

Research Article

Rolling Bearing Fault Diagnosis Method Based on Multisynchrosqueezing S Transform and Faster Dictionary Learning

Guodong Sun ¹, Ye Hu ¹, Bo Wu ², and Hongyu Zhou ¹

¹School of Mechanical Engineering, Hubei University of Technology, Wuhan, Hubei, China

²Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China

Correspondence should be addressed to Guodong Sun; sgdeagle@163.com and Hongyu Zhou; engzhy@126.com

Received 9 April 2021; Revised 28 April 2021; Accepted 4 May 2021; Published 19 May 2021

Academic Editor: Haiyang Pan

Copyright © 2021 Guodong Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Addressing the problem that it is difficult to extract the features of vibration signal and diagnose the fault of rolling bearing, we propose a novel diagnosis method combining multisynchrosqueezing S transform and faster dictionary learning (MSSST-FDL). Firstly, MSSST is adopted to transform vibration signals into high-resolution time-frequency images. Then, the local binary pattern (LBP) operator is introduced to extract the low-dimensional texture features of time-frequency images, which improves the speed of fault recognition. Finally, nonnegative matrix factorization (NMF) with only one hyperparameter and nonnegative linear equation are used to solve the dictionary learning and feature coding, respectively. The feature coding is input into the classifier for training and recognition. Experiments show that our method performs well on the rolling bearing dataset of Case Western Reserve University (CWRU) and the Society for Machinery Failure Prevention Technology (MFPT). Further, the proposed method is applied to the loudspeaker pure-tone detection dataset, and the loudspeaker anomaly diagnosis is achieved. The diagnosis results verify that our method can meet the needs of practical engineering.

1. Introduction

Rotating machinery plays an increasing role in electric manufacturing, transportation, power, and other industries. As the core component of rotating machinery, rolling bearings directly affect the operation of the entire equipment. However, the damage of rolling bearings is inevitable, which may cause serious economic losses and even safety accidents [1–3]. Thus, it is of great significance to detect bearing faults in time and take appropriate maintenance measures according to the diagnosis results [4, 5].

Vibration signal, which contains rich information and can reflect the running state of rotating machinery well, has become the most commonly used signal source in fault diagnosis of rotating machinery. The vibration signals of rolling bearings have obvious nonlinear and nonstationary characteristics so that it is difficult to recognize their faults directly. By using the appropriate time-frequency analysis

method to process the vibration signal, we can obtain the variation law of its spectrum with time. The idea of time-frequency analysis is originated from the Gabor transform [6]. Thereafter, short-time Fourier transform (STFT) [7], continuous wavelet transform (CWT) [8], and S transform (ST) [9] appear successively. Although these methods are easy to implement, the limitations of Heisenberg's uncertainty principle [10, 11] prevent them from improving both time and frequency resolutions. To obtain time-frequency images with better energy concentration of vibration signals, Daubechies et al. [12] proposed synchrosqueezed wavelet transform (SSWT). In essence, it is a time-frequency analysis method of energy rearrangement. Namely, based on CWT, the spectrum energy is redistributed and concentrated on the instantaneous frequency [13]. Based on this idea, Huang et al. [14] proposed the synchrosqueezing S transform (SSST) and Yu et al. [15] proposed the multi-synchrosqueezing transform (MSST). SSST is an energy

rearrangement algorithm based on ST, and ST has better performance than STFT. MSST is a multiple energy rearrangement algorithm based on STFT. In other words, multiple iterations of synchrosqueezing transform are performed. The more iterations, the better the time-frequency energy concentration. Combining the advantages of both SSST and MSST, multisynchrosqueezing S transform (MSSST) [16] is proposed as a new time-frequency analysis method for rolling bearing vibration signals.

Although the time-frequency image of the vibration signal of the rolling bearing is more intuitive than the raw signal, the dimension of the time-frequency image is too large to be directly inputted into the classifier for training and recognition. Therefore, it is necessary to extract the low-dimensional and valuable features of the time-frequency image. Over recent years, many scholars have studied this problem. In reference [17–19], a convolutional neural network (CNN) was used to extract the features of CWT, STFT, and HHT time-frequency images, respectively. Li et al. [20] proposed a convolution sparse self-learning (CSSL) to extract the defective bearing morphological feature. Wang et al. [21] designed a one-dimensional vision ConvNet (VCN) to extract multiscale sensitive features of bearings in complex operating environments. These deep learning-based methods are relatively new, but they have three major problems as follows: (1) they require a large number of training samples which are expensive and difficult to obtain; (2) it is challenging and time-consuming to train an excellent CNN from scratch with only a small number of samples; (3) plenty of hyperparameters have to be predetermined for CNN, such as activation functions, epoch number, learning rate, momentum, kernel sizes, and numbers of layers.

Therefore, the traditional machine learning method based on feature engineering is still worth further study. In reference [22], two-dimensional nonnegative matrix factorization (2DNMF) is developed to extract more informative features from the ST time-frequency images for accurate fault recognition. In reference [23], the features of ST time-frequency images were first extracted by nonnegative matrix factorization (NMF) [24, 25], and then nondominated sorting genetic algorithms (NSGA-II) were proposed to make secondary the selection of features. Yu et al. [26] proposed a rolling bearing fault diagnosis method based on Hilbert–Huang transform (HHT) and supervised sparse coding (SSC). This method adopted SSC to obtain a sparse representation of the marginal spectrum generated by HHT and used the support vector machine (SVM) to achieve fault recognition. Although these methods can obtain high recognition precision on their own datasets, they may not be suitable for simultaneous diagnosis of fault location and damage degree.

In this regard, Sun et al. [27] proposed a method combining MSST and sparse feature coding based on dictionary learning (SFC-DL). Li et al. [28] designed a symplectic weighted sparse SMM (SWSSMM) model with the sparsity constraint and low-rank constraint, and Li et al. [29] developed the discriminative manifold random vector functional link neural network (DMRVFLNN) model. The above fault diagnosis methods cannot overcome the problem of how to select the optimal parameters for their models.

More importantly, the method proposed by Sun et al. would consume many computer resources and running time because all the elements of time-frequency images are taken for nonnegative matrix factorization with sparseness constraints (NMFSCs) [30] and sparse coding.

To overcome the abovementioned problems, we propose to first extract the texture features of time-frequency images by the local binary pattern (LBP) [31, 32] operator and then use these texture features for dictionary learning. The LBP algorithm with linear order time complexity and space complexity is simple to calculate. Meanwhile, the texture features of time-frequency images have rich information and low dimension, which can greatly improve the performance of dictionary learning and feature coding. In addition, we use NMF instead of NMFSC for dictionary learning and reduce the hyperparameter to only one, which makes the optimal fault diagnosis model easier to be obtained. We name these optimizations faster dictionary learning (FDL).

In summary, the main contributions of this article are as follows:

- (1) A new method for rolling bearing fault diagnosis is proposed by combining MSSST and FDL, which is named MSSST-FDL
- (2) MSSST is adopted to obtain high-resolution time-frequency images of vibration signals, which can promote the accuracy of fault diagnosis
- (3) To improve the speed of feature extraction from time-frequency images, we design the FDL algorithm by introducing LBP and NMF
- (4) Experiment results on two rolling bearing datasets and one loudspeaker dataset show that the proposed method performs well and has the potential to be applied to different types of equipment

The remaining of this article is mainly described as follows. Section 2 introduces the theory of MSSST and FDL. Section 3 presents the experimental comparisons. An extended application of the proposed method is shown in Section 4. Finally, the conclusions are presented in Section 5.

