

Article

# A Data-Driven Approach for the Diagnosis of Mechanical Systems Using Trained Subtracted Signal Spectrograms

Jiung Huh <sup>1</sup>, Huan Pham Van <sup>1</sup>, Soonyoung Han <sup>1</sup>, Hae-Jin Choi <sup>1,\*</sup> and Seung-Kyum Choi <sup>2,\*</sup>

<sup>1</sup> School of Mechanical Engineering, Chung-Ang University, Seoul 06974, Korea; puraria@cau.ac.kr (J.H.); pvh134@cau.ac.kr (H.P.V.); hsy4190@cau.ac.kr (S.H.)

<sup>2</sup> The George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

\* Correspondence: hjchoi@cau.ac.kr (H.-J.C.); schoi@me.gatech.edu (S.-K.C.)

Received: 30 January 2019; Accepted: 25 February 2019; Published: 1 March 2019



**Abstract:** Toward the prognostic and health management of mechanical systems, we propose and validate a novel effective, data-driven fault diagnosis method. In this method, we develop a trained subtracted spectrogram, the so called critical information map (CIM), identifying the difference between the signal spectrograms of normal and abnormal status. We believe this diagnosis process may be implemented in an autonomous manner so that an engineer employs it without expert knowledge in signal processing or mechanical analyses. Firstly, the CIM method applies sequential and autonomous procedures of time-synchronization, time frequency conversion, and spectral subtraction on raw signal. Secondly, the subtracted spectrogram is then trained to be a CIM for a specific mechanical system failure by finding out the optimal parameters and abstracted information of the spectrogram. Finally, the status of a system health can be monitored accurately by comparing the CIM with an acquired signal map in an automated and timely manner. The effectiveness of the proposed method is successfully validated by employing a diagnosis problem of six-degree-of-freedom industrial robot, which is the diagnosis of a non-stationary system with a small amount of training datasets.

**Keywords:** prognostics and health management (PHM); industrial robot; critical information map (CIM); data-driven; non-stationary signal; smart factory; wavelet package decomposition (WPD)

## 1. Introduction

Uncertainty in engineering has been considered as a critical problem as it could result in serious financial losses or catastrophic accidents. Specifically, if a machine in a production line suddenly malfunctions, the line will stop, which leads to significant financial losses. Therefore, it is critical to predict malfunctions or the life of manufacturing machines. This requirement has resulted in active research regarding the prognostics and health management of machinery with the benefit of the fourth industrial revolution from the significant progress in data science, computer performance, and communication.

For the health monitoring or diagnosis of a machine, we need sensors to measure the status of the machine, including sensors to measure temperature [1], pressure [2], volume pressure [3,4], and acceleration [5–7]. Among these sensors, acoustic emission sensors and accelerometers are the most commonly used for monitoring machines as they can provide instantaneously the status of a machine with high data sampling frequencies, which is not possible with temperature or pressure sensors. In other words, we can obtain large amounts of information over a relatively short measurement time. Accelerometer signals that measure structural vibration are especially useful in monitoring rotary

machines, such as gear sets and bearings, as their frequencies are closely related to the frequency of the mechanical mesh and the machine structure [8,9]. In this study, we investigate the techniques for monitoring machinery with the acceleration signals.

The approaches for the fault diagnosis are categorized into model-based, signal-based, knowledge-based, hybrid-based [10], and active approaches [11,12]. Among the five approaches, model-based, signal-based and knowledge-based approaches are mostly applied in the machine fault diagnosis. Model-based methods employ specific dynamic models or theories which simulate the real condition. Luo [13] proposed a shape-independent method to model different kinds of tooth spalls and validated by finite element analyses. Endo [14] diagnosed spalls and cracks on the gear tooth based on simulated signals. Park [15] diagnosed gear faults in planetary gears using transmission errors simulated from a dynamic model.

