovic, Z.; Ioana, C. A local entropy-based algorithm for information content extraction from time-frequency distributions of noisy signals. *Digit. Signal Prog.* **2017**, *70*, 155–165. [CrossRef]
20. Wang, Y.; Fan, Z.; Liu, H.; Gao, X. Planetary Gearbox Fault Diagnosis Based on ICEEMD-Time-Frequency Information Entropy and VPMCD. *Appl. Sci.* **2020**, *10*, 6376. [CrossRef]
21. Zhao, D.; Wang, T.; Chu, F. Deep convolutional neural network based planet bearing fault classification. *Comput. Ind.* **2019**, *107*, 59–66. [CrossRef]

22. Ma, S.; Chu, F.; Han, Q. Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions. *Mech. Syst. Signal Process.* **2019**, *127*, 190–201. [[CrossRef](#)]
23. Chen, R.; Huang, X.; Yang, L.; Xu, X.; Zhang, X.; Zhang, Y. Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Comput. Ind.* **2019**, *106*, 48–59. [[CrossRef](#)]
24. Wang, H.; Xu, J.; Sun, C.; Yan, R.; Chen, X. Intelligent Fault Diagnosis for Planetary Gearbox Using Time-Frequency Representation and Deep Reinforcement Learning. *IEEE ASME Trans. Mechatron.* **2022**, *27*, 985–998. [[CrossRef](#)]
25. Vibration Time-Frequency Images of Wind Turbine Planetary Gearboxes. Available online: <https://iee-dataport.org/documents/vibration-time-frequency-images-planetary-gearboxes> (accessed on 2 May 2022).
26. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 5–9 October 2015.
27. Falk, T.; Mai, D.; Bensch, R.; Cicek, O.; Abdulkadir, A.; Marrakchi, Y.; Bohm, A.; Deubner, J.; Jackel, Z.; Seiwald, K.; et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **2019**, *16*, 67–72. [[CrossRef](#)] [[PubMed](#)]
28. Liu, Z.; Cao, Y.; Wang, Y.; Wang, W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom. Constr.* **2019**, *104*, 129–139. [[CrossRef](#)]
29. Gao, K.; Huang, L.; Zheng, Y. Fault Detection on Seismic Structural Images Using a Nested Residual U-Net. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4502215. [[CrossRef](#)]
30. Lau, M.; Lim, K. Review of Adaptive Activation Function in Deep Neural Network. In Proceedings of the IEEE-EMBS Conference on Biomedical Engineering and Sciences, Sarawak, Malaysia, 3–6 December 2018.
31. Zhang, D. Shape Representation. In *Fundamentals of Image Data Mining*, 2nd ed.; Springer: Gewerbestrasse, Switzerland, 2021; pp. 133–175.
32. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, CA, USA, 1993; pp. 1–10.
33. Breiman, L.; Friedman, J.; Olshen, R. *Classification and Regression Trees*; Wadsworth: Monterey, CA, USA, 1984; pp. 1–5.
34. Gehrke, J.; Ganti, V.; Ramakrishnan, R.; Loh, W.Y. BOAT-Optimistic decision tree construction. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Philadelphia, PA, USA, 31 May–3 June 1999.
35. Shafer, J.; Agrawal, R.; Mehta, M. SPRINT: A scalable parallel classifier for data mining. In Proceedings of the 22nd International Conference on Very Large Data Bases, Mumbai, India, 3–6 September 1996.
36. Xu, Y.; Shmaliy, Y.S.; Hua, L.; Ma, L.; Zhuang, Y. Decision tree-extended finite impulse response filtering for pedestrian tracking over tightly integrated inertial navigation system/ultra wide band data. *Meas. Sci. Technol.* **2021**, *33*, 034007. [[CrossRef](#)]
37. Young, W.A.; Weckman, G.R.; Hari, V.; Whiting, H.S.; Snow, A.P. Using artificial neural networks to enhance CART. *Neural Comput. Appl.* **2012**, *21*, 1477–1489. [[CrossRef](#)]
38. Luna, J.M.; Gennatas, E.D.; Ungar, L.H.; Eaton, E.; Diffenderfer, E.S.; Jensen, S.T.; Simone, C.B.; Friedman, J.H.; Solberg, T.D.; Valdes, G. Building more accurate decision trees with the additive tree. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19887–19893. [[CrossRef](#)] [[PubMed](#)]
39. Pytorch. Available online: <https://pytorch.org> (accessed on 16 June 2021).
40. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
41. Adams, M.D.; Kossentini, F.; Ward, R.K. Generalized S transform. *IEEE Trans. Signal Process.* **2002**, *50*, 2831–2842. [[CrossRef](#)]
42. Bloice, M.D.; Roth, P.M.; Holzinger, A. Biomedical image augmentation using Augmentor. *Bioinformatics* **2019**, *35*, 4522–4524. [[CrossRef](#)] [[PubMed](#)]
43. Groschner, C.K.; Choi, C.; Scott, M.C. Machine Learning Pipeline for Segmentation and Defect Identification from High-Resolution Transmission Electron Microscopy Data. *Microsc. Microanal.* **2021**, *27*, 549–556. [[CrossRef](#)] [[PubMed](#)]

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.