2. The Proposed Method

In this paper, a new method based on time-frequency analysis and improved dictionary learning is proposed for fault diagnosis of rolling bearing. The main procedures are described as follows:

Step 1. Raw vibration signals of rolling bearing are collected under different working conditions, and their states are noted.

Step 2. The raw signals are segmented and ensured that each sample signal contains one complete period at least.

Step 3. MSSST is performed for sample signals to obtain the time-frequency images with high resolution.

Step 4. FDL is used to process the time-frequency images, and we can get the effective feature coding of each sample quickly.

Step 5. The feature coding set is divided into the training, validation, and testing sets. Then, the diagnosis model is obtained by cross-validation on the training set and the validation set.

Step 6. Testing set is input into the diagnosis model for fault diagnosis.

Figure 1 shows the overall framework of the proposed method, and the following subsections provide details of MSSST and FDL.

2.1. Multisynchrosqueezing S Transform. In practice, raw vibration signals are always nonlinear and nonstationary. It is necessary to process these complex signals. Fortunately, time-frequency analysis is an effective approach to reveal the frequency components and time-variation features of vibration signals. In various time-frequency analysis methods, MSSST combines the advantages of SSST and MSST to generate better energy concentration and suppress the cross-terms over the time-frequency plane.

Let the vibration signal be $s(t)$, then the expression of its ST is as follows:

$$ST(t, \omega) = \frac{|\omega|}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} s(u) e^{-(u-t)^2 \omega^2 / 2} e^{-i2\pi \omega u} du, \quad (1)$$

where ω is the frequency, t represents the time axis displacement parameter, i is an imaginary unit, and $e^{-(u-t)^2 \omega^2 / 2}$ is the window function.

It can be seen from equation (1) that the window function of ST is flexible. Its window width can change according to the change of the frequency ω . The window width is wider in the low-frequency part and narrower in the high-frequency part. This not only improves the shortcomings of STFT but also inherits the multiresolution characteristics of CWT. Therefore, combining the advantages of ST and the idea of iterative compression of MSST, the MSSST can be expressed as

$$\begin{aligned} \text{MSSST}^{[1]}(t, \omega) &= \int_{-\infty}^{+\infty} ST(t, \eta) \delta(\omega - \hat{\omega}_s(t, \eta)) d\eta, \\ \text{MSSST}^{[2]}(t, \omega) &= \int_{-\infty}^{+\infty} \text{MSSST}^{[1]}(t, \eta) \delta(\omega - \hat{\omega}_s(t, \eta)) d\eta, \\ &\vdots \\ \text{MSSST}^{[N]}(t, \omega) &= \int_{-\infty}^{+\infty} \text{MSSST}^{[N-1]}(t, \eta) \delta(\omega - \hat{\omega}_s(t, \eta)) d\eta, \end{aligned} \quad (2)$$

where $ST(t, \eta)$ is the ST of signal $s(t)$, $\text{MSSST}^{[N]}(t, \omega)$ is the MSSST after N iterations, $\hat{\omega}_s(t, \eta)$ is the instantaneous frequency (IF) estimate based on ST, and its expression is defined as follows [14]:

$$\hat{\omega}_s(t, \omega) = \omega + [i2\pi ST(t, \omega)]^{-1} \frac{\partial ST(t, \omega)}{\partial t}. \quad (3)$$

We substitute $\text{MSSST}^{[1]}(t, \eta)$ into $\text{MSSST}^{[2]}(t, \eta)$, and then the $\text{MSSST}^{[2]}(t, \eta)$ can be expressed as follows:

$$\begin{aligned} \text{MSSST}^{[2]}(t, \omega) &= \int_{-\infty}^{+\infty} \text{MSSST}^{[1]}(t, \eta) \delta(\omega - \hat{\omega}_s(t, \eta)) d\eta \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} ST(t, \xi) \delta(\omega - \hat{\omega}_s(t, \xi)) d\xi \right] \delta(\omega - \hat{\omega}_s(t, \eta)) d\eta \\ &= \int_{-\infty}^{+\infty} ST(t, \xi) \int_{-\infty}^{+\infty} \delta(\omega - \hat{\omega}_s(t, \xi)) \delta(\omega - \hat{\omega}_s(t, \eta)) d\eta d\xi \\ &= \int_{-\infty}^{+\infty} ST(t, \xi) \delta(\omega - \hat{\omega}_s(t, \hat{\omega}_s(t, \xi))) d\xi \\ &= \int_{-\infty}^{+\infty} ST(t, \eta) \delta(\omega - \hat{\omega}_s(t, \hat{\omega}_s(t, \eta))) d\eta. \end{aligned} \quad (4)$$

The rest can be done in the same manner, and then the $\text{MSSST}^{[N]}(t, \omega)$ can be expressed as follows:

$$\text{MSSST}^{[N]}(t, \omega) = \int_{-\infty}^{+\infty} ST(t, \eta) \delta(\omega - \hat{\omega}_s^{[N]}(t, \eta)) d\eta \quad (5)$$

where $\hat{\omega}_s^{[N]}(t, \omega) = \hat{\omega}_s(t, \hat{\omega}_s^{[N-1]}(t, \omega))$.

The MSSST uses the S transform of the SSST to obtain time-frequency coefficients with better energy concentration. At the same time, combined with the idea of multiple iterations in the MSST, the time-frequency results can be further sharpened. After one iteration, MSSST will construct a new IF estimate to reassign the blurry ST energy. Therefore, after several iterations, the IF estimation in the MSSST will get closer and closer to the real IF of the vibration signal.

Namely, the energy of the time-frequency distribution can be gradually concentrated.

2.2. Faster Dictionary Learning. The dimension of the time-frequency image generated by MSSST is 1600×800 , which is too high. If all the elements of the time-frequency image are directly input into the classifier for training and recognition, serious overfitting will occur. Therefore, we propose the FDL algorithm to extract the effective features of time-frequency images. The procedures of FDL are shown in Figure 2. First of all, the texture feature vectors of MSSST time-frequency images are extracted by the LBP operator, and each feature vector is taken as one column of V . After that, one-tenth of the samples are uniformly and randomly selected to