Signal-based methods use the aforementioned signals with prior knowledge. Lebold [16] reported a signal processing procedure to extract features from the vibration signal for monitoring a gear set. The first step in this method was extracting statistical parameters, such as root mean square, kurtosis, and skewness, from raw vibrational signal that could include important information regarding mechanical faults. The second step was extracting features from the cyclic signal from which noise is removed by the time synchronous averaging (TSA) technique [8]. Additional signal features could be obtained from the residual signal remaining after removing the gear mesh frequency, the difference signal obtained by further removing the sideband of the mesh frequency, or the band-pass mesh signal, which is the band frequency signal extracted from the TSA signal. He et al. [17] employed bearing characteristic frequencies as the representative features for the analysis of bearing faults. The bearing characteristic frequencies are the metrics of frequency densities related to a specific part of a bearing. Statistical parameters, such as kurtosis, are also used to measure degradation and the degree of bearing defects.

The model-based and signal-based methods are commonly employed for the diagnosis of gear sets and bearings. However, these approaches are available only when the signal is measured in a stationary status. It is also virtually impossible for engineers to apply the approaches to a system of which structure is complex and dynamic characteristics, such as operational frequency of the gear sets and bearings, is not sufficient. In order to overcome these limitations and meet the demands of diagnoses of various system types, knowledge-based methods, also called data-driven methods, have been proposed. A considerable number of research works [18–21] based on the approach have been carried out recently. In the knowledge-based methods, professional information or knowledge of the systems are not necessary. Instead, a machine-learning algorithm recognizes patterns of each class and diagnoses the systems by themselves. However, it requires a large amount of historical data to find important information for the fault diagnoses [22]. In order to overcome the limitations of the three different methods, we propose a new fault diagnosis method called a critical information map (CIM), that works with (1) relatively small amounts of data, (2) non-stationary signals, and (3) unknown system structures and dynamics as discussed in this study.

The details of our idea are presented in the following sections: Section 2 explains pre-processing processes, which apply to the raw signals, include data synchronization, time frequency representation (TFR) transformation, and spectral subtraction. Section 3 includes an optimization process that addresses the important parameters to achieve critical information map. A case study of data acquiring from manipulator is introduced and the CIM method is validated in Section 4. Another case study based on the data from the National Aeronautics and Space Administration (NASA) repository has been employed to validate the usefulness of the CIM, comparing performance with other methods in Section 5. Finally, in Section 6, we discuss the validation results and the usefulness of this technique in a prognostics and health management (PHM) problem.

## 2. Critical Information Map (CIM): Pre-Processing

In this study, we propose a TFR-based CIM that includes the information of the locations within the TFR spectrograms that shows clear (or large) difference between normal and fault conditions. An engineer can diagnose the system fault by identifying differences of the parameters at the specific regions of the CIM of a system of interest. Using this approach, we believe that detailed structural or dynamic knowledge of a system is not necessary since we employ TFR-based spectrograms. In addition, the proposed approach is applicable to the non-stationary signal-based diagnoses as we preserve and make use of the domain of the TFRs. Finally, this approach requires relatively small amounts of training data compared to that required by the TFR and convolutional neural network (CNN) approaches.

The overall procedure of creating a CIM with raw measured vibrational signals is illustrated in Figure 1. It comprises three pre-processes: data synchronization that is discussed in Section 2.1, wavelet packet transformation discussed in Section 2.2, and spectral subtraction explained in Section 2.3. Without the pre-processing steps, the following step (i.e., optimization process) cannot be converged. The final optimization process for creating CIMs is discussed in Section 3.

