

Article

Rotating Machinery State Recognition Based on Mel-Spectrum and Transfer Learning

Fan Li ¹, Zixiao Lu ^{1,*}, Junyue Tang ^{2,*}, Weiwei Zhang ², Yahui Tian ³, Zhongyu Cui ², Fei Jiang ⁴, Honglang Li ¹ and Shengyuan Jiang ²

¹ National Center for Nanoscience and Technology, Beijing 100190, China

² The State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China

³ Institute of Acoustics Chinese Academy of Sciences, Beijing 100190, China

⁴ School of Mechatronics, Beijing Institute of Technology, Beijing 100081, China

* Correspondence: luzx2019@nanoctr.cn (Z.L.); tangjunyue@hit.edu.cn (J.T.)

Abstract: During drilling into the soil, the rotating mechanical structure will be affected by soil particles and external disturbances, affecting the health of the rotating mechanical structure. Therefore, real-time monitoring of the operational status of rotating mechanical structures is of great significance. This paper proposes a working state recognition method based on Mel-spectrum and transfer learning, which uses the mechanical vibration signal's time domain and frequency domain information to identify the mechanical structure's working state. Firstly, we cut the signal at window length, and then the Mel-spectrum of the truncated signal is obtained through the Fourier transform and Mel-scale filter bank. Finally, we adopted the method of transfer learning. The pre-trained model VGG16 is adjusted to extract and classify the features of the Mel-spectrum. Experimental results show that the framework maintains an accuracy of more than 90% for vibration signals under minor window conditions, which verifies the real-time reliability of the method.

Keywords: rotating machinery; state recognition; Mel-spectrum; transfer learning



Citation: Li, F.; Lu, Z.; Tang, J.; Zhang, W.; Tian, Y.; Cui, Z.; Jiang, F.; Li, H.; Jiang, S. Rotating Machinery State Recognition Based on Mel-Spectrum and Transfer Learning. *Aerospace* **2023**, *10*, 480. <https://doi.org/10.3390/aerospace10050480>

Academic Editor: M. Reza Emami

Received: 17 March 2023

Revised: 12 May 2023

Accepted: 15 May 2023

Published: 18 May 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

Operating mechanical structures, such as bearings, gears, and fans, are widely used in industrial manufacturing. As a vital power transmission mechanism, it plays an essential role in the regular operation of the mechanical system. When the rotating mechanical structure works in different states, its rotation and vibration change. By installing a vibration sensor on the structure's surface, we can obtain the vibration signals generated by it under different conditions. In the era of the intelligent industry, the generation of massive data and the development of artificial intelligence technology have promoted the development of industrial monitoring technology. Using historical monitoring data and artificial intelligence algorithms can classify vibration signals without manual judgment. Improving algorithm recognition accuracy, real-time, and practicality has become an important research goal.

The vibration signal can reflect the working state of the rotating mechanical structure. Therefore, the problem of working state recognition transforms into the problem of vibration signal classification. The typical task flow for the vibration signal classification problem includes data preprocessing, feature extraction, pattern recognition, and model application. Among them, feature extraction and algorithm identification are the core steps of the process. Signal features can be extracted from time, frequency, time-frequency, or other transformed domains. Time-domain features are extracted directly from the time series of vibration signals. Commonly used indicators include peak value, root means square value, variance, etc. B. Sreejith et al. used the normalized negative logarithm likelihood and kurtosis values in the time domain to classify the bearing vibration signal [1]. Tahsin Doguer et al. extracted 32 features from the time domain. They screened them

by studying the higher-order derivatives of the time series signal and some parameters that characterize the randomness of the peak position of the signal [2]. The frequency domain features commonly used techniques, such as Fourier transform, power spectrum, cepstrum, etc. C. Mishra et al. used a method based on envelope spectrum analysis and thresholded wavelet denoising based on the Sigmoid function to extract frequency domain features from bearing signals [3]. David et al. extracted the cepstrum from the faulty bearing's vibration and acoustic emission signals, performed whitening preprocessing, and compared the application effects of the two signals [4]. The time-frequency domain analysis method simultaneously gives information about the signal in the time and frequency domains. The commonly used methods include short-time Fourier transform, empirical mode decomposition, and wavelet transform. Gao et al. used the short-time Fourier transform to display the local information of the vibration signal of the fault bearing and then used the non-negative matrix decomposition (NFD) method to extract fault features from the time-frequency distribution [5]. Jaouher et al. used a feature extraction method based on empirical mode decomposition to obtain energy entropy and classify the vibration signals of rolling bearing units with defects [6]. In addition, there are other signal methods, such as the Hilbert–Huang transform and cosine transform, which can obtain the characteristics of signals in different transform domains [7,8]. Since time and frequency domains can reflect the vibration signal containing fault information, it is necessary to consider the real-time changes in the information over time. The short-time Fourier transform can convert the original signal into a time-frequency spectrum, which can reflect the change information of the signal in the time domain and frequency domain. For the vibration signal studied in this paper, the low frequency of the signal can reflect the working state more. Compared with the traditional time-frequency diagram, the Mel-spectrum has a higher resolution in the low-frequency band. This helps to magnify the distinction of different kinds of signals and improve the recognition performance of the classification model. So this paper considers using the Mel-spectrum as the feature of the vibration signal.