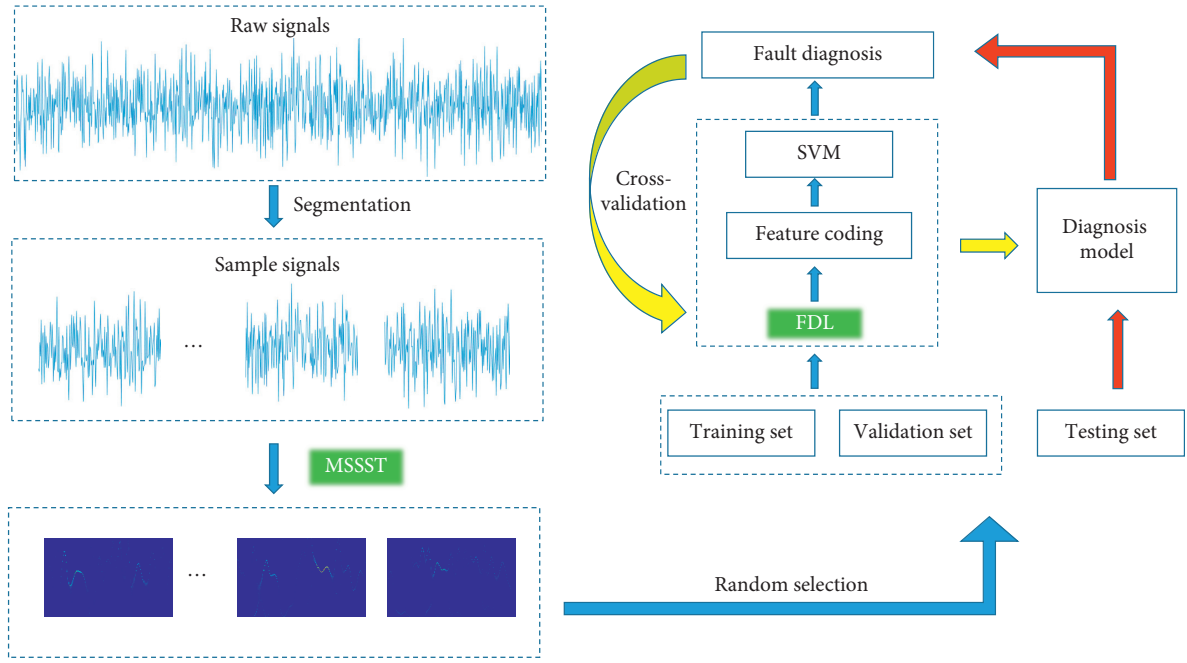


FIGURE 1: The overall framework of the presented method.

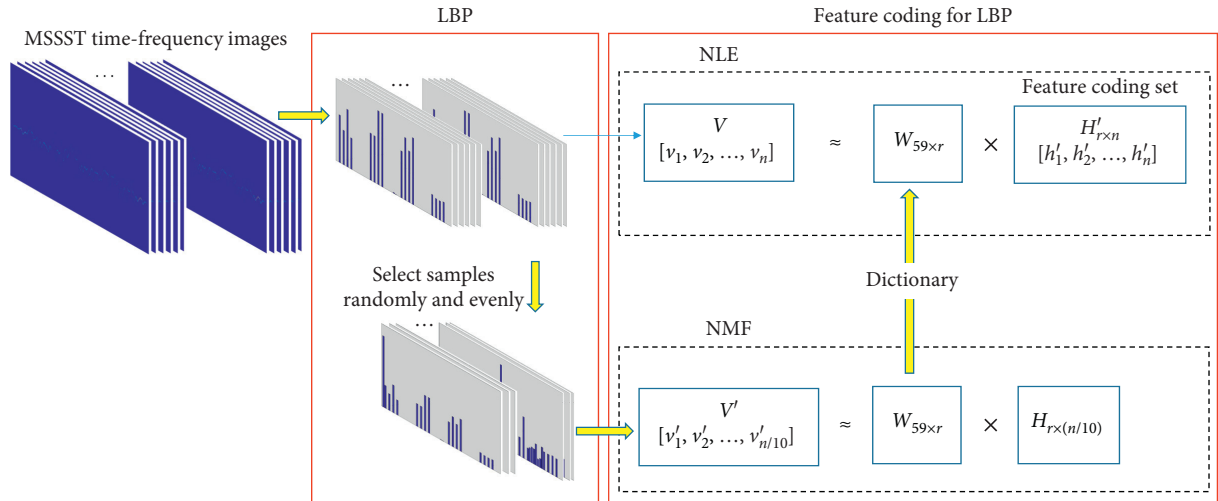


FIGURE 2: The procedures of FDL.

compose matrix V' , which is decomposed by NMF to generate the dictionary W . At last, in combination with W , the feature coding set H' of all samples can be solved by NLE. More details are described in the following subsections.

2.2.1. Local Binary Pattern Operator. As shown in Figure 3, the original LBP operator is defined as a 3×3 square window, and the center point of the window is taken as the threshold value to compare it with 8 adjacent pixels. If the surrounding

pixel value is greater than the value of the center point, this pixel is marked as 1; otherwise, it is 0. In the end, we can obtain 256 types of binary patterns.

To further reduce the number of binary patterns and improve the statistics, a uniform pattern [32] is designed, which recorded the jump times of binary numbers 0 and 1 of the LBP operator. If the number of jumps is less than or equal to 2, it is called uniform pattern, and all except the uniform pattern are classified into one class. As a result, the number of patterns has been reduced from 256 to 59. Namely, the dimension of the feature vector of the time-

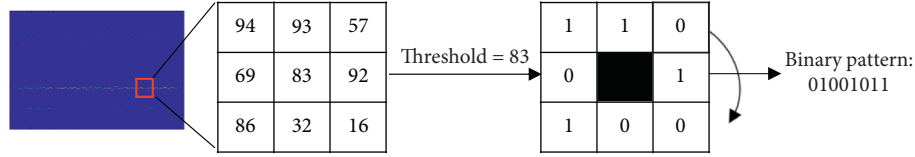


FIGURE 3: The original LBP operator.

frequency image is 59. The low-dimensional texture features of time-frequency images extracted by LBP can speed up dictionary learning and feature coding.

2.2.2. Feature Coding for LBP. By extracting the LBP features of the time-frequency images, the feature size can be reduced to 59. However, there are still some redundant features that affect the recognition accuracy. The optimal dimension can be further determined by feature coding so as to achieve the best recognition.

Before feature coding for LBP, it is necessary to find its basis dictionary. The method is NMF which is defined as follows:

$$V_{n \times m} \approx W_{n \times r} H_{r \times m}, \quad (6)$$

where $V_{n \times m}$, $W_{n \times r}$, and $H_{r \times m}$ are nonnegative matrices, V is the small-batch set of the LBP features, W is the basis dictionary, and H is the feature coding set. Each column of V is the LBP feature of one time-frequency image, and each column of H is corresponding to the feature coding of each LBP feature. Then, r can control the dimension of the feature coding.

NMF has only one hyperparameter, which can be used for dictionary learning to obtain the optimal model more efficiently. All the remaining samples can be represented by different linear combinations of column vectors of the basis dictionary. The linear combinations are called feature coding. We can solve the feature coding by nonnegative linear equation (NLE), which is expressed as follows:

$$\begin{aligned} \min \|v_i - Wh_i\|^2 \\ \text{s.t. } h_{ij}' \geq 0, \end{aligned} \quad (7)$$

where v_i represents the i th sample, h_i' represents the corresponding feature coding of v_i , and h_{ij}' represents the j th element in the feature coding h_i' .

By equation (7), we can obtain the vector set $[h_1', h_2', \dots, h_n']$ in Figure 2. The set $[h_1', h_2', \dots, h_n']$ is the feature coding set of all LBP features, which is denoted as H' and input into the classifier for training and fault recognition.

3. Experimental Study

To verify the performance of the proposed method, the Case Western Reserve University (CWRU) [33] and the Machinery Failure Prevention Technology (MFPT) [34] rolling bearing vibration signal datasets are selected for experiments. All experiments are carried out with Windows 10 (64 bit), CPU Intel Xeon E5-2640@2.40 GHz, memory 64 GB, and MATLAB 2017b.