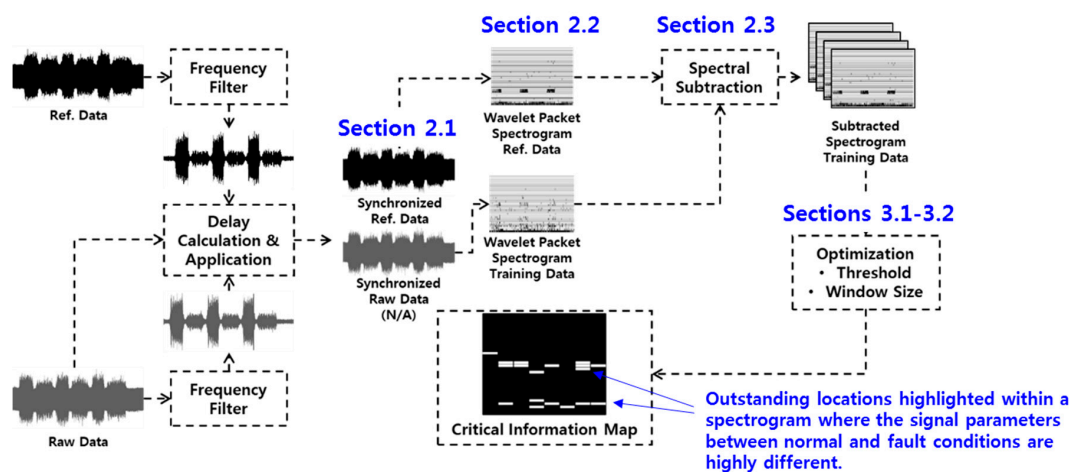


Figure 1. The process of building the critical information map (CIM).

### 2.1. Data Synchronization

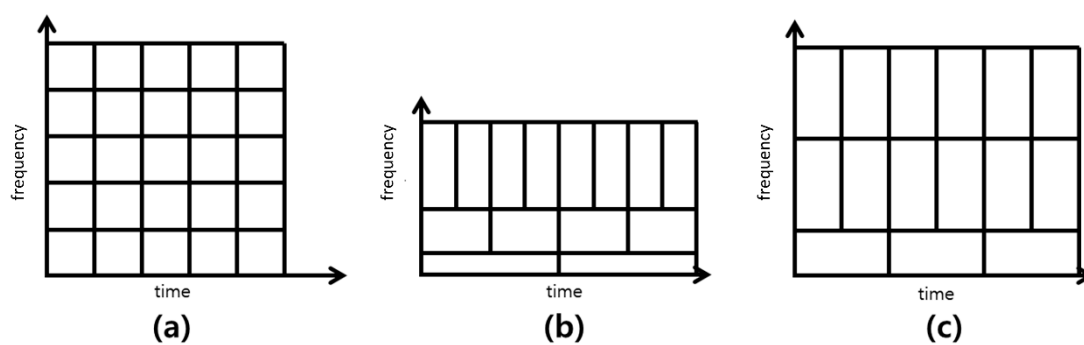
Measured data typically includes a time delay that needs to be eliminated by synchronization. In this study, we employ cross-correlation for the time synchronization. Before the synchronization, we first extract non-stationary movement signals of interest using a band-pass filter as other unrelated signals become noise and hinder the time synchronization.

Cross-correlation is typically used to synchronize two different time-series signals. The conformity degree between the two signals is defined as  $(f \times g)(\tau) = \int_{-\infty}^{\infty} f(t)g(t + \tau)dt$  where  $f$  is the original function,  $g$  is the filtering function, and  $\tau$  is the time delay between the two signals [23]. We determine the time delay ( $\tau_{max}$ ) that maximizes the cross-correlation,  $(f \times g)(\tau)$ , which is  $\tau_{max} = Argmax\{(f \times g)(\tau)\}$ . The raw signal is then synchronized with the original signal for comparison by applying the  $\tau_{max}$ .

### 2.2. Time Frequency Representation (TFR) Transformation

For developing CIM, we convert raw signals to TFRs (i.e., two-dimensional data plots) for the effective representation of the signal information for humans as well as computers. Short-time Fourier transform (STFT), wavelet transform (WT), and wavelet packet decomposition (WPD) are the linear methods commonly used for the transformation [24]. The primary differences between the three methods are the form of the filter and the shapes of the spectrum tile (i.e., window) delivered from the filter.