The purpose of feature extraction is to provide a dataset for pattern recognition. The pattern recognition methods on vibration signals consist of machine learning and deep learning. Machine learning generally constructs features from the original signal to form a new data set and then uses machine learning algorithms and data sets for model training and testing. Commonly used machine learning algorithms include KNN, support vector machine (SVM), etc. Dong et al. used the particle swarm algorithm to optimize the KNN model optimization and used the Shannon entropy obtained from the local average decomposition as a feature for bearing fault diagnosis [9]. D. H. Pandya et al. used the APF-KNN algorithm based on the non-axisymmetric approximation function as the classifier. They used the Hilbert–Huang transform technology to extract the features of the acoustic emission signal for the pattern recognition [10]. Li et al. introduced a binary tree-based SVM as a pattern recognition method and used multiscale permutation entropy as a feature to realize fault recognition [11].

However, machine learning methods have a strong dependence on manual feature extraction. The birth of deep learning has solved this problem very well. The convolutional neural network is potent in deep learning, which can automatically extract features and perform pattern recognition directly from the original signal. It showed high performance in the identification of one-dimensional time series signals and two-dimensional image signals. Wang and Huang et al. embedded a layer of singular value decomposition (SVD) in the one-dimensional convolutional network to avoid the loss of signal feature information during the training [12]. Amin Khorram et al. used a new type of convolutional long-short-term memory neural network to diagnose bearing faults, avoiding the preprocessing step of the original signal [13]. Wang et al. used the short-time Fourier transform to obtain the time-frequency map and used the convolutional neural network for the classification [14]. The Mel-spectrum extracted in this paper is a two-dimensional image. The convolution kernel can scan the image and automatically extract sufficient information from the image, thereby replacing manual operation. At present, many teams in the world have built and

trained convolutional neural networks and achieved superior performance on large image datasets. Using a pre-trained network model to classify images saves effort compared to building a convolutional neural network from scratch. The idea used in it is transfer learning [15]. Among them, the VGG16's high accuracy on significant image datasets classification problems verifies its power. Its model structure is straightforward. Therefore, this paper first uses the pre-trained convolutional neural network VGG16 to extract and recognize image features and then compares it with the other network models. We use the accuracy, precision, recall, and F1-score to evaluate the effect of the model and draw the confusion matrix to observe the model's classification performance on each category of signals. In this paper, the idea of transfer learning is applied to the process of identifying the working state of the drill pipe, which is helpful to reduce the steps of model construction compared with the traditional method and achieve high accuracy.

2. Data Collection and Preprocessing

Vibration sensing technology is a standard method to detect the state of structures and has a wide range of applications in detecting internal or external disturbances, such as structural cracks, wear, knocking, etc. The function of rotary mechanical structure is for soil drilling. In order to obtain the vibration signals of the motor under different working conditions, this paper builds a vibration measuring platform. It attaches a sensor to the vibration measuring platform to collect the signals generated by the vibration of the drill pipe. The vibration-measuring platform consists of a fixture, a motor, a drill pipe, a motor rack, and a platform rack. The container with soil is below the drill pipe. Figure 1 shows the platform. We installed the sensor on the motor. This structure can set different drilling procedures during the drilling process, such as rotation speed, impact, and other methods.