3.1. Description of Dataset. The CWRU test platform is mainly composed of a motor, torque sensor, dynamometer, and electronic control equipment. The designation of the tested bearing is 6205-2RS JEM SKF, which is located at the drive end. The EDM technology is adopted to set the faults at three different positions (inner race, outer race, and ball) of the bearing. Each fault location has three different degrees of damage (fault diameter of 0.007, 0.014, and 0.021 inches, respectively). Therefore, after adding the normal state, there are 10 kinds of health conditions of rolling bearings. Vibration signals under each health condition are collected at four different motor loads (0, 1, 2, and 3 hp) and speeds (1797, 1772, 1750, and 1730 r/min) with a sampling frequency of 12 kHz. To retain data features as much as possible and increase the number of samples, the 40 raw samples are obtained under 10 health conditions and 4 working conditions are further divided. As shown in Figure 4, each raw sample is continuously divided into 150 samples, and each sample has 800 sampling points. Finally, the total number of samples is 6000, including 600 samples for each health condition and 150 samples for each working condition, which are divided into the training set, validation set, and testing set according to the ratio of 6:2:2. To improve the robustness of the diagnosis method and meet the needs of practical engineering, the influence of working conditions on fault recognition is not considered. At the same time, to avoid contingency, 150 samples under each working condition are randomly divided into the training set (90), validation set (30), and testing set (30). The details of the CWRU dataset are shown in Table 1.

MFPT dataset includes inner race, outer race, and normal health conditions with a motor speed of 25 Hz. The sampling frequencies of the inner race and outer race data are both 97656 Hz. The inner race data are collected at 7 different motor loads (0, 50, 100, 150, 200, 250, and 300 lbs). The outer race data are collected at 7 different motor loads (25, 50, 100, 150, 200, 250, and 300 lbs). The sampling frequency of normal data is 97656 Hz, and the motor load is 270 lbs. MFPT data are segmented in the same way as CWRU. However, each segmented sample has 4000 sampling points. To reduce the redundant information of each sample and facilitate the subsequent operation, it is sampled down every 5 points to get the final sample length of 800 sampling points. In the end, a total of 2,100 samples are obtained, with 700 samples for each health condition. In the case of inner and outer race faults, 100 samples are corresponding to each motor load. The ratio of training set validation set and testing set is 6:2:2. They are selected at random in the same way as the CWRU. The details of the MFPT dataset are shown in Table 2.

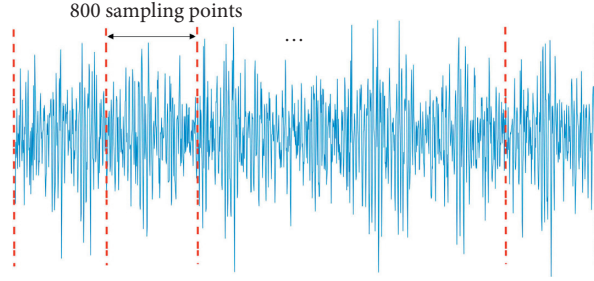


FIGURE 4: Segmentation of raw sample data of CWRU.

TABLE 1: Description of CWRU dataset.

Fault type	Fault diameter (inches)	Motor load (hp)	Motor speed (r/min)	Dataset			Fault label
				Training set	Validation set	Testing set	
Normal	—	0/1/2/3	1797/1772/1750/1730	360	120	120	1
Inner race	0.007	0/1/2/3	1797/1772/1750/1730	360	120	120	2
	0.014	0/1/2/3	1797/1772/1750/1730	360	120	120	3
	0.021	0/1/2/3	1797/1772/1750/1730	360	120	120	4
Ball	0.007	0/1/2/3	1797/1772/1750/1730	360	120	120	5
	0.014	0/1/2/3	1797/1772/1750/1730	360	120	120	6
	0.021	0/1/2/3	1797/1772/1750/1730	360	120	120	7
Outer race	0.007	0/1/2/3	1797/1772/1750/1730	360	120	120	8
	0.014	0/1/2/3	1797/1772/1750/1730	360	120	120	9
	0.021	0/1/2/3	1797/1772/1750/1730	360	120	120	10

TABLE 2: Description of MFPT dataset.

Fault type	Motor load (hp)	Motor speed (Hz)	Dataset			Fault label
			Training set	Validation set	Testing set	
Normal	270	25	420	140	140	1
Inner race	0/50/100/150/200/250/300	25	420	140	140	2
Outer race	25/50/100/150/200/250/300	25	420	140	140	3

3.2. *Time-Frequency Analysis of Vibration Signals.* To prove the superiority of MSSST, it is compared with several popular time-frequency analysis methods. Each type of faults in two datasets provides one sample randomly for our study. Table 3 shows the Rényi entropy (RE) [35, 36] of the samples. The time-domain waveform of the vibration signal of the CWRU inner race fault is shown in Figure 5(a). Figures 5(b)–5(g), respectively, show the time-frequency distribution of STFT, ST, Wigner–Ville transformation (WVT) [37], SSST, MSST, and MSSST.

It can be seen from Figure 5 that the energy concentration of time-frequency images by STFT and ST is poor, and WVT has a serious cross-term. Further comparing SSST, MSST, and MSSST, we find that MSST and MSSST have better energy concentration, which verifies that iterative compression can improve the time-frequency analysis algorithm, and we also find that the time-frequency image of MSSST has less redundant information than that of MSST. Meanwhile, it can be seen from Table 3 that the RE of MSSST is always the lowest, which further indicates that the energy of the time-frequency distribution by MSSST is more concentrated. Therefore, in this paper, MSSST is adopted as

the time-frequency analysis method of the rolling bearing vibration signal.

3.3. Ablation Study

3.3.1. *Hyperparameter r .* From Section 2.2.2, it can be seen that the hyperparameter r directly determines the dimension of feature coding of the sample. In order to obtain the algorithm model, it is necessary to determine the value of the hyperparameter r . We change the value of r and combine different time-frequency analysis methods with FDL to carry out the experiments. To avoid contingency and particularity, each method is performed for ten repeated runs under different r . We determine the value of r by the average recognition accuracy of the validation set.

Detailed results are available in Figure 6. Table 4 records the highest average recognition accuracy and its corresponding standard deviation and r . According to Figure 6, from the overall trend, the average accuracy of MSSST is higher than that of other time-frequency analysis methods. From Table 4, it can be observed that the standard deviations

TABLE 3: Rényi entropy of different frequency analysis methods.

Dataset	Fault type	Method					
		STFT	ST	WVT	SSST	MSST	MSSST
CWRU	Normal	17.9889	16.9369	16.3525	12.6381	11.5885	10.6908
	Inner race	17.7992	16.8018	16.9188	12.7815	12.7316	11.8420
	Ball	17.1826	16.6566	16.4919	12.1864	11.2877	10.6046
	Outer race	17.6826	17.1549	16.3112	12.8543	12.2866	11.4123
MFPT	Normal	19.5311	18.5683	17.8579	14.8126	14.5465	12.4836
	Inner race	19.0094	18.3162	18.2588	14.6791	13.7552	11.1386
	Outer race	19.9868	19.3678	19.0506	14.4431	13.6602	11.9499

of diagnostic accuracy are relatively small, which indicates that the obtained models are relatively stable.