In the STFT method, the size of the windows is pre-determined and identical for all frequency and time domains, as shown in Figure 2a. In addition, one frequency density value is assigned for each window. The pre-determination of the window size can, at times, increase the degree of uncertainty in the high- and low-frequency domains of the transformed diagrams [25]. Verstraete [26] reported that CNN models that were trained based on the two-dimensional image data transformed using the STFT method exhibited low algorithm reliability. The WT method overcomes this limitation by varying the size of the window of each time and frequency domain, as shown in Figure 2b. However, in WT method, the resolution of the time and frequency domains are determined by the number of windows, which makes the window sizes uneven in the domain. Additionally, WT decomposes only the low frequency component at subsequent levels whereas WPD decomposes both low- and high-frequency components at each level. WT is not desirable for creating a CIM for fault identification where signal contains high frequency information. The WPD method addresses the limitations of the two methods.



**Figure 2.** Window shapes: (a) short-time Fourier transform (STFT); (b) wavelet transform (WT); (c) wavelet packet decomposition (WPD).

As can be seen in Figure 2c, the WPD method assigns a different filter at each divided domain, which provides a high degree of freedom in setting the size of tiles. Compared to the STFT method that uses only one filter, we can obtain more reliable frequency density values with the WPD method. In this study, we employ WPD as the TFR method for creating the CIM. In WPD, the kernel function of the wavelet packet to decompose the signal to several frequencies is:

$$W_{j,k}^n(t) = 2^{\frac{j}{2}} W^n(2^j t - k), \quad (1)$$

Equation (1) includes three positive integer constants:  $j$  is the index scale,  $k$  is the translation operation, and  $n$  is the modulation parameter or oscillation parameter. The first and second wavelet packet functions need to be predefined, as shown in Equations (2) and (3), which are known as the usual scaling function and mother function, respectively.

$$W_{0,0}^0(t) = \phi(t), \quad (2)$$

$$W_{0,0}^1(t) = \psi(t), \quad (3)$$

additional functions may be created as follows:

$$W_{0,0}^{2n}(t) = \sqrt{2} \sum_k h(k) W_{1,k}^n(2t - k), \quad (4)$$

$$W_{0,0}^{2n+1}(t) = \sqrt{2} \sum_k g(k) W_{1,k}^n(2t - k), \quad (5)$$

where  $h(k)$  and  $g(k)$  are the quadrature mirror filters, which are orthogonal to each other. The wavelet packet coefficient in a  $j, n, k$  state that represents the density of a filter is:

$$\omega_{j,n,k}(t) = \langle f, W_{j,k}^n \rangle = \int f(t) W_{j,k}^n(t) dt, \quad (6)$$

where  $f(t)$  is the time signal that will be analyzed. The values of the wavelet packet coefficients that will be used for developing the CIMs are displayed in the WPD spectrogram in Figure 2c. The details of the WPD are beyond the scope of this study, and can be found in [27]

### 2.3. Spectral Subtraction

In this study, we use the spectral subtraction technique to determine the critical area on the WPD by which we can screen fault conditions from the samples. The spectral subtraction technique has been employed by acoustics engineers to filter out unwanted noise. The basic principle is to extract signals in the domain where speech signals do not exist, and then using the extracted signal to enhance the voices.

In the spectral subtraction, the modified signal spectrum is as follows:

$$P'_s(w) = \begin{cases} P_s(w) - P_n(w), & \text{if } P_s(w) - P_n(w) > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (7)$$

where  $P_s(w)$  is the spectrum with speech and noise, and  $P_n(w)$  is the noise spectrum without speech. Berouti et al. [28] obtained speech signals without noise by the Fourier inversion of  $P'_s(w)$ . Denda [29] employed spectral subtraction based on a wavelet transformed spectrum, which is represented as follows:

$$|\hat{X}(b, a)| = |Y_s(b, a)| - \alpha \left| \overline{N_n(b, a)} \right|, \quad (8)$$

where  $|\hat{X}(b, a)|$  is the wavelet spectrum of enhanced speech,  $|Y_s(b, a)|$  is the wavelet spectrum of the observed signal, and  $\left| \overline{N_n(b, a)} \right|$  the wavelet spectrum of noise. In addition,  $\alpha$  is the reduction factor, and  $a$  and  $b$  are the filter location and scale parameters in the wavelet transformation spectrum, respectively. El Bouchikhi [30] assumed the spectrum domain without speech as the signal from normal bearings and the spectrum domain with speech as the signal from faulty bearings. After the spectral subtraction, El Bouchikhi obtained improved diagnosis results.