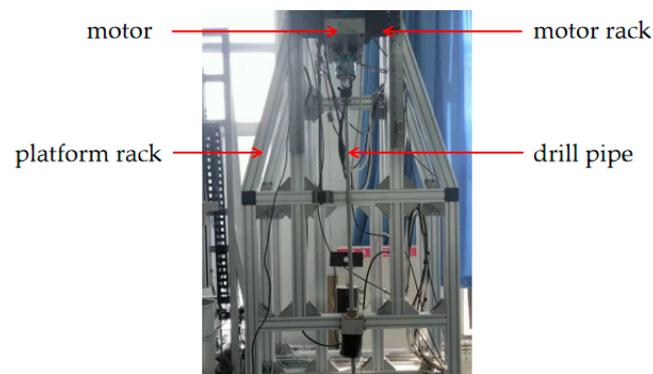
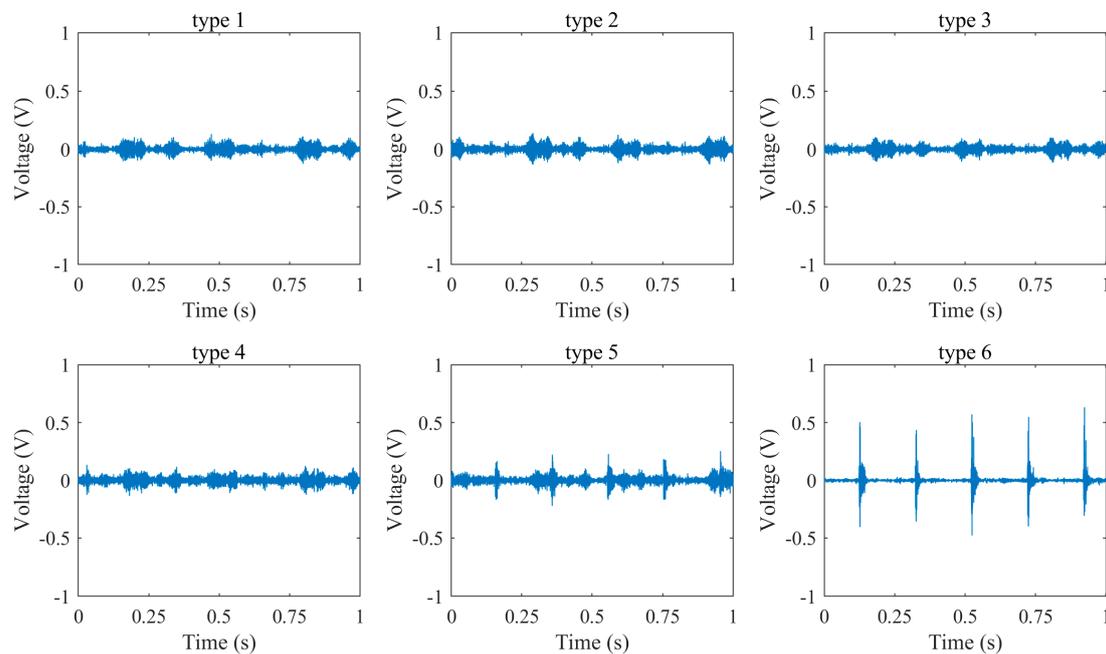


Figure 1. Vibration measuring platform.

In the process of soil drilling, drilling procedures and external interference mainly determine the working state of the drilling rig. In this paper, we set up six drilling procedures for the mechanism. Table 1 shows the details. Every kind of procedure corresponds to different Vibration signals. The experiment was repeated thrice in each working state, and we obtained a signal of 120 s in one experiment. In order to realize the real-time identification of the working state of the drilling rig, we divide the original signal into a signal of 1 s, and the overlapping time of the signal is 0.4 s. So, the delay of a single prediction is 0.6 s. All categories sum up to 3582 samples, and each has 597 samples. Figure 2 shows the six kinds of truncated signals. Type 1–4 in Figure 2 shows the aperiodic fluctuations of the signal of the rotating mechanism when the rotating mechanism is idling, turning, or applying artificial percussion. Time domain signals are difficult to distinguish. Type 5 and 6 in Figure 2 show periodic peaks under the impact state, which reflects the intermittency.

Table 1. Drill Procedure Description.

Type	State	Description
1	Idling	The drill rotates but does not touch the soil.
2	Idling + knocking platform rack	Knock the surface of the test bench while idling to simulate disturbances to the support facility.
3	Idling + knocking motor rack	Knock the motor rack while idling to simulate a motor being disturbed
4	Rotation (120 rpm)	The drill enters the soil and rotates smoothly, and advances at a speed of 120 rpm
5	Rotation (120 rpm) + Impact	Add impact based on 120 rpm rotation to simulate accelerated digging state.
6	Rotation (60 rpm) + Impact	Add impact based on 60 rpm rotation

**Figure 2.** Six types of truncated signals.

3. Feature Extraction and Recognition

3.1. Mel-Spectrum

There are many signal feature extraction methods, such as time domain, frequency domain, and time-frequency domain. Figure 2 shows that the vibration signal contains discriminative information in both the time domain and the frequency domain. Therefore, the feature extraction method in the time-frequency domain is adopted. Mel-spectrum is a powerful tool for classifying audio signals, including speech, music, and vibration signals. The calculation of the Mel-spectrum requires the operation of frame division, windowing, Fourier transform, Mel filter bank and logarithmic transformation. Firstly, the signal is divided into several frames according to a certain length of time, and then the window function is used to process the signal in each frame. The frame signal processed by the window function is converted into a frequency domain signal by Fourier transform. Then it is filtered through the Mel filter bank and converted into the Mel frequency domain. Finally, the log of the output of each Mel filter is calculated to get the Mel spectrum. Mel-spectrum is a powerful tool for classifying audio signals, including speech, music, and vibration signals. The Mel-spectrum is similar to the spectrogram obtained by the short-time Fourier transform, except that the Mel-spectrum converts the ordinary frequency scale into the Mel-frequency scale. As shown in the following formula:

$$\text{mel} = 2595 \lg(1 + f/700) \quad (1)$$

The Mel-frequency scale can show higher resolution to the low-frequency band of the signal. Figure 3 shows the Mel-spectrum of six kinds of truncated signals. The Mel spectrogram shows that the energy of the vibration signal is mainly around 2.1 kHz and 0.5 kHz, respectively. The difference between the first three pictures is weak, while the last three pictures are more pronounced. The influence of external interference on vibration signals is mainly concentrated in the low-frequency band, and the Mel spectrum improves the signal resolution in the low-frequency band through the Mel filter bank and logarithmic transformation, which will be beneficial to the feature extraction and recognition accuracy of the neural network model.

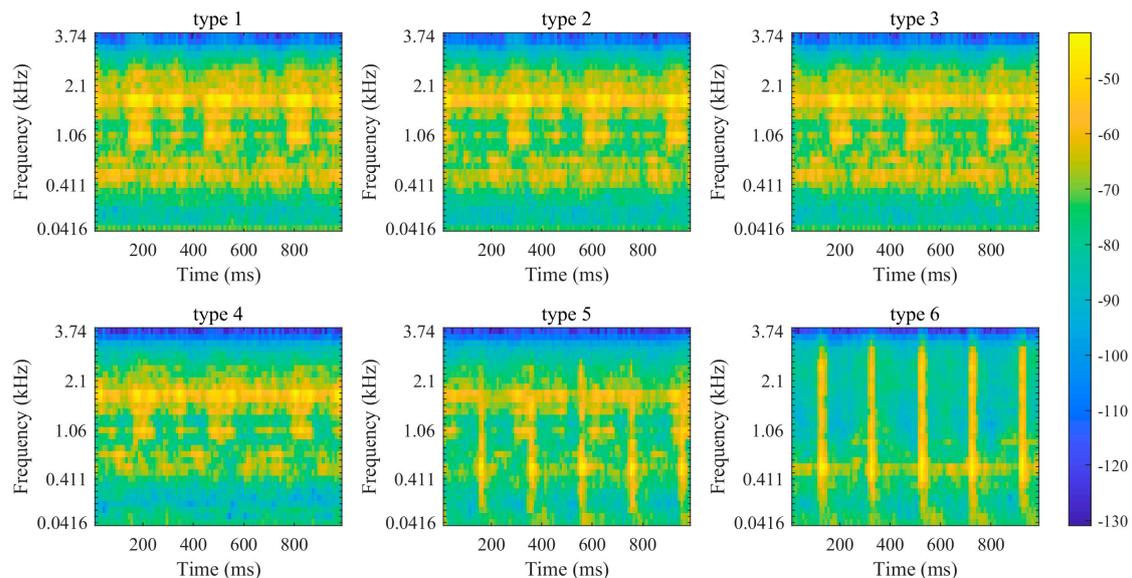


Figure 3. Mel–spectrum of six types of truncated signals.

3.2. Transfer Learning

Transfer learning is the extension of deep learning. The traditional deep learning approach is to train a neural network from scratch using a training set and test on a testing set. This process is the learning process, and the result of learning is to make it able to deal with new data of the same task. The disadvantage of this method is that when faced with different tasks, it is necessary to use new data to re-establish and train the model, which consumes computational costs. The emergence of transfer learning effectively solves this problem. It simulates how humans learn across tasks, applying the knowledge learned from one task to tasks in other related fields. The significant advantage of this deep learning method is that it can reduce the time and cost of learning. The premise of transfer learning is to have a trained neural network model. The problem studied in this paper is image classification. Many organizations and teams worldwide have developed models suitable for image feature extraction and classification based on large image data sets [16]. These models have different structures and have a large number of parameters. They have shown superior performance in image classification problems. After the original vibration signal is transformed into the Mel spectrum, the problem is transformed from time series classification to image recognition. Therefore, migrating it to the vibration signal Mel-spectrum recognition problem in this paper can reduce the steps of feature extraction and model establishment.

