applied sciences

MDPI

Article

The Average Coding Length of Huffman Coding Based Signal Processing and Its Application in Fault Severity Recognition

Jiancheng Yin, Mingjia Lei, Huailiang Zheng, Yuantao Yang, Yuqing Li and Minqiang Xu *

Deep Space Exploration Research Center, Harbin Institute of Technology, Harbin 150001, China; wdyydy@163.com (J.Y.); hitleimingjia@163.com (M.L.); hlzhenghit@126.com (H.Z.); yangyuantaohit@163.com (Y.Y.); bradley@hit.edu.cn (Y.L.)

* Correspondence: xumq@hit.edu.cn

Received: 23 October 2019; Accepted: 20 November 2019; Published: 22 November 2019

Abstract: The transient impact components in vibration signal, which are the major information for bearing fault severity recognition, are often interfered with by ambient noise. Meanwhile, for bearing fault severity recognition, the frequency band selection methods which are employed to pre-process the contaminated vibration signal only select the partial frequency band of the vibration signal and cause information loss of other frequency band. Aiming at this issue, this paper proposes a novel fault severity recognition method based on Huffman coding, which can retain all the information of the frequency band, and is applied for the first time to bearing fault severity recognition. Specifically, the average coding length of Huffman coding (ACLHC) of the original vibration signal is first calculated to reduce the noise and highlight the impact components of the signal. Then, the ACLHC is encoded by symbolic aggregate approximation (SAX) to reflect the modulation information of bearing. Finally, the Lempel-Ziv indicator (LZ indicator) of the symbol sequence is calculated to reflect the fault severity. The proposed method is verified by the bearing datasets under different working conditions. Compared with the methods based on frequency band selection, the proposed method effectively recognizes the fault severity of bearing for more working conditions.

Keywords: signal processing; Lempel-Ziv indicator; average coding length of Huffman coding; symbolic aggregate approximation; bearing fault severity

1. Introduction

As an important part of rotating machinery, monitoring the health condition of bearings is very necessary during the industrial productions. Once a local fault occurs on the surface of a bearing, a series of impacts are aroused by the strikes between the rolling elements and the local fault on the outer or inner race [1]. However, the impacts caused by the local fault may be submerged in the noise as the severe environment. Therefore, effectively highlighting the impact components of the signal is beneficial to improving the recognition results of bearing fault severity.

Recently, the importance of bearing fault severity recognition has been paid more and more attention [2,3]. In general, the damage severity of bearings is usually measured from the aspect of signal complexity [4]. Currently, as the effective method for fault severity recognition, the Lempel-Ziv indicator (LZ indicator) is frequently used for the recognition of bearing fault severity. The LZ indicator can effectively reflect the change in frequency component in signals, and the different fault severities of the inner and outer race of the bearing will cause the change in the frequency component [4]. Thus, The LZ indicator can be effectively used for the recognition of bearing fault severity. However, although the LZ indicator can effectively recognize the fault severity of the inner and outer race of the bearing, the recognition results are often affected by the environmental noise. Therefore, based on Lempel-Ziv complexity, various methods of noise reduction have been used to improve the recognition results of bearing fault severity. To date, continuous wavelet transform [5], Empirical Mode Decomposition [6], local mean decomposition [7,8], intrinsic time-scale decomposition [9], double-dictionary matching pursuit [10], and sparsogram [11] have been used in combination with the Lempel-Ziv complexity to reduce the noise and improve the recognition results of fault severity. Therefore, reducing noise and improving signal-to-noise ratio are helpful for improving the fault severity recognition results of the LZ indicator.

Recently, the frequency band selection methods have been widely used to improve the signal-to-noise ratio and to highlight the impact components of signal. As a common time–frequency method, wavelet transform is used to decompose and extract the effective frequency components of the signals. For examples, Wang et al. [12] combined the Hilbert and wavelet transforms to improve the results of bearing fault identification. In order to extract the effective information of the signal effectively, spectral kurtosis was first used to select that of the signal. For example, Tian et al. [13] applied the K-Nearest Neighbor (KNN) distance measurement to bearing fault detection based on the multiple features which, extracted by cross-correlation, improve spectral kurtosis. Saidi et al. [14] applied support vector regression to estimate the residual useful life (RUL) of the high-speed shaft bearings based on the time indicators derived from spectral kurtosis. Based on the high potential of spectral kurtosis in detecting and characterizing non-stationary signals, Antoni et al. [15] proposed a new method called Kurtogram, which improved spectral kurtosis using short-time Fourier transform, depending on a series of different window lengths. In addition, Kurtogram has been continuously improved to improve the ability to select the effective fault information. For example, Wang et al. [16] applied manifold learning to overcome the drawbacks of Kurtogram, in which the in-band noise was left unprocessed. Wang et al. [17] improved Kurtogram based on a meshing frequency modulation index which utilized the particular gearbox related phenomenon to extract the bearing fault-induced impact components under the background noise of the planetary gearbox. Besides this, other new methods have been used in combination with the original methods to improve the extraction results. Barszcz et al. [18] proposed a new method called Protrugram, which calculates the kurtosis of the envelope spectrum amplitudes of demodulated signals instead of the kurtosis of the filtered temporal signals when the signal to noise ratio is low. Based on the sparsity measurement, Tse et al. [19] proposed Sparsogram, which can detect the high resonant frequency to extract bearing fault feature and then apply the genetic algorithm and Morlet wavelet to enhance the detection results of Sparsogram. Miao et al. [20] identified the optimal frequency band under the interferences of the motor and industrial field based on the singular value negentropy. These methods have achieved good results in fault diagnosis and feature frequency selection.

Although these frequency band selection methods can reduce noise and have a good effect in feature frequency selection and fault diagnosis, they also ignore some effective information out of the selected frequency band. Meanwhile, as the transient impact caused by bearing fault can spread within a wide frequency band [16], the above methods cannot retain all the information caused by bearing fault. Although these frequency band selection methods can improve the recognition results of the LZ indicator to some extent, there could be some deviations in fault severity recognition based on the LZ indicator.

Aiming to resolve this problem, in this paper, a fault severity recognition method of bearing based on the full frequency band of vibration signal is proposed. According to the average coding length of Huffman coding (ACLHC), the original signal is processed to reduce the noise and highlight the impact components of the signal. The signals processed by ACLHC not only reduce the influence of noise, but also contain all the fault information. Then, the ACLHC of the signal is further encoded by symbolic aggregate approximation (SAX) to reflect the modulation information of bearing vibration signal. Finally, the LZ indicator is obtained to reflect the fault severity. After the above processing and calculation, the bearing fault severity under different working conditions can be effectively recognized.