3.3.2. Module Combination. We carry out a series of experiments to study the effect of time-frequency analysis and feature extraction methods. The hyperparameters r applied to the testing set are set according to Table 4. The experimental results are shown in Figure 7. We have the following three findings. (1) Under the same conditions, the recognition accuracy of MSSST is always higher than that of MSST. Investigating its reason, MSST is based on STFT and the width of its window function is fixed. However, it is based on ST that MSSST has a variable window function, which improves the self-adaptability of spectrum analysis and can extract more detailed time-frequency characteristics in vibration signals. (2) The features extracted using NMF + NLE have higher recognition accuracy than those extracted directly using the LBP operator. That is because NMF + NLE gets the optimal feature dimension, while LBP has a fixed value of 59. (3) When LBP is combined with NMF + NLE, the recognition accuracy reaches the highest, which indicates that when NMF + NLE is directly used to extract the features of the time-frequency image, its dimension is too large. Therefore, too much redundant information is also extracted, which affects the recognition effect.

Taken together, these results suggest that whether in the CWRU dataset or the MFPT dataset, MSSST + FDL is the best combination.

3.3.3. Time for Feature Extraction. To verify the superiority of the time efficiency of the feature extraction algorithm in this paper, four different methods are designed to extract the features of the MSSST time-frequency image, and the time taken by them is recorded, respectively. We define T_d as the time spent in dictionary learning and T_f as the time spent in feature extraction of one sample.

As shown in Table 5, in CWRU and MFPT datasets, the dictionary learning time of texture feature vectors of time-frequency images is 0.037 and 0.012 hours, respectively. And yet, the time of dictionary learning on time-frequency images is 16.469 hours and 3.897 hours separately, which is very time-consuming. More importantly, in the process of fault recognition of the signal, although the time of only using LBP is the shortest, it can be seen from Section 3.3.2 that the recognition accuracy of this method is low. In

addition, the feature coding of a time-frequency image takes 39,525 and 37,413.6 milliseconds, respectively.

By contrast, the proposed algorithm only consumes less than 100 milliseconds, saving more than 300 times of time, which can better meet the real-time requirements in practical engineering applications. At the same time, the feature coding algorithm in reference [27] is also very time-consuming, which further indicates that the proposed algorithm has high timeliness.

3.4. Comparisons with Other Methods. To prove the effectiveness of the proposed method, Table 6 shows the recognition accuracy of different fault diagnosis methods for bearing faults.

Reference [27] directly adopted NMFSC + NLE to obtain the sparse coding of MSST time-frequency images and trained SVM to diagnose bearing faults. The parameter sparsity was set to 0.7; the parameter rank was set to 25 and 100 in datasets CWRU and MFPT, respectively. In reference [38, 39], Hilbert–Huang transform and convolutional neural network (HHT + CNN) were combined to recognize the bearing state. The former input the CWRU time-frequency images of 32×32 pixels into CNN, while the latter input MFPT time-frequency images of 32×32 and 96×96 pixels. In reference [40], the wavelet packet energy features combined with multifractal features (WPE-MFs), which are of feature size 33, were used to train SVM.

It can be seen from the results that, compared with other methods, our method has only one hyperparameter and can achieve higher recognition accuracy with fewer features.

4. Extended Application

To demonstrate the practical engineering application value of the proposed method, we try to apply the proposed method to loudspeaker fault diagnosis.

4.1. Data Description and Analysis. The loudspeaker signal acquisition system consists of a microphone, acquisition card, sweep generator, and acquisition software. First, the sweep generator touches the signal collection points of the loudspeaker. Then, the signal acquisition software synchronously collects the data of the loudspeaker sound signal. Finally, the label of the collected sound signal is marked.

The sampling frequency of the loudspeaker sound signal is 8 kHz. To increase the number of samples and ensure that

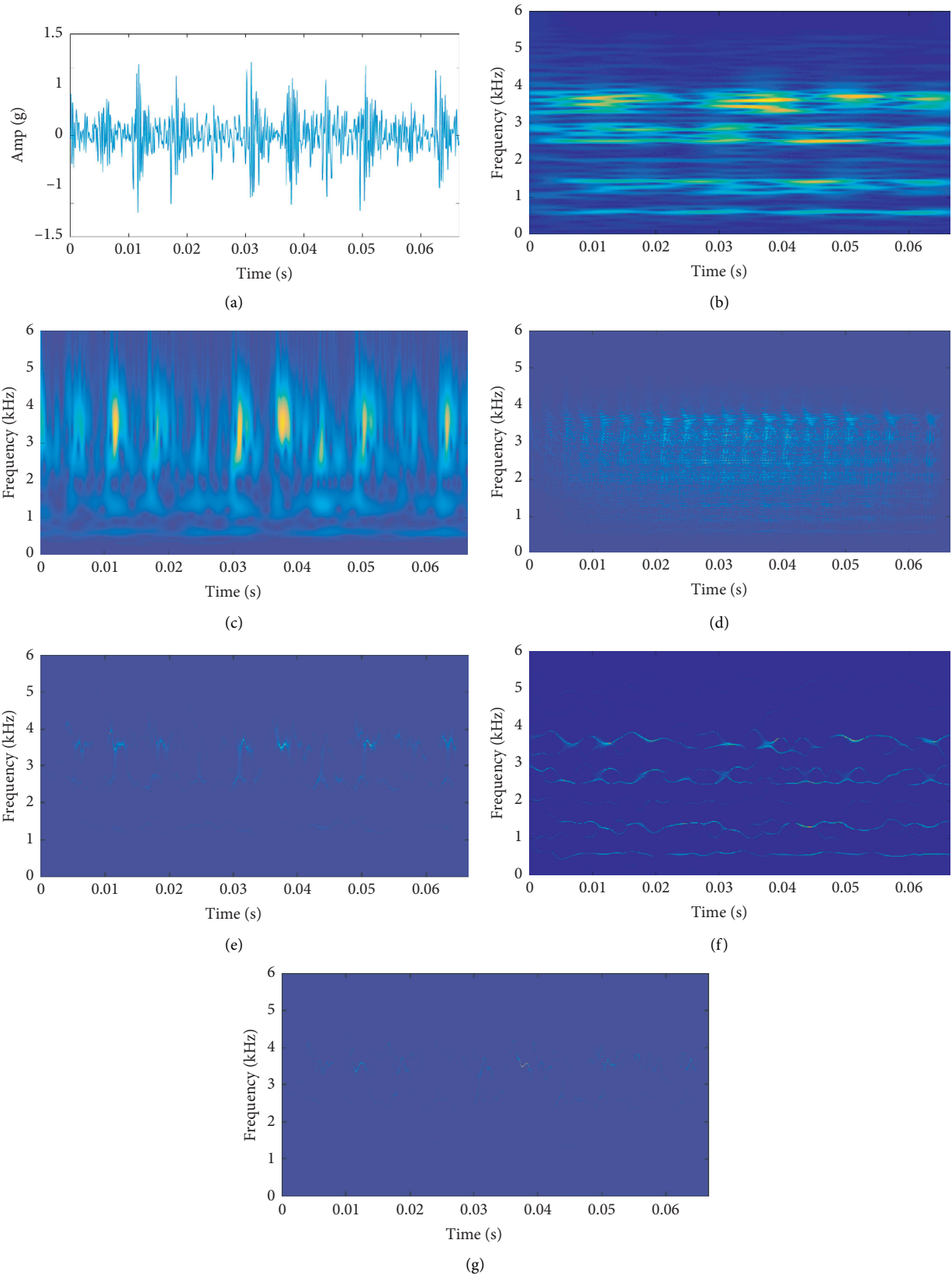


FIGURE 5: Time-frequency images of different frequency analysis methods: (a) raw signal; (b) STFT; (c) ST; (d) WVT; (e) SSST; (f) MSST; (g) MSSST.