The pre-trained convolutional network model is the basis of our transfer learning. The structure of the convolutional neural network can be divided into three parts: convolutional layer, pooling layer and fully connected layer. The convolutional layer is responsible for the extraction of image features. Its main parameters include the number of convolution kernels, the size of the convolution kernel and the stride. The pooling layer is responsible for receiving the features processed by the convolution layer and reducing the dimensions

to remove redundant information. Its main parameters include the pool size, pooling mode and stride. Common pooling modes include maximum pooling, average pooling, global pooling and so on. The fully connected layer is equivalent to an ordinary neural network model, which receives the image features from the convolutional layers and the pooling layers for classification. The last layer of the fully connected layer is the output layer, which is responsible for the output of the corresponding category of samples. The main parameters of the fully connected layer include the number of neurons and activation function. The activation functions commonly used in the full connection layer include ReLU, sigmoid, tanh, softmax functions, etc., which realize the nonlinear transformation of input data to enhance the expression ability of the network. The size and category of images will vary for different image recognition problems. Adjusting the input size of the convolutional layer can adapt to different image sizes. The number of output neurons of the fully connected layer depends on the number of categories.

This paper uses the VGG16 model as the pre-trained neural network for transfer learning. The VGG16 model is a convolutional network proposed by Simonyan and Zisserman [17]. The VGG16 model performs well on the ImageNet dataset, which contains more than 1 million images of 100 categories. Therefore, the convolutional layer of the network has superior image feature extraction capabilities. VGG16 consists of 13 convolutional layers, three fully connected layers, and five pooling layers. The first convolution layer of VGG16 has 64 convolution kernels. The number of convolution kernels in the middle layers gradually increases, and the last layer has 512 convolution kernels. The size of the convolution kernels is 3×3 , and the stride is 1×1 . The kernel size of the pooling layers of VGG16 is 2×2 , the stride is 2×2 , and the pooling mode is maximum pooling. We froze the parameters of the convolutional layer to ensure that the weights of this part will not be updated. The classification problem on vibration signals requires adjusting the fully connected layer's parameters. The image size obtained in this paper is a three-channel image of 150×150 , and the number of output categories in six categories. In this paper, we set the fully connected layer of the original VGG16 model to 5 layers, and the output layer contains six neurons. The activation function of the first four layers of neurons is ReLU, and the last layer is softmax. The number of neurons in each layer is [1024, 1024, 512, 512, 6]. Figure 4 shows the fine-tuning VGG16 model.

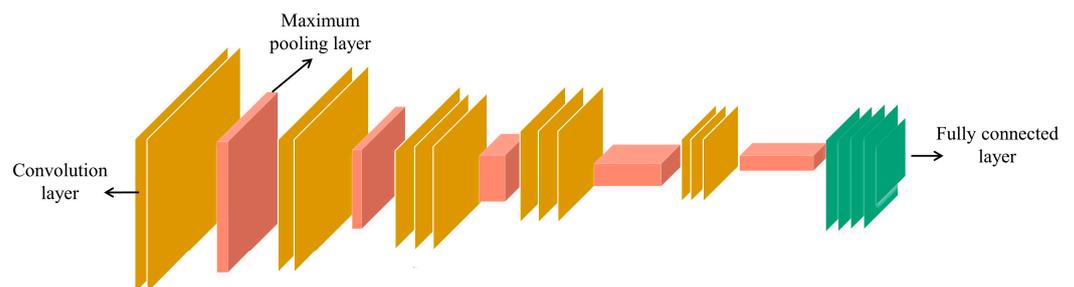


Figure 4. Fine-tuning VGG16 Model.

3.3. Recognition Process

Figure 5 shows the working state identification process of a rotating mechanical structure based on the Mel-spectrum and transfer learning method proposed in this paper. The whole framework consists of three parts. The first part is feature engineering, which aims to provide feature data for input in transfer learning. Firstly, the vibration signal data is collected from the vibration sensor, and then the collected vibration signal is truncated to make the sample length consistent and reduce prediction delay. In this paper, the truncated signal time is set at 1 s, and the overlap is set at 0.4 s. Then the processed signal is transformed into the Mel-spectrum. In the second part, we use the convolutional layer and pooling layers of the VGG16 model to extract the one-dimensional features of the Mel-spectrum. Then we modify the fully connected layer part of VGG16 and use the received features to output classification results. This paper locks the parameters of the convolutional

layer and pooling layer of the VGG16 model and trains the network model by updating the parameters of the fully connected layer. In the third part, we compare the predicted labels with the true labels and use the confusion matrix, accuracy, precision, recall, and f1-score to evaluate the quality of the model. Table 2 briefly describes the four indicators. TP (true positive) indicates the positive example of the correct prediction, TN (true negative) indicates the negative example of the correct prediction, and FP (false positive) and FN (false negative) are similar to TP and TN. The whole data set has one accuracy value, and the last three indicators can separately evaluate the classification performance of each type of signal. The four indicators comprehensively reflect the classification performance of the model. The confusion matrix is drawn based on the classification results of the test set. It can clearly reflect the classification of each type of sample.