The rest of this paper is organized as follows. Section 2 introduces the basic theory of ACLHC and symbolic aggregate approximation Lempel-Ziv indicator (SAX-LZ). In Section 3, the new recognition method of bearing fault severity is proposed and the characteristic of the ACLHC is discussed. In Section 4, the effectiveness of the proposed method is validated by the bearing datasets under different working conditions. Section 5 concludes this paper.

2. Theory Background

2.1. The Theory of ACLHC

Huffman coding is a variable-length source coding method which was proposed by Huffman in 1951 [21]. Huffman coding has been widely used in computer, data encryption, and communication fields because of its high efficiency. Huffman coding is a prefix code, which minimizes the average coding length [22]. The symbols with high probability are represented by short code words, while the symbols with low probability are represented by long code words. So, Huffman coding depends on the occurring probability of symbols. The closer the probability of each symbol appears, the higher the average coding length of Huffman coding is. For a symbol sequence, the ACLHC can be obtained as follows [23,24].

Step 1: Count the occurring probability $\{p_1, p_2, \cdots, p_m\}$ of each symbol in the symbol sequence. Construct the binary tree set $F = \{T_1, T_2, \cdots, T_m\}$ based on the statistical probability $\{p_1, p_2, \cdots, p_m\}$. In each binary tree T_i, there is only one root node with the probability p_i. The left and right subtrees of the binary tree are empty;

Step 2: The two trees with minimum root node probability in the binary tree set F are selected as the left and right subtrees to construct a new binary tree. The probability of the root node of the new binary tree is the sum of the probabilities of the left and right root nodes of the subtree;

Step 3: Delete the two trees with the minimum probability in the binary tree set F and add the new binary tree to F;

Step 4: Repeat Step 2 and Step 3 until there is only one tree in the binary tree set F;

Step 5: After the Huffman tree is constructed, the left child nodes of each father node are coded as 1 (0) and the right child nodes of each father node are coded as 0 (1). Search forward from the last binary tree, the Huffman coding of each symbol $\{H_1, H_2, \cdots, H_m\}$ is the set of all codes on the corresponding path;

Step 6: The average coding length of Huffman coding is obtained according to the length of the Huffman coding of each symbol and the corresponding probability, as follows:

$$ACL = \sum_{i=1}^{m} p_i \cdot length(H_i), \tag{1}$$

where ACL is the average coding length of Huffman coding, $length(H_i)$ is the number of 0 or 1 contained in H_i.

To better illustrate the calculation process of ACLHC, suppose a probability sequence {0.35, 0.30, 0.15, 0.10, 0.10}, the Huffman tree and corresponding Huffman coding is shown in Figure 1.

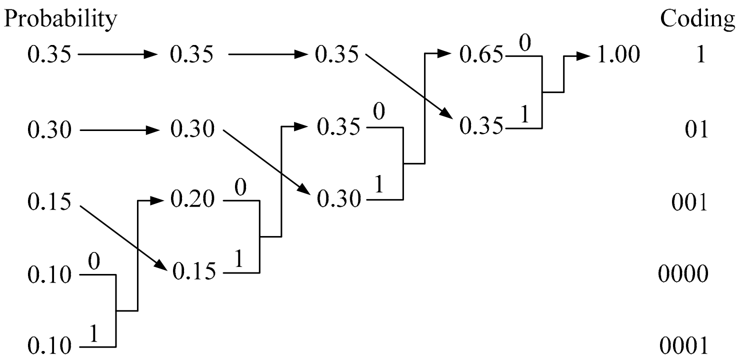


Figure 1. The Huffman tree and Huffman coding.

The average coding length of the probability sequence is calculated by (1). Therefore, the average coding length of the probability sequence is 2.2:

$$ACL = 0.35 \cdot 1 + 0.30 \cdot 2 + 0.15 \cdot 3 + 0.10 \cdot 4 + 0.10 \cdot 4 = 2.2.$$

2.2. The Theory of SAX-LZ

The Lempel-Ziv complexity, which was proposed by Lempel and Ziv [25], has been widely used to recognize the fault severity of bearing, because signals of bearing with different fault severities have different complexities. In general, the LZ indicator is calculated based on a 0–1 sequence [26]. However, the application of SAX to code the time series in the calculation of LZ cannot only describe the details of time series, but also improve the computational efficiency. For a given time series $X = \{x_1, x_2, \cdots, x_n\}$, the process of the SAX-LZ is given as follows [27].

Step 1: The time series is normalized according to the equation, as follows:

$$NX = \frac{X - \mu}{\sigma}, \tag{2}$$

where NX is the normalized series of X, μ is the mean of the time series, and σ is its standard deviation;

Step 2: According to the below equation, the normalized time series is represented by the average of each segment, which is divided according to Piecewise Aggregate Approximation (PAA):

$$\overline{x_i} = \frac{N}{n} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j, \tag{3}$$

where $\overline{x_i}$ is the average of the i^{th} segment, x_j is one point of time series X, j is the sequence number for each segment, N is the number of equal sized segments;

Step 3: According to Table 1, determine the breakpoints β_i according to the equiprobable regions α of the distribution space to be divided;

Table 1. The lookup table of breakpoints.

β \ α	3	4	5	6	7	8	9	10
β1	−0.43	−0.67	−0.84	−0.97	−1.07	−1.15	−1.22	−1.28
β2	0.43	0	−0.25	−0.43	−0.57	−0.67	−0.76	−0.84
β3		0.67	0.25	0	−0.18	−0.32	−0.43	−0.52
β4			0.84	0.43	0.18	0	−0.14	−0.25
β5				0.97	0.57	0.32	0.14	0
β6					1.07	0.67	0.43	0.25
β7						1.15	0.76	0.52
β8							1.22	0.84
β9								1.28

Step 4: The symbolic sequence $\{s_i\}_{i=1}^{N}$ is obtained based on the symbol which is assigned according to the region divided by the breakpoints;

Step 5: Initialization, set $S_{v,0} = \{\}$, $Q_0 = \{\}$, $C_N = 0$, $r = 1$;

Step 6: Take $Q_r = (Q_{r-1}s_r)$ and judge whether Q_r belongs to $S_{v,r-1} = \{S_{v,r-2}s_{r-1}\}$. If not, $Q_r = \{\}$, then $C_N(r) = C_N(r-1) + 1$, $r = r + 1$. If so, $Q_r = (Q_{r-1}s_r)$, $C_N(r) = C_N(r-1)$, $r = r + 1$, and repeat this step until symbolic sequence is completely covered;

Step 7: The Lempel-Ziv complexity is normalized by:

$$0 \le C_{nN} = \frac{C_N(N)}{C_{UL}} \le 1, \tag{4}$$

where,

$$C_{UL} = \lim_{N\to\infty} C_N(N) = \lim_{N\to\infty} \frac{N}{(1-\beta)\log_k N} \approx \frac{N}{\log_k N}, \tag{5}$$

where k is the number of the alphabets (for binary string, $k = 2$).