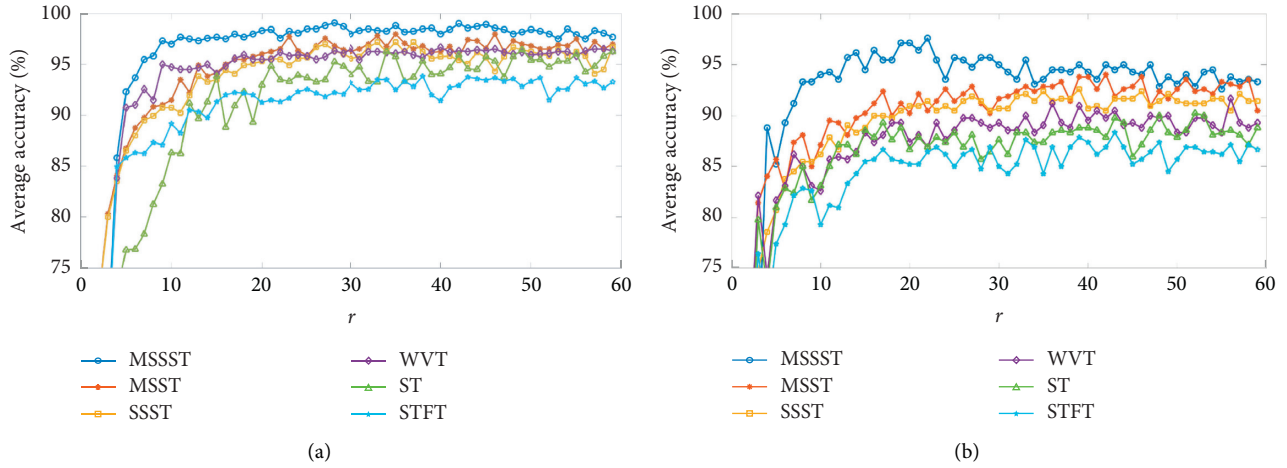


FIGURE 6: Average recognition accuracy of validation set under different r (a) in the CWRU dataset; (b) in the MFPT dataset.

TABLE 4: The highest average accuracy of validation set and its corresponding standard deviation and r .

Dataset	Data name	Methods					
		MSSST	MSST	SSST	WVT	ST	STFT
CWRU	Max average accuracy	99.08	98	97.16	96.66	96.41	93.83
	Standard deviation	0.32	0.42	0.46	0.43	0.51	0.42
	Optimal r	28	35	33	40	49	38
MFPT	Max average accuracy	97.61	94.04	92.61	91.66	90.24	88.33
	Standard deviation	0.36	0.44	0.42	0.48	0.53	0.50
	Optimal r	22	42	39	56	52	43

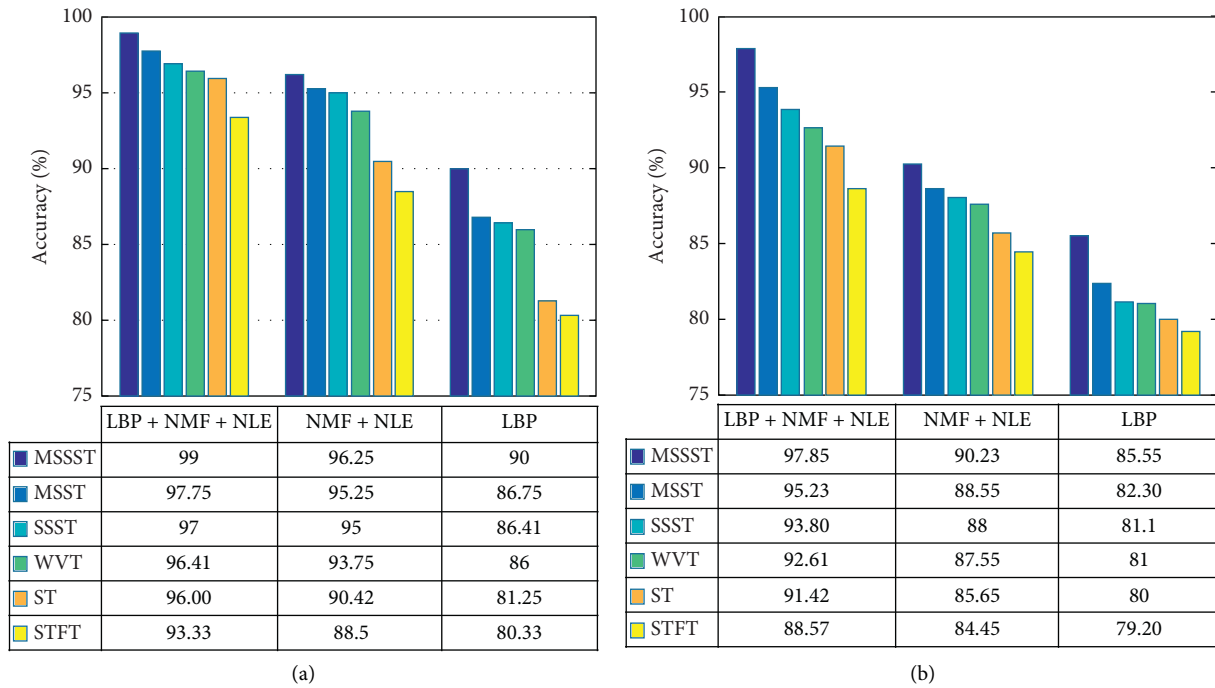


FIGURE 7: The recognition accuracy of testing set under different methods (a) in the CWRU dataset; (b) in the MFPT dataset.

TABLE 5: Time taken by different feature extraction methods.

Dataset	Feature extraction method	Feature size	Time	
			T_d (h)	T_f (ms)
CWRU	LBP	59	—	87.6
	NMF + NLE	28	16.469	39525.0
	LBP + NMF + NLE (FDL)	28	0.037	88.8
	NMFSC + NLE [27]	25	8.089	19413.0
MFPT	LBP	59	—	93.6
	NMF + NLE	22	3.897	37413.6
	LBP + NMF + NLE (FDL)	22	0.012	95.4
	NMFSC + NLE [27]	100	2.247	21573.6

TABLE 6: Comparison of the proposed method with other methods.