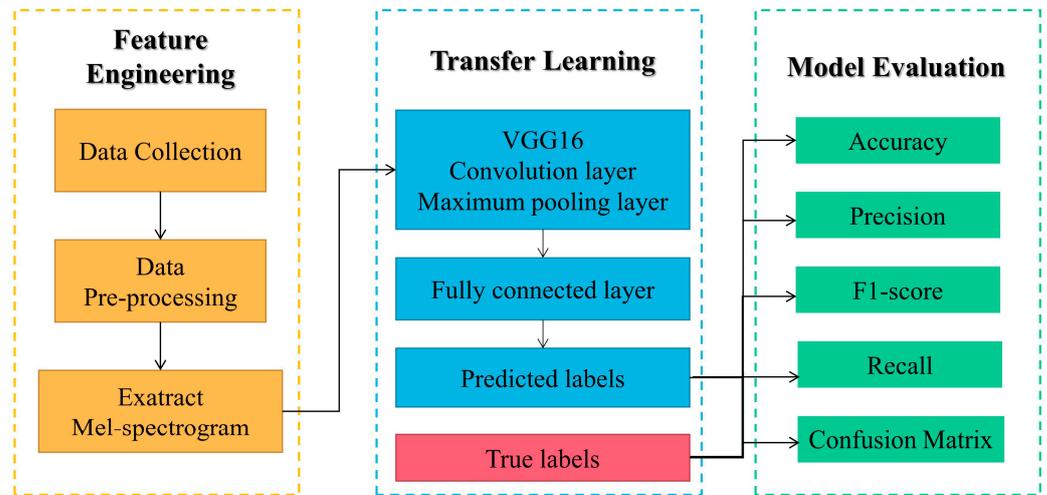


Figure 5. Recognition process.

Table 2. Evaluation indicators.

Indicators	Equation
accuracy	$(TP + TN)/(TP + TN + FP + FN)$
precision	$TP/(TP + FP)$
recall	$TP/(TP + FN)$
f1-score	$2/(1/precision + 1/recall)$

4. Experimental Results

In this paper, the vibration signals of rotating mechanical structures are classified based on the feature extraction and pattern recognition framework described above. This paper divides the data set into 70% training and 30% test sets. The convolutional and pooling layers of the model are used for feature extraction and remain unchanged. The training process only updates the weights of the fully connected layer, and the number of iterations is 50. Figure 6 shows the confusion matrix on the test set, and Table 3 lists the evaluation indicators obtained by the model on the test set. The overall accuracy rate is 96.37%. From the confusion matrix, 26 type 3 signals were classified into Type 2 signals. It indicates that the characteristics of the two procedures are similar. Overall, the framework adopted in this paper shows superior performance on the classification problem of vibration signals.

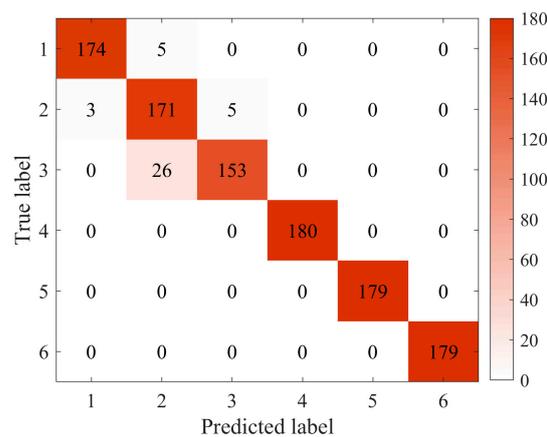


Figure 6. Confusion Matrix on the test set.

Table 3. VGG16 classification performance.

Type	State	Precision	Recall	F1-Score
1	Idling	98.31%	97.21%	97.75%
2	Idling + knocking platform rack	84.65%	95.53%	89.76%
3	Idling + knocking motor rack	96.84%	85.47%	90.80%
4	Rotation (120 rpm)	100.00%	100.00%	100.00%
5	Rotation (120 rpm) + Impact	100.00%	100.00%	100.00%
6	Rotation (60 rpm) + Impact	100.00%	100.00%	100.00%

5. Comparative Study

There are many convolutional neural networks pre-trained on large datasets. In addition to VGG16, network models, such as VGG19, MobileNet, MobileNetV2, Xception, and Inception V3, also show superior performance on the ImageNet dataset [18]. The difference among them is the network structure and model parameters. VGG19 has three more convolutional layers than VGG16. VGG16 and VGG19 increase the depth of the network by using more small convolution kernels and fewer pooling layers. This kind of network mainly proves that network depth can affect network performance to some extent. MobileNet and MobileNetV2 are lightweight network models. In this kind of network, deep separable convolution technology is used instead of standard convolution, which can greatly reduce the number of parameters in the model. MobileNetV2 introduces the residual bottleneck module on the basis of MobileNet, which can effectively reduce the complexity of the model. Xception and InceptionV3 networks introduce the inception module. By means of multi-scale feature extraction, dimension reduction, maximum pooling and Concatenate operation, the module can improve the expressiveness and precision of the model without increasing the amount of computation. Xception uses a deeply separable convolution and an extreme Inception module, while InceptionV3 uses branch-based inception modules and auxiliary classifiers. All these designs are helpful in improving the expressiveness and generalization ability of the model so as to improve its accuracy of the model. All of these models have been proven to have superior classification performance on large image classification data sets. Therefore, the idea of transfer learning is beneficial to simplify the model construction process and improve classification accuracy.