3. The Proposed Method

SAX-LZ can effectively recognize the fault severity of a bearing. Nevertheless, the noise in the vibration signal limits the accuracy of SAX-LZ in the fault severity recognition of bearing. However, the ACLHC can effectively highlight the mutation components caused by bearing fault in the signal and reduce the noise in the signals. Therefore, the ACLHC can be effectively used to process the vibration signal of bearings to highlight the impact components and reduce the interference of noise. The recognition accuracy of SAX-LZ can be effectively improved after processing by ACLHC.

Given a signal $X = \{x_1, x_2, \cdots, x_n\}$, the SAX-LZ of the signal processed by ACLHC is calculated as follows:

Step 1: According to Equation (6), transform the signal amplitude to the positive value $X' = \{x'_1, x'_2, \cdots, x'_n\}$:

$$x'_i = x_i - \min(X); \tag{6}$$

Step 2: Then, the signal $X' = \{x'_1, x'_2, \cdots, x'_n\}$ is divided into different segments $SX = \{sx_1, sx_2, \cdots, sx_{n-l+1}\}$ according to the sliding window. Where $sx_i = \{x'_i, x'_{i+1}, \cdots, x'_{i+l-1}\}$, l is the length of the sliding window;

Step 3: Normalized segment data $SP = \{P_1, P_2, \cdots, P_{n-l+1}\}$ are calculated by Equation (7), where $P_i = \{p_i, p_{i+1}, \cdots, p_{i+l-1}\}$:

$$p_i = \frac{x'_i}{\max(sx_i)}; \tag{7}$$

Step 4: Calculate the ACLHC $C = \{c_1, c_2, \cdots, c_{n-l+1}\}$ of the normalized segment data $SP = \{P_1, P_2, \cdots, P_{n-l+1}\}$ according to Section 2.1;

Step 5: The ACLHC C is normalized to obtain the normalized series $NC = \{Nc_1, Nc_2, \cdots, Nc_{n-l+1}\}$ by the Equation (2);

Step 6: The SAX-LZ of the normalized series NC is obtained according to the description in Section 2.2.

The calculation process is illustrated in Figure 2.

To illustrate the characteristics of the ACLHC in the description of impact, two sets of simulation signals with different SNR are constructed, as follows:

$$X(t) = x(t) + n(t), \tag{8}$$

where $X(t)$ is the simulation signal with different signal-to-noise ratios (SNR), the waveform of the $X(t)$ with SNR of 13dB is shown in Figure 3b, while that with SNR of 3dB is shown in Figure 4b. $n(t)$ is the noise, $x(t)$ is the simulation signal without noise, the $x(t)$ is shown as follows:

$$x(t) = \left(\sum_{k=-\infty}^{+\infty} d \cdot \delta(t - zT_o) * e(t) \right) \cdot \sin(2\pi f_n), \tag{9}$$

where d is the intensity of the impact forces, $\delta(t)$ is the unit impact function, z is the number of the impacts, $T_o = 1/f_o$ is the interval between two impacts, f_o is the characteristic frequency of the bearing, $e(t)$ is the attenuation function, f_n is the resonance frequency. The waveform of $x(t)$ is shown in Figures 3a and 4a.

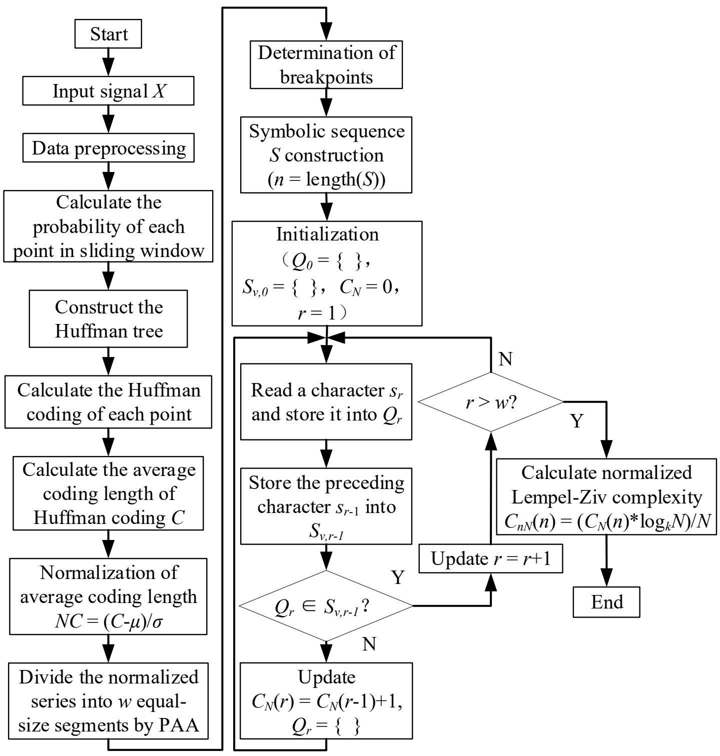


Figure 2. The calculation process of the proposed method.

In order to better illustrate the effect of ACLHC in noise reduction, the ACLHC was compared with other three frequency band selection methods (protrugram, sparsogram, and genetic algorithm sparsogram (GA-sparsogram)). The results of the simulation signals processed by the above four methods are shown in Figures 3 and 4.

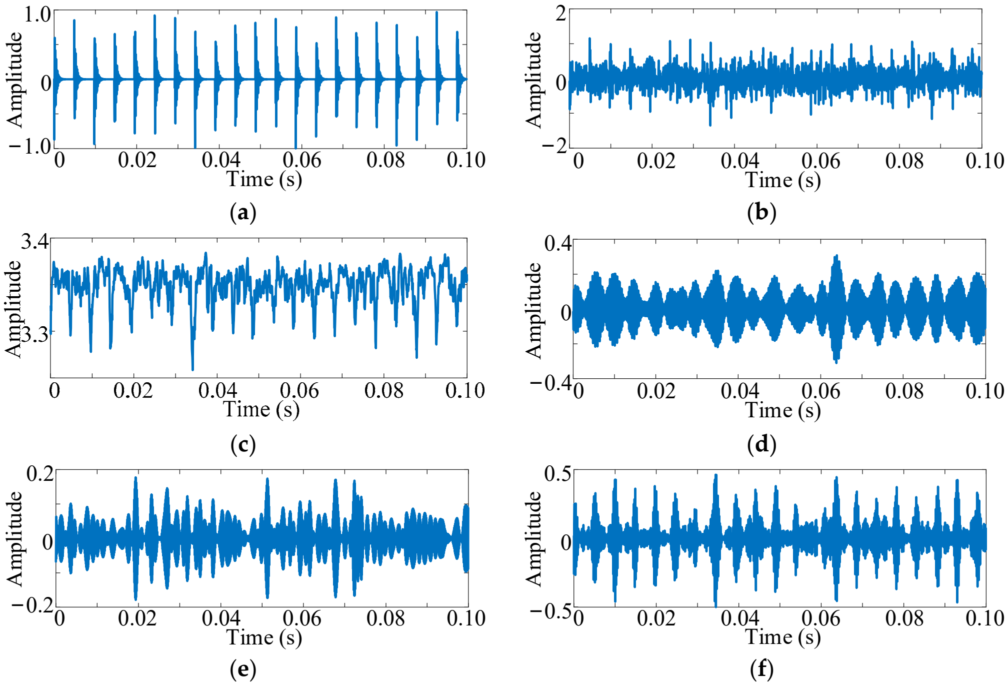


Figure 3. The waveforms at the signal-to-noise ratio (SNR) of 13dB. (**a**) Impact signal; (**b**) the impact signal with noise; (**c**) average coding length of Huffman coding (ACLHC); (**d**) protrugram; (**e**) sparsogram; (**f**) genetic algorithm sparsogram (GA-sparsogram).

For the simulation signal with high SNR, as shown in Figure 3a,c, as the amplitude of the signal was more dispersed when the impact occurred, the probability of the signal was dispersed at this time. So, the ACLHC of the impact component is lower than that of other components when there are impacts in the signals. Although there are some differences between Figure 3a,c, the location of each impact can be restored by ACLHC well. However, although the approximate shape of the impact could be restored by GA-sparsogram, as shown in Figure 3f, the number of impacts was less than that of the impact signal in Figure 3a. In addition, the location and shape of the impact in Figure 3d,e were completely different from that in Figure 3a. Therefore, compared with protrugram, sparsogram, and GA-sparsogram, the ACLHC can effectively restore the location of the impact and highlight the impact components of signal.

For the simulation signal with low SNR, the time-domain waveforms and frequency spectrums of the simulation signal processed by different methods are shown in Figure 4.

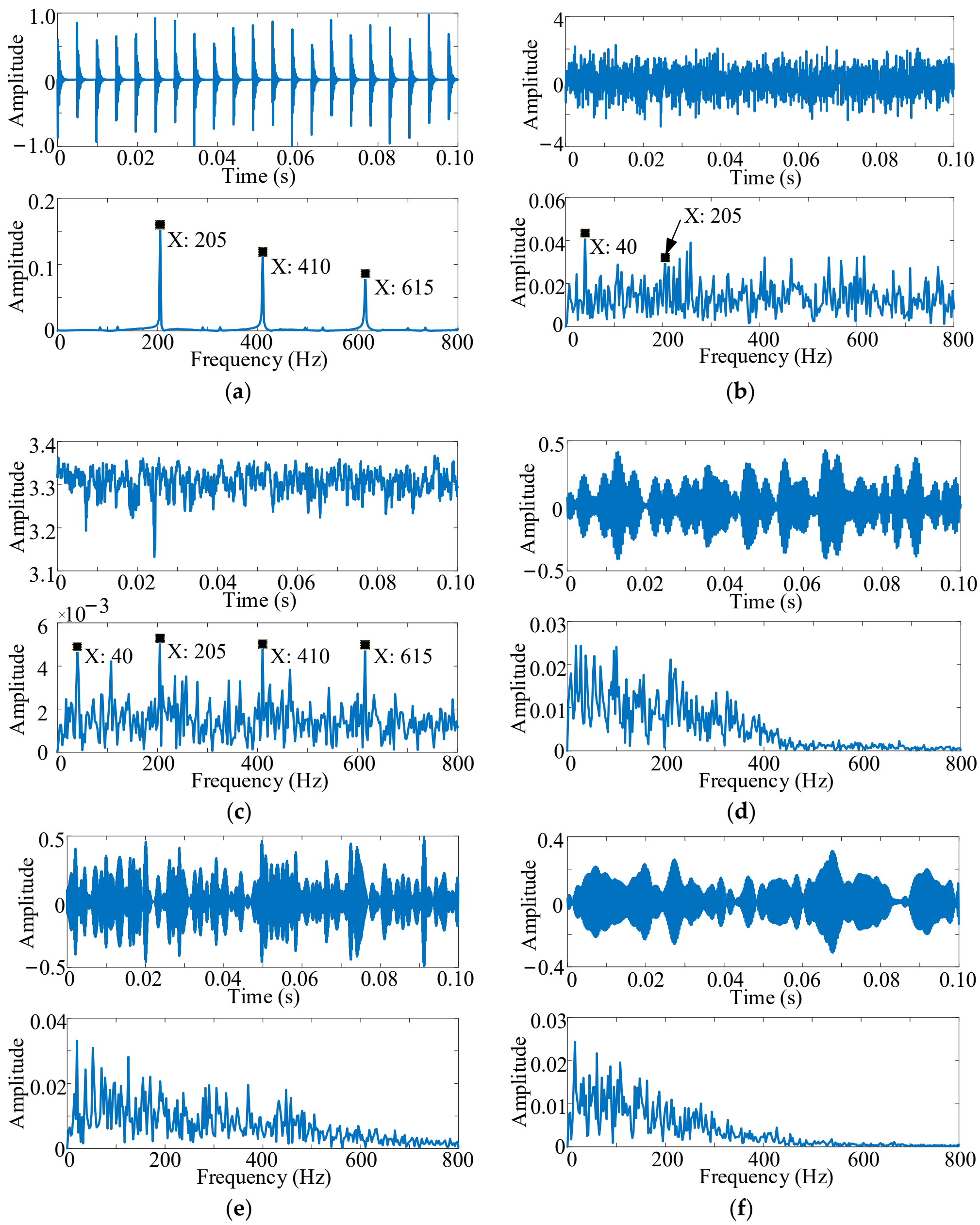


Figure 4. The waveforms and frequency spectrums at the SNR of 3 dB. (**a**) Impact signal; (**b**) the impact signal with noise; (**c**) average coding length of Huffman coding (ACLHC); (**d**) protrugram; (**e**) sparsogram; (**f**) genetic algorithm sparsogram (GA-sparsogram).

As shown in Figure 4b, due to the noise interference, the impact signal was completely submerged in the noise in the time-domain waveform. Besides this, the obvious characteristic frequency of the impact signal also could not be found in the frequency spectrum. The additional frequency component of 40 Hz was introduced by the noise. Although the obvious impact component cannot be found in the time-domain waveform by ACLHC in Figure 4c, the obvious characteristic frequency of the impact signal can already be found in the frequency spectrum. However, the characteristic frequency of the impact signal could not be found by protrugram, sparsogram, and GA-sparsogram, as shown in Figure 4d–f. Besides this, the high-frequency part of the frequency spectrum was also lost, while the frequency spectrum of ACLHC contained the whole frequency components. The information after 400 Hz of the frequency band processed by protrugram and GA-sparsogram was lost, as shown in Figure 4d,f. After 600 Hz, the frequency band processed by Sparsogram was lost, as shown in Figure 4e.

Therefore, the ACLHC not only highlights the impact components of the signal, but also retains the whole frequency components of signal.

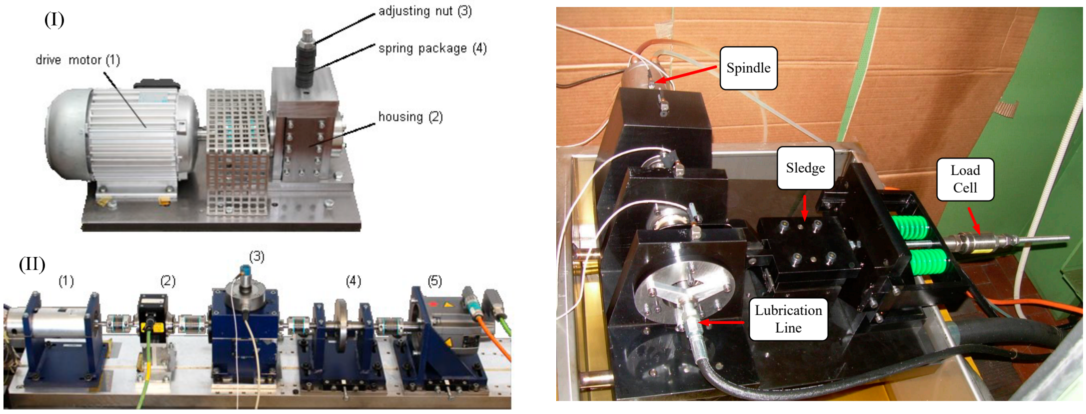
Altogether, compared with protrugram, sparsogram, and GA-sparsogram, the ACLHC can effectively restore the location of impact in the signal when the SNR is high. Even if the ACLHC cannot obviously restore the location of impact in the signal when the SNR is low, it can also effectively reflect the characteristic frequency of the impact in the frequency spectrum.

4. Result and Discussion

4.1. The Description of the Datasets

4.1.1. Case Study 1

The data are from the Konstruktions-und Antriebstechnik (KAt), in the school of Mechanical Engineering at University Paderborn [28]. The bearing test rig is shown in Figure 5a (II). The modular test rig consisted of an electric motor, a torque-measurement shaft, a rolling bearing test module, a flywheel, and a load motor. The bearings 6203—with different types of damage caused by an accelerated lifetime test on the apparatus, as shown in Figure 5a (I)—were mounted in the bearing test module (as shown in Figure 5a (II)) to generate the experimental data. The severity of the damage was described by the percentage of length relative to pitch circumference. The three levels for 6203 are shown in Table 2 [28].



(**a**) (**b**)

Figure 5. The bearing test rig. (**a**) The test rig of Konstruktions-und Antriebstechnik (Kat); (**b**) The test rig of Dynamic and Identification Research Group (DIRG): (I) apparatus for accelerated life time test to obtain the damaged bearings; (II) modular test rig to obtain the vibration signals under different working conditions: (1) an electric motor; (2) a torque-measurement shaft; (3) a rolling bearing test module; (4) a flywheel; and (5) a load motor.

Table 2. The damage levels to determine the severity of damage.

Damage Level	Percentage Values	Limits for Bearing 6203
1	0–2%	≤ 2 mm
2	2–5%	> 2 mm
3	5–15%	> 4.5 mm

The information of the test bearings with real damages caused by the accelerated lifetime test is shown in Table 3. The rotational speed of the drive system, the radial force onto the test bearing, and the load torque in the drive train were the main operating parameters. The parameters were defined as shown in Table 4. All three parameters were kept constant for the time of each measurement.

Table 3. The test bearings with real damages caused by accelerated lifetime test.

Severity of Damage	Location of Fault	Characteristic of Damage	Arrangement	Damage
1	Inner race	Single point	No repetition	Fatigue: pitting
2	Inner race	Single point	No repetition	Fatigue: pitting
3	Inner race	Single point	No repetition	Fatigue: pitting

Table 4. The operating parameters.

No.	Rotational Speed (rpm)	Load Torque (Nm)	Radial Force (N)	Name of Setting
1	900	0.7	1000	N09-M07-F10
2	1500	0.7	400	N15-M07-F04

4.1.2. Case Study 2

The experimental data came from the Dynamic and Identification Research Group (DIRG), in the Department of Mechanical and Aerospace Engineering at Politecnico di Torino [29]. The bearing test rig is shown in Figure 5b. The test rig mainly contained three major parts: a high-speed spindle, a load cell, and a lubrication part. The bearings of the spindle, for which the main geometrical properties of the bearings are listed in Table 5, were grease lubricated and their temperature was limited by a liquid (glycol/water) refrigeration circuit. Two accelerometers were installed on the key position. The sampling frequency was 51,200 Hz.

Table 5. The main properties of the rolling bearings.

	Pitch Diameter D (mm)	Rollers Diameter D (mm)	Contact Angle Φ (°)	Rolling Elements Z
Size	40.5	9.0	0	10

The bearing fault with artificial damage occurred on the inner race. The diameter of an indentation on the inner race was 150, 250, and 450 μm. The application of the static load was 1000, 1400, and 1800 N. The rotational frequency of the shaft increased from 100 Hz to 400 Hz with steps 100 Hz. The speed–load combinations of the bearing data used below are shown in Table 6.

Table 6. The list of the used load and speed cases.

Nominal Load (N)	Nominal Speed (Hz)			
1000	100	200	300	400
1400	-	200	300	400
1800	100	200	300	-

4.2. *Fault Severity Recognition by the Proposed Method*

In order to verify the effectiveness of the proposed method in fault severity recognition, the proposed method was used to calculate the complexity of the above bearing datasets. The results of LZ and SAX-LZ were also used to compare with the proposed method. The calculated results of the two datasets are shown in Figures 6 and 7, respectively.

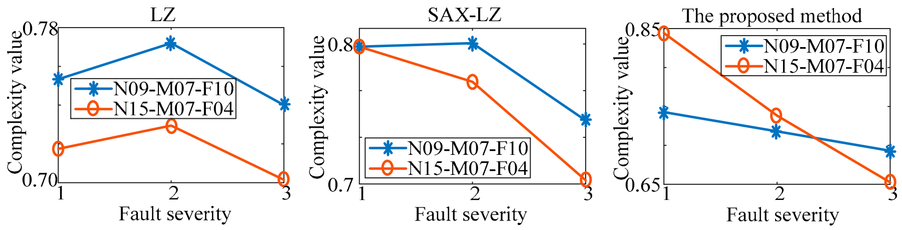


Figure 6. The complexity values of Konstruktions-und Antriebstechnik datasets under different working conditions.

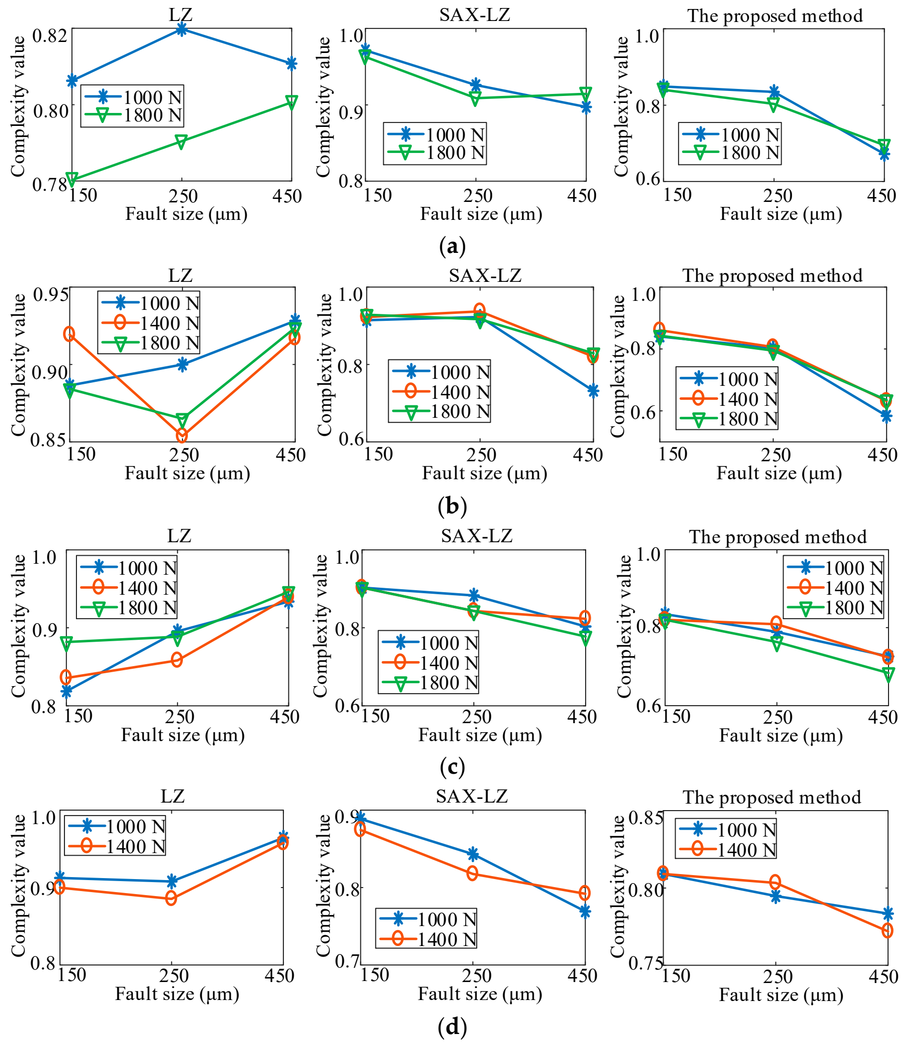


Figure 7. The complexity values of Dynamic and Identification Research Group datasets under different working conditions. (**a**) 100 Hz; (**b**) 200 Hz; (**c**) 300 Hz; and (**d**) 400 Hz.

According to Yan [4], the complexity value should decrease with an increase in the fault severity of the bearing inner race. However, as shown in Figures 6 and 7, the complexity values of the traditional LZ method did not decrease with the increase in the fault severity of bearing inner race. Although the complexity values of the SAX-LZ decreased with the increase in the fault severity of bearing inner race under most working conditions, there were also some non-monotonous trends in the complexity values obtained by SAX-LZ, such as the N09-M07-F10 in KAt datasets and the 100 Hz and 200 Hz in DIRG datasets. However, the complexity values of the proposed method showed all monotonous decreasing trends under all working conditions. Therefore, for the problem of fault severity recognition, although the monotonicity of SAX-LZ was better than that of LZ, that of SAX-LZ was also affected by noise and interference. However, the SAX-LZ could effectively recognize the fault severity of bearing after the signal was processed by ACLHC.

Therefore, the ACLHC can effectively reduce the influence of noise and interference and highlight the fault information. The proposed method can be effectively used to recognize the fault severity of bearings.

4.3. Comparison with Other Frequency Band Selection Methods

In order to further illustrate the advantages of the proposed method in the fault severity recognition of bearings, the SAX-LZ of the bearing vibration signals processed by three frequency band selection methods (protrugram, sparsogram, and GA-sparsogram) was employed to compare with the proposed method. The calculated results of the two datasets are shown in Figures 8 and 9.

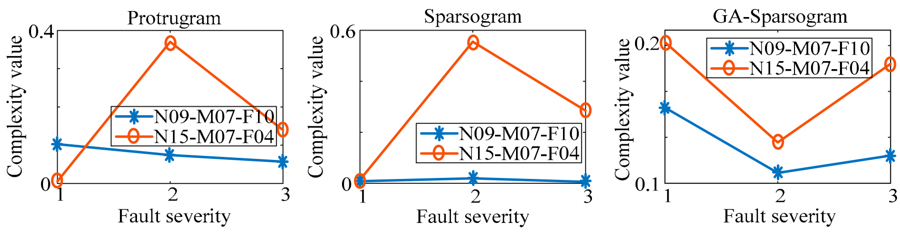


Figure 8. The complexity values of Konstruktions-und Antriebstechnik datasets after different treatment.

As shown in Figures 8 and 9, compared with SAX-LZ, although the monotonicity of SAX-LZ could be improved by the three frequency band selection methods in some cases, the overall monotonicity was not better than that of SAX-LZ without treatment. The reason for this could be that although the above frequency band selection methods could effectively select the frequency band which contained the most fault information, some useful information was also lost. Therefore, the calculated results based on the frequency band selection methods might not be better than those of SAX-LZ without treatment. However, the ACLHC did not need to select the frequency band of the signal in the process of reducing the influence of noise and highlighting the impact components of signal. Therefore, the signals processed by ACLHC cannot only reduce the influence of noise, but also contain all the fault information. The calculated results of the proposed method were better than those of the SAX-LZ of the bearing vibration signals processed by the three frequency band selection methods. Therefore, the signals processed by ACLCH can effectively improve the recognition results of bearing fault severity. The proposed method can be effectively used to recognize the fault severity of bearings under different working conditions.

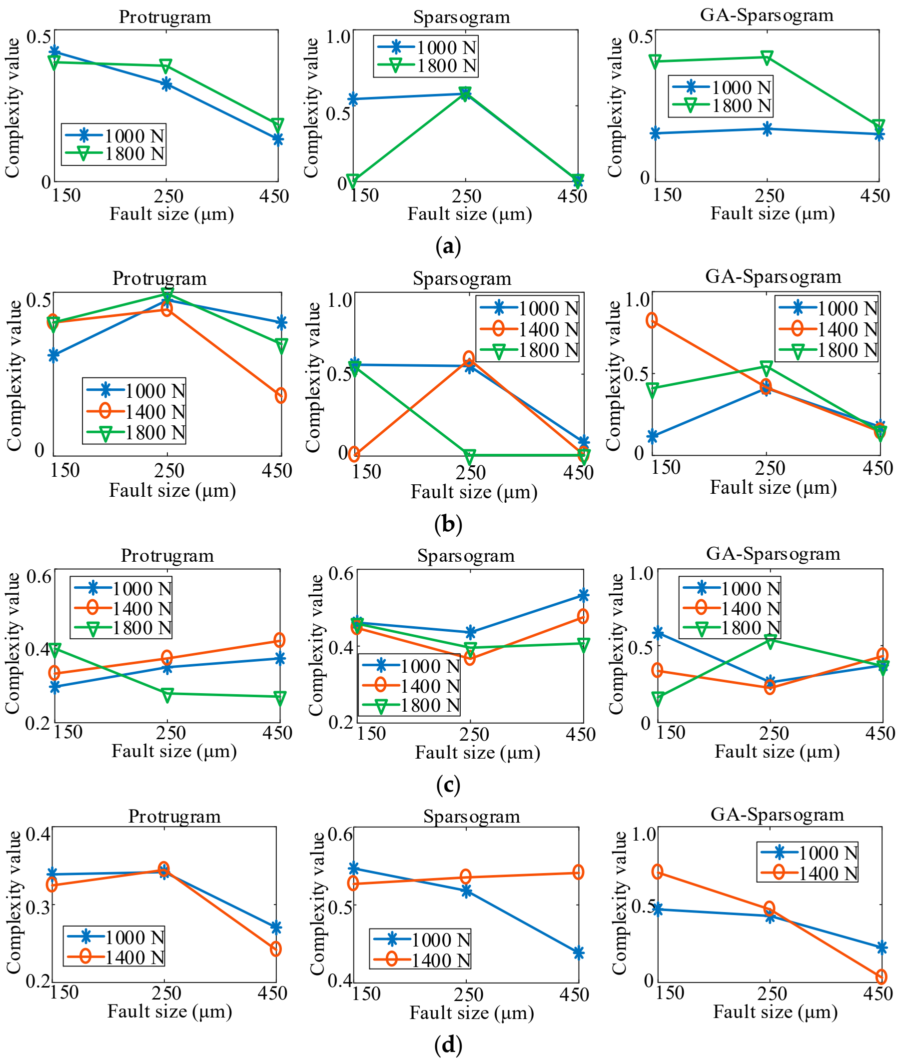


Figure 9. The complexity values of Dynamic and Identification Research Group datasets after different treatment. (**a**) 100 Hz; (**b**) 200 Hz; (**c**) 300 Hz; and (**d**) 400 Hz.

5. Conclusions

This paper proposed an improved method for fault severity recognition. This paper, for the first time, proposed a noise reduction method based on ACLHC for processing bearing vibration signals. The main finding was that, compared with the frequency band selection methods, the ACLHC did not need to select the frequency band of the signal in the process of reducing the influence of noise and highlighting the impact components of signal. Therefore, the signals processed by ACLHC could not only reduce the influence of noise, but also contained all the fault information. Then, the modulation information of the processed vibration signal, which is an essential fault characteristic of bearing, was reflected by SAX. Finally, the complexity of the symbol sequence coded by SAX was calculated by the LZ indicator. Through the verification of the single-point fault dataset of KAt and that of DIRG, it was proven that the ACLHC can reduce the influence of noise and highlight the impact components of signal without losing the fault information. Also, the signals processed by ACLHC can effectively improve the recognition results of bearing fault severity. In conclusion, the proposed method can effectively improve the recognition results of fault severity by reducing noise while retaining all frequency band information, and can be used in different working conditions.

Although some important problems associated with the proposed method have been investigated in this paper, there are still a few questions that are worthy of further consideration, such as the application scope of ACLHC.

Author Contributions: Conceptualization, J.Y.; Formal analysis, J.Y.; Investigation, J.Y., H.Z. and Y.Y.; Methodology, J.Y. and M.L.; Project administration, M.X.; Software, J.Y. and Y.Y.; Supervision, M.X.; Validation, H.Z.; Writing—original draft, J.Y.; Writing—review and editing, J.Y., H.Z., Y.L. and M.X.

Funding: This research was funded by National Natural Science Foundation of China, grant number 51805434 and 61203151.

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

References

1. Mcfadden, P.D.; Smith, J.D. Model for the vibration produced by a single point defect in a rolling element bearing. *J. Sound Vib.* **1984**, *96*, 69–82. [CrossRef]
2. Du, Y.; Chen, Y.; Meng, G.; Ding, J.; Xiao, Y. Fault Severity Monitoring of Rolling Bearings Based on Texture Feature Extraction of Sparse Time–Frequency Images. *Appl. Sci. Basel.* **2018**, *8*, 1538. [CrossRef]
3. Chen, Y.; Zhang, T.; Luo, Z.; Sun, K. A Novel Rolling Bearing Fault Diagnosis and Severity Analysis Method. *Appl. Sci. Basel* **2019**, *9*, 2356. [CrossRef]
4. Yan, R.; Gao, R.X. Complexity as a Measure for Machine Health Evaluation. *IEEE Trans. Instrum. Meas.* **2004**, *53*, 1327–1334. [CrossRef]
5. Hong, H.; Liang, M. Fault severity assessment for rolling element bearings using the Lempel-Ziv complexity and continuous wavelet transform. *J. Sound Vib.* **2009**, *320*, 452–468. [CrossRef]
6. Dou, D.; Zhao, Y. Fault severity assessment for rolling element bearings based on EMD and Lempel-Ziv index. *J. Vib. Shock* **2010**, *29*, 5–8. [CrossRef]
7. Zhang, C.; Chen, J. Fault severity assessment for rolling element bearings based on LMD and Lempel-Ziv index. *J. Vib. Shock* **2012**, *31*, 77–82. [CrossRef]
8. He, L.; Tan, J.; Yin, F.; Ding, C. Radial Wear Degree Recognition of Bearing based on LMD and Lempel-Ziv Index. *J. Mech. Tran.* **2014**, *38*, 34–38.
9. Zhang, X.; Zhang, Q.; Qin, X.; Sun, Y. Rolling bearing fault diagnosis based on ITD Lempel-Ziv complexity and PSO-SVM. *J. Vib. Shock* **2016**, *35*, 102–107. [CrossRef]
10. Cui, L.; Gong, X.; Zhang, J.; Wang, H. Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel-Ziv complexity. *J. Sound Vib.* **2016**, *385*, 372–388. [CrossRef]
11. Cui, L.; Li, B.; Ma, J.; Jin, Z. Quantitative trend fault diagnosis of a rolling bearing based on Sparsogram and Lempel-Ziv. *Measurement* **2018**, *128*, 410–418. [CrossRef]
12. Wang, D.; Miao, Q.; Fan, X.; Huang, H.Z. Rolling element bearing fault detection using an improved combination of Hilbert and wavelet transforms. *J. Mech. Sci. Technol.* **2009**, *23*, 3292–3301. [CrossRef]
13. Tian, J.; Morillo, C.; Azarian, M.H.; Pecht, M. Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis. *IEEE Trans. Ind. Electron.* **2016**, *63*, 1793–1803. [CrossRef]
14. Saidi, L.; Ali, J.B.; Bechhoefer, E.; Benbouzid, M. Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR. *Appl. Acoust.* **2017**, *120*, 1–8. [CrossRef]
15. Antoni, J.; Randall, R.B. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. *Mech. Syst. Signal Process.* **2006**, *20*, 308–331. [CrossRef]
16. Yi, W.; Tse, P.W.; Tang, B.; Yi, Q.; Lei, D.; Tao, H. Kurtogram manifold learning and its application to rolling bearing weak signal detection. *Measurement* **2018**, *127*, 533–545. [CrossRef]
17. Wang, T.; Chu, F.; Feng, Z. Meshing frequency modulation (MFM) index-based kurtogram for planet bearing fault detection. *J. Sound Vib.* **2018**, *432*, 437–453. [CrossRef]
18. Barszcz, T.; Jabłoński, A. A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. *Mech. Syst. Signal Process.* **2011**, *25*, 431–451. [CrossRef]
19. Tse, P.W.; Dong, W. The design of a new sparsogram for fast bearing fault diagnosis. *Mech. Syst. Signal Process.* **2013**, *40*, 499–519. [CrossRef]
20. Miao, Y.; Zhao, M.; Lin, J. Periodicity-impulsiveness spectrum based on singular value negentropy and its application for identification of optimal frequency band. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3127–3138. [CrossRef]
21. Huffman, D.A. A Method for the Construction of Minimum-Redundancy Codes. In *Proceedings of Institute of Radio Engineers*; IEEE: Piscataway, NJ, USA, 1951; pp. 1098–1101. [CrossRef]

22. Petrini, A.C.; Ionescu, V.M. Study of Huffman Coding Performance in Linux and Windows 10 IoT for Different Frameworks. In Proceedings of the International Conference on European Transnational Education International Workshop on Soft Computing Models in Industrial and Environmental Applications Computational Intelligence in Security for Information Systems Conference, San Sebastian, Spain, 19–21 October 2016; pp. 766–775. [CrossRef]
23. Larmore, L.L.; Hirschberg, D.S. A Fast Algorithm for Optimal Length-Limited Huffman Codes. *J. ACM* **1990**, *37*, 464–473. [CrossRef]
24. Wei, W.; Wei, Z. Huffman Coding Based Adaptive Spatial Modulation. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 5090–5101. [CrossRef]
25. Lempel, A.; Ziv, J. On the Complexity of Finite Sequences. *IEEE Trans. Inform. Theory* **1976**, *22*, 75–81. [CrossRef]
26. Zhang, X.; Roy, R.J.; Jensen, E.W. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans. Biomed. Eng.* **2001**, *48*, 1424–1433. [CrossRef]
27. Yin, J.; Xu, M.; Zheng, H. Fault diagnosis of bearing based on Symbolic Aggregate approXimation and Lempel-Ziv. *Measurement* **2019**, *138*, 206–216. [CrossRef]
28. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In Proceedings of the European Conference of the Prognostics and Health Management Society, Bilbao, Spain, 5–8 July 2016; pp. 1–17.
29. Daga, A.P.; Fasana, A.; Marchesiello, S.; Garibaldi, L. The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data. *Mech. Syst. Signal Process.* **2019**, *120*, 252–273. [CrossRef]