Dataset	Method	Sparsity	Feature size	Fault types	Accuracy (%)
CWRU	The proposed method	—	28	10	99
	SFC-DL + SVM [27]	0.7	25	10	98.03
	HHT + CNN [38]	—	32×32	10	95
	WPE-MF + SVM [40]	—	33	10	88.9
MFPT	The proposed method	—	22	3	97.85
	SFC-DL + SVM [27]	0.7	100	3	95.83
	HHT + CNN [39]	—	96×96	3	92.9
		—	32×32	3	75.9

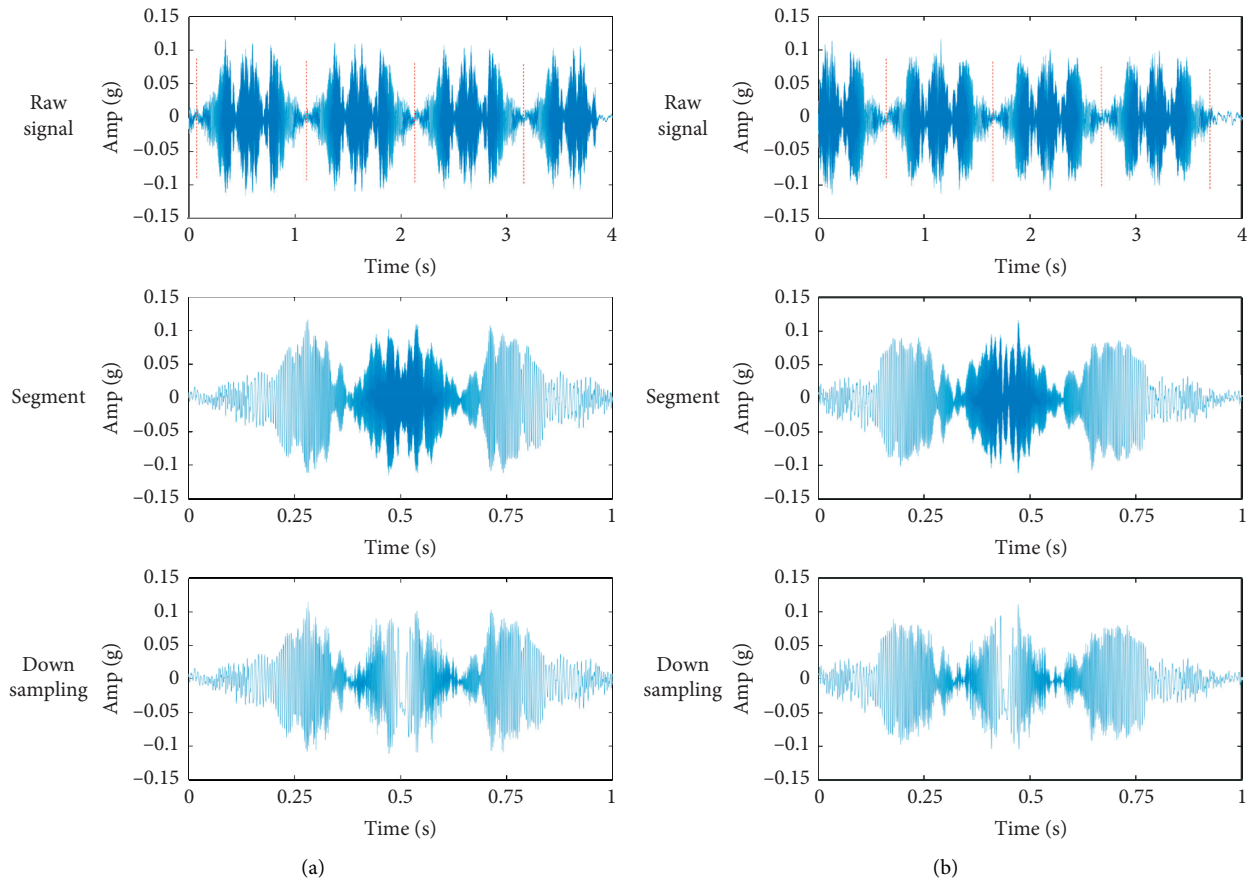


FIGURE 8: Time-domain waveform of loudspeaker signals: (a) normal; (b) abnormal.

TABLE 7: Description of loudspeaker pure-tone detection dataset.

Fault type	Training set	Validation set	Testing set	Fault label
Normal	600	200	200	1
Abnormal	600	200	200	2

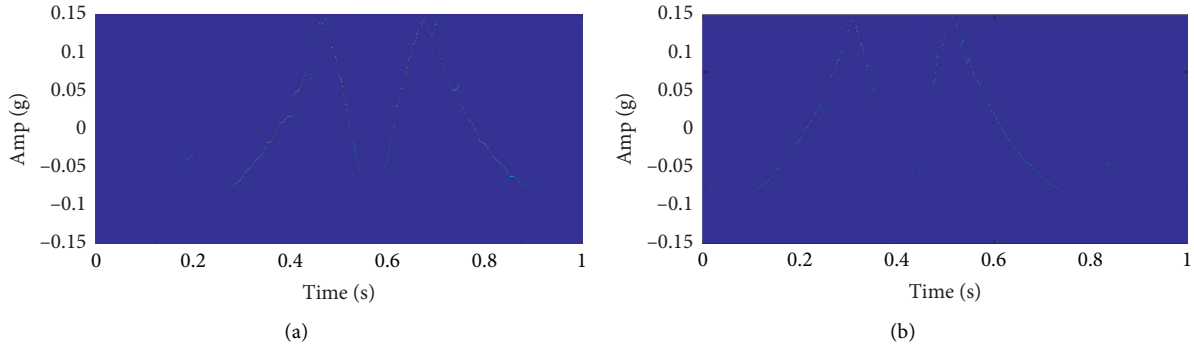
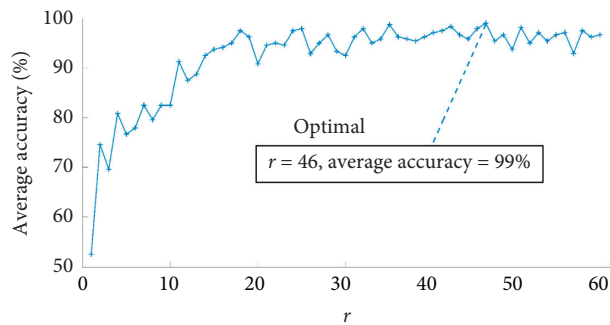


FIGURE 9: Time-frequency images of loudspeaker signals: (a) normal; (b) abnormal.

FIGURE 10: Average recognition accuracy of loudspeaker validation set under different r .

each sample contains one complete period at least, as shown in Figure 8, the raw sound signal is randomly segmented to obtain the segmented samples with 8000 sampling points. Then, the down-sampling process is performed on the segmented samples and each sample is of 1600 sampling points. And it can be seen from Figure 8 that the down-sampling sample not only reduces the computational complexity but also retains the waveform of the segmented sample. Finally, we can obtain a total of 2000 samples, which are divided into the training set, validation set, and testing set at a ratio of 6 : 2 : 2. More sample information is shown in Table 7.

4.2. Fault Diagnosis. The time-frequency images obtained by MSSST are shown in Figure 9. They have concentrated energy, but the dimension is 1600×3200 , and there is a lot of irrelevant background information. Using the whole image as a raw sample for dictionary learning will consume a lot of computing resources and time, and redundant information will be added to the generated dictionary. Therefore, we can extract the optimal features of time-frequency images by FDL.

Figure 10 shows the average recognition accuracy of the validation set under different r values. It can be seen that the optimal value of r is 46. At this time, the average recognition accuracy of the validation set is 99%. This model is applied to the testing set. The experimental results show that 2 normal samples are misdiagnosed as the abnormal state and 1 abnormal sample is misdiagnosed as the normal state. Namely, the recognition accuracy is 98.5%.