This paper uses the Mel-spectrum dataset of vibration signals to compare the application performance of these six models. The models all use a custom five-layer fully connected layer, and the number of iterations is 50 steps. Figure 7 shows the confusion matrix of the six models on the test set. Table 4 shows the f1-score and accuracy of all models. The classification accuracy of the six models for the type 2 and 3 signals is lower, while other categories are close to or reach 100%. Among them, the overall accuracy rates of VGG16 and MobileNet reached 96.37% and 96.28%, respectively.

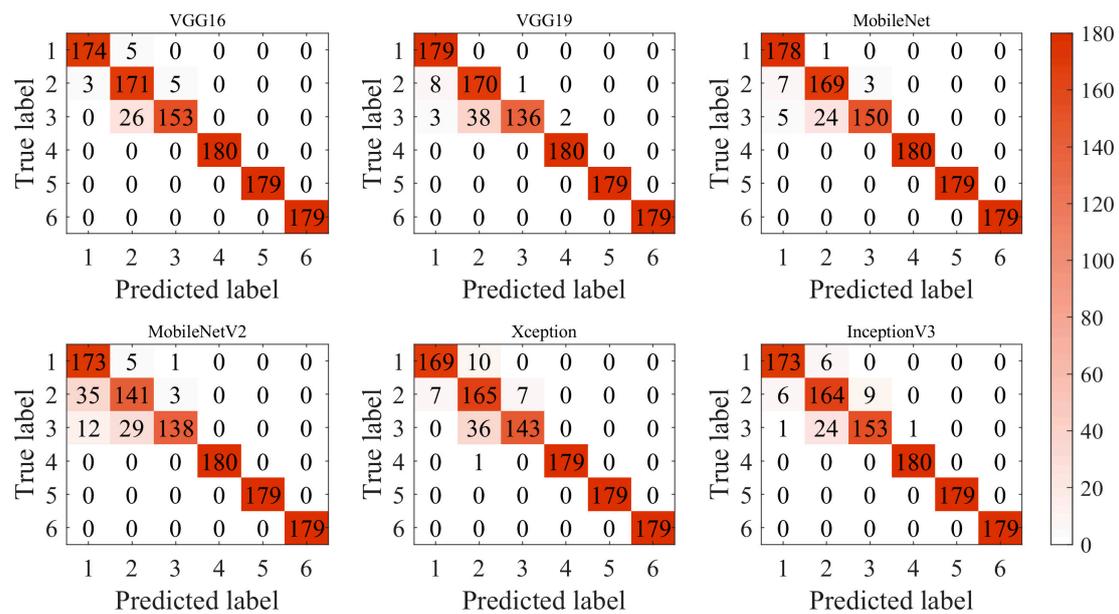


Figure 7. Confusion matrix of six models on the test set.

Table 4. F1-score and accuracy on six pre-trained models.

Type	VGG16	VGG19	MobileNet	MobileNetV2	Xception	InceptionV3
1	97.75%	97.02%	96.48%	86.72%	95.21%	96.38%
2	89.76%	87.86%	90.62%	79.66%	84.40%	87.94%
3	90.80%	86.08%	90.36%	85.98%	86.93%	89.74%
4	100.00%	99.45%	100.00%	100.00%	99.72%	99.72%
5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
6	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Accuracy	96.37%	95.16%	96.28%	92.09%	94.33%	95.63%

6. Conclusions

With the improvement of industrial intelligence, the generation of massive data and the development of artificial intelligence technology will play an essential role in equipment monitoring and fault diagnosis. This paper adopts a framework based on the Mel-spectrum and transfer learning. Firstly, the original signal is truncated by windowing and converted into a Mel-spectrum. Combined with the idea of transfer learning, we adopt the VGG16 model for feature extraction and pattern recognition of Mel-spectrum. At the same time, compared with the other five types of pre-trained convolutional networks, the VGG16 model achieved the highest recognition accuracy rate of 96.37% on the test set. The recognition accuracy of other models is also above 90%. This framework has two advantages. On the one hand, The Mel-spectrum preserves the time-domain and frequency-domain features of the original signal and highlights the low-frequency part.