In this study, we performed the spectral subtraction in Equation (8) based on the spectrogram obtained by WPD to identify the critical region of the spectrogram for fault identification. We subtracted the signal spectrum of the normal condition from that of the abnormal condition, which indicates that specific features of the normal condition data are removed. In this study, normal-reference (N-R) indicates the spectrogram obtained by subtracting the reference signal from the normal condition signal. A normal dataset is randomly selected as the reference signal for this spectral subtraction to fix a ground value of the spectral coefficient. Abnormal-reference (A-R) is then the spectrogram derived by subtracting the reference signal from the abnormal one. We mutually subtract between the spectrums of N-R and A-R to obtain the degree of random error, such as measurement error. We discuss the technique to deal with the random error for robust decision-making in diagnosis in Section 3.

## 3. Critical Information Map: Optimization

### 3.1. CIM Optimization (Training) Parameters

For developing the CIM, we need to identify the parameters that determine the location and size of important regions for each fault condition. Figure 3 is an example of a spectrogram (i.e., map) obtained after the spectral subtraction process discussed in Section 2.3. The resolution of the map may be determined by setting the number of windows (i.e., tiles) in the time domain, but the resolution in the frequency domain is fixed because the variation of the frequency domains increases the non-linearity by changing the number of samples in the time domain. A number of density values may exist in each window. There could be specific windows whose density values for normal and fault conditions

are significantly different. These windows may be beneficial for identifying a fault condition. If we know the location of these windows, we can compare the values of those specific windows only. This could reduce the noise for fault diagnosis. Therefore, we need to create an optimal diagnosis map that includes the location and the threshold value of important windows for each fault condition.

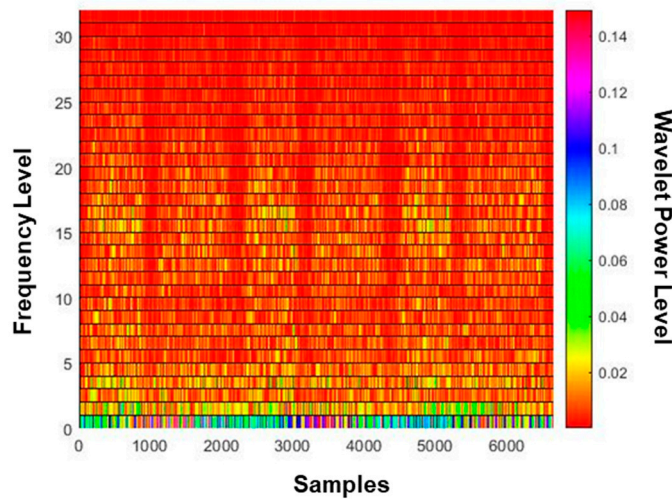


Figure 3. Example of subtracted spectrogram.

In order to formulate an optimization problem for creating the CIM, we first define three design variables to determine. The first variable ( $x_1$ ) is the number of windows in the time domain that also determines the size of windows (or resolution). The second variable ( $x_2$ ) is the threshold value to be an outstanding parameter in a window. If a density parameter is greater than  $x_2$ , it is then an outstanding parameter. The last variable ( $x_3$ ) is the threshold of the number of outstanding parameters in a window. If the number of outstanding parameters in a window is greater than  $x_3$ , then the window becomes an important window to be observed for diagnosis.

A simple example to better understanding the definition of the variables,  $x_2$  and  $x_3$ , is shown in Figure 4. The window in this example comprises 10 parameters (i.e., cells) with the assigned values as shown in the figure. If  $x_2 = 2$ , then the outstanding parameters are the parameters of which values are greater than or equal to 2. In this case, the number of outstanding parameters in the window is 6. If  $x_2 = 3$ , then the number of outstanding parameters in the window becomes 3. If  $x_2 = 3$  and  $x_3 = 2$ , then the window is outstanding. However, if  $x_2 = 2$  and  $x_3 = 5$ , then the window is not outstanding.

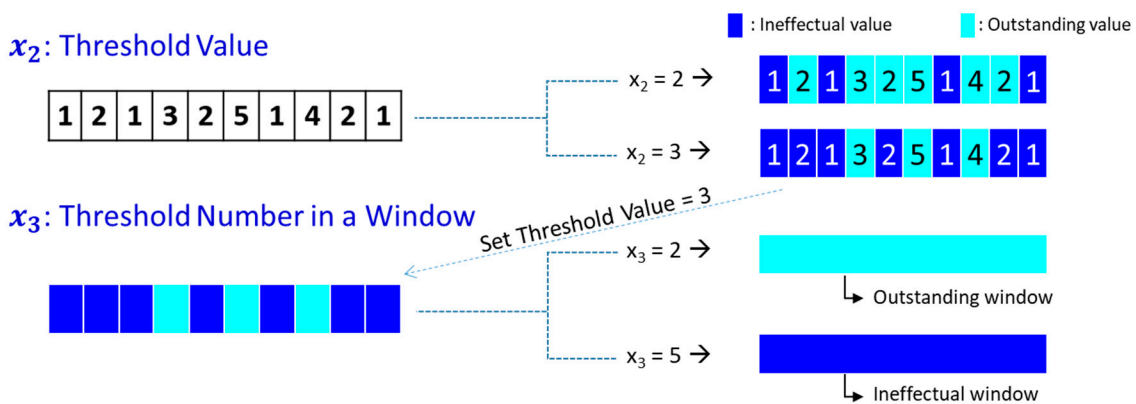


Figure 4. Example of design variables:  $x_2$  = threshold value for parameter to be outstanding, and  $x_3$  = threshold number of outstanding parameters for a window to be outstanding.

### 3.2. Optimization Problem Formulation

Figure 5 shows the optimization problem formulation [31] to build the CIM. Our goal in this problem is to minimize the loss of important information. Therefore, the objective function is to maximize the normalized outstanding area (i.e., important regions), which is the number of outstanding windows normalized by the division number in the time domain. As mentioned in Section 2.3, it is important to identify the area where N-R is dominant, which will result in creating a map highly sensitive to random noise (or measurement error). For this reason, we employed the outstanding degree of a window as  $R_c = N_{a-r,ij}/N_a - C_p N_{n-r,ij}/N_n$  where  $C_p N_{a-r,ij}/N_a$  is the term that represents the disturbing errors. It is recommended that the  $R_c$  of a window be close to one (i.e., greater than 0.95 or 0.99) to be selected as an outstanding window.

**Given**

- $N_n$  = the number of normal data sets
- $N_a$  = the number of abnormal data sets
- $D_f$  = the number of divisions in frequency domain
- $C_p$  = penalty weight for N-R subtraction
- $R_c$  = outstanding degree of a window ( $-C_p \leq R_c \leq 1$ )
- $L_S$  = length of synchronized data
- $N_{n-r,ij}$  = the number of normal data sets that make the window at  $i$ -th row and  $j$ -th column outstanding in N-R spectrogram
- $N_{a-r,ij}$  = the number of normal data sets that make the window at  $i$ -th row and  $j$ -th column outstanding in A-R spectrogram

**Find**

- $x_1$  = the number of divisions in time domain
- $x_2$  = threshold value for a parameter to be outstanding
- $x_3$  = threshold number of the outstanding parameters for a window to be outstanding

**Satisfy**

- $5 \leq x_1 \leq \frac{L_S}{D_f} (x_1 \in \mathbb{N})$
- $0 < x_2 \leq 1$  (usually 0.1 upper limit)
- $1 \leq x_3 \leq \frac{L_S}{x_1 D_f} (x_3