In reference [41], the second-order time-reassigned multisynchrosqueezing transform was used to obtain the time-frequency images of the loudspeaker sound signals, and CNN was adopted to implement the feature extraction and fault recognition. The accuracy was 98.25%. By comparison, the proposed method is not inferior to the deep learning method.

5. Conclusions

In this paper, a new fault diagnosis method for the rolling bearing, which is called MSSST-FDL, is proposed to boost the speed and accuracy of recognition. Experiments show that the MSSST has better energy concentration than other time-frequency analysis methods; time-frequency images

with better energy concentration can improve the quality of fault diagnosis; the dictionary learning and feature coding of LBP feature vectors are faster than those of the whole time-frequency images, which can not only quickly determine the optimal hyperparameter but also meet the real-time requirement of fault diagnosis. The effectiveness of the proposed method is verified in CWRU and MFPT datasets, and the fault recognition accuracy is 99% and 97.85%, respectively. Furthermore, we apply the proposed method to loudspeaker anomaly diagnosis, and the recognition accuracy reaches 98.5%, which indicates that our method has the potential to be applied to other equipment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant no. 51775177).

References

- [1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106587, 2020.
- [2] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization gan for unbalanced data," *Knowledge-Based Systems*, vol. 187, Article ID 104837, 2020.
- [3] Z. He, H. Shao, P. Wang, J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowledge-Based Systems*, vol. 191, Article ID 105313, 2020.
- [4] M. Zeng, W. Zhang, and Z. Chen, "Group-based k-svd denoising for bearing fault diagnosis," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6335–6343, 2019.
- [5] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, 2018.
- [6] D. Gabor, "Theory of communication. part I: the analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.
- [7] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, pp. 235–238, 1977.
- [8] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory-part I: complex signal and scattering in multilayered media," *Geophysics*, vol. 47, no. 2, pp. 203–221, 1982.
- [9] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: the S transform," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 998–1001, 1996.
- [10] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.
- [11] L. Cohen, *Time-Frequency Analysis: Theory and Applications*, Prentice-Hall, Upper Saddle River, NJ, USA, 1995.
- [12] I. Daubechies, J. Lu, and H. Wu, "Synchrosqueezed wavelet transforms: a tool for empirical mode decomposition," 2009, <http://arxiv.org/abs/0912.2437>.
- [13] F. Auger, P. Flandrin, Y.-T. Lin et al., "Time-frequency reassignment and synchrosqueezing: an overview," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013.
- [14] Z.-l. Huang, J. Zhang, T.-h. Zhao, and Y. Sun, "Synchrosqueezing s-transform and its application in seismic spectral decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 817–825, 2016.
- [15] G. Yu, Z. Wang, and P. Zhao, "Multisynchrosqueezing transform," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 7, pp. 5441–5455, 2019.
- [16] K. Lin, *Research on fine-grained fault identification algorithm of rotating machinery vibration signal based on time-frequency image and multi-level sparse coding*, Ph.D. thesis, Hubei University of Technology, Wuhan, China, 2020.
- [17] S. Guo, T. Yang, W. Guo, and C. Zhang, "A novel fault diagnosis method for rotating machinery based on a convolutional neural network," *Sensors*, vol. 18, no. 5, 2018.
- [18] M.-F. Guo, N.-C. Yang, and W.-F. Chen, "Deep-Learning-based fault classification using hhlbert-huang transform and convolutional neural network in power distribution systems," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6905–6913, 2019.
- [19] S. Duan, H. Zheng, and J. Liu, "A novel classification method for flutter signals based on the CNN and STFT," *International Journal of Aerospace Engineering*, vol. 2019, 8 pages, Article ID 9375437, 2019.
- [20] Q. Li, X. Ding, Q. He, W. Huang, and Y. Shao, "Manifold sensing-based convolution sparse self-learning for defective bearing morphological feature extraction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3069–3078, 2021.
- [21] Y. Wang, X. Ding, Q. Zeng, L. Wang, and Y. Shao, "Intelligent rolling bearing fault diagnosis via vision convnet," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6600–6609, 2021.
- [22] B. Li, P.-l. Zhang, D.S. Liu, S.-S. Mi, G.-Q. Ren, and H. Tian, "Feature extraction for rolling element bearing fault diagnosis utilizing generalized s transform and two-dimensional non-negative matrix factorization," *Journal of Sound and Vibration*, vol. 330, no. 10, pp. 2388–2399, 2011.
- [23] B. Li, P.-l. Zhang, H. Tian, S.-s. Mi, D.-s. Liu, and G.-q. Ren, "A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10000–10009, 2011.
- [24] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.
- [25] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [26] X. Yu, E. Ding, C. Chen, and L. Li, "Rolling bearing fault state recognition method based on hht and supervised sparse coding," *Journal of China Coal Society*, vol. 40, no. 11, 2015.
- [27] G. D. Sun, Y. Gao, K. Lin, and Y. Hu, "Fine-grained fault diagnosis method of rolling bearing combining multi-synchrosqueezing transform and sparse feature coding based on dictionary learning," *Shock and Vibration*, vol. 2019, 13 pages, Article ID 1531079, 2019.

- [28] X. Li, Y. Yu, H. D. Shao, X. Zhong, J. Cheng, and J. s. Cheng, "Symplectic weighted sparse support matrix machine for gear fault diagnosis," *Measurement*, vol. 168, Article ID 108392, 2020.
- [29] X. Li, Y. Yu, N. Q. Hu, Z. Cheng, and J. Cheng, "Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis," *Knowledge-Based Systems*, vol. 211, Article ID 106507, 2021.
- [30] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [31] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [32] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [33] K. A. Loparo and K. A. Loparo, *Bearing Data Center*, Case Western Reserve University, Cleveland, OH, USA, 2013, <http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-datacenter-website..>
- [34] E. Bechhoefer, "A quick introduction to bearing envelope analysis, MFPT Data," 2016, <http://www.mfpt.org/FaultData/FaultData.htm>.
- [35] S. Aviyente and W. J. Williams, "Minimum entropy time-frequency distributions," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 37–40, 2005.
- [36] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "measuring time-frequency information content using the renyi entropies," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1391–1409, 2001.
- [37] T. A. Classen and M. Wolfgang, "The wigner distribution-a tool for time-frequency signal analysis, part ii: discrete-time signals," *Philips Journal of Research*, vol. 35, no. 3, pp. 276–300, 1980.
- [38] Z. Yuan, L. B. Zhang, L. X. Duan, and T. Li, "Intelligent fault diagnosis of rolling element bearings based on hht and cnn," in *Proceedings of the Prognostics and System Health Management Conference*, pp. 292–296, Qingdao, China, 2019.
- [39] W. L. Du, J. F. Tao, Y. M. Li, and C. L. Liu, "Wavelet leaders multifractal features based fault diagnosis of rotating mechanism," *Mechanical Systems and Signal Processing*, vol. 43, no. 1-2, pp. 57–75, 2014.
- [40] V. David, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock and Vibration*, vol. 2017, Article ID 5067651, 17 pages, 2017.
- [41] G. D. Sun, Y. Gao, Y. Xu, and W. Feng, "Data-driven fault diagnosis method based on second-order time-reassigned multisynchrosqueezing transform and evenly mini-batch training," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.