On the other hand, the transfer learning method eliminates the model-building step. The work in this paper reflects the effectiveness of the Mel-spectrum representation of vibration signal features and the expansion ability of transfer learning. It indicates that Mel-spectrum and transfer learning are promising technology in rotating mechanical structures. In the future, with the accumulation of data in the industrial field, the migration and integration of data and models in different fields will become an important driving force for promoting industrial intelligence.

Author Contributions: Methodology, F.J. and F.L.; Data Curation, Z.C.; Investigation, F.L.; Writing—Original Draft Preparation, F.L.; Writing—review and editing, Z.L., J.T., W.Z., Y.T., H.L. and S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the National Key R&D Program of China (2020YFB2008803), National Natural Science Foundation of China (12004413, 52105549, U2013603), Beijing Nova Program (Z201100006820012 and 20220484172), China Postdoctoral Science Foundation (2021M690828), Heilongjiang Postdoctoral Grant (LBH-Z20145), and Self Planned Task of State Key Laboratory of Robotics and System (HIT) (SKLRS202113B).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Sreejith, B.; Verma, A.K.; Srividya, A. Fault diagnosis of rolling element bearing using time-domain features and neural networks. In Proceedings of the 2008 IEEE Region 10 and the Third International Conference on Industrial and Information Systems, Kharagpur, India, 8–10 December 2008; pp. 1–6.
2. Doguer, T.; Strackeljan, J. Vibration Analysis using Time Domain Methods for the Detection of small Roller Bearing Defects. In Proceedings of the SIRM 2009-8th International Conference on Vibrations in Rotating Machines, Vienna, Austria, 23–25 February 2009; p. 16.
3. Mishra, C.; Samantaray, A.K.; Chakraborty, G. Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising. *Measurement* **2017**, *103*, 77–86. [[CrossRef](#)]
4. Ibarra-Zarate, D.; Tamayo-Pazos, O.; Vallejo-Guevara, A. Bearing fault diagnosis in rotating machinery based on cepstrum pre-whitening of vibration and acoustic emission. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 4155–4168. [[CrossRef](#)]
5. Gao, H.; Liang, L.; Chen, X.; Xu, G. Feature extraction and recognition for rolling element bearing fault utilizing short-time Fourier transform and non-negative matrix factorization. *Chin. J. Mech. Eng.* **2014**, *28*, 96–105. [[CrossRef](#)]
6. Ben Ali, J.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27. [[CrossRef](#)]
7. Zhang, S.; Zhou, J.; Wang, E.; Zhang, H.; Gu, M.; Pirttikangas, S. State of the art on vibration signal processing towards data-driven gear fault diagnosis. *IET Collab. Intell. Manuf.* **2022**, *4*, 249–266. [[CrossRef](#)]
8. Anwarsha, A.; Narendiranath Babu, T. Recent advancements of signal processing and artificial intelligence in the fault detection of rolling element bearings: A review. *J. Vibroeng.* **2022**, *24*, 1027–1055. [[CrossRef](#)]
9. Dong, S.; Xu, X.; Chen, R. Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing. *J. Braz. Soc. Mech. Sci. Eng.* **2015**, *38*, 2255–2263. [[CrossRef](#)]
10. Pandya, D.H.; Upadhyay, S.H.; Harsha, S.P. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. *Expert Syst. Appl.* **2013**, *40*, 4137–4145. [[CrossRef](#)]
11. Li, Y.; Xu, M.; Wei, Y.; Huang, W. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* **2016**, *77*, 80–94. [[CrossRef](#)]
12. Wang, Y.; Huang, S.; Dai, J.; Tang, J. A Novel Bearing Fault Diagnosis Methodology Based on SVD and One-Dimensional Convolutional Neural Network. *Shock. Vib.* **2020**, *2020*, 1850286. [[CrossRef](#)]
13. Khorram, A.; Khalooei, M.; Rezghi, M. End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis. *Appl. Intell.* **2021**, *51*, 736–751. [[CrossRef](#)]
14. Wang, L.-H.; Zhao, X.-P.; Wu, J.-X.; Xie, Y.-Y.; Zhang, Y.-H. Motor Fault Diagnosis Based on Short-time Fourier Transform and Convolutional Neural Network. *Chin. J. Mech. Eng.* **2017**, *30*, 1357–1368. [[CrossRef](#)]
15. Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A systematic review of deep transfer learning for machinery fault diagnosis. *Neurocomputing* **2020**, *407*, 121–135. [[CrossRef](#)]
16. Park, J.; Jung, Y. A review and comparison of convolution neural network models under a unified framework. *Commun. Stat. Appl. Methods* **2022**, *29*, 161–176. [[CrossRef](#)]
17. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
18. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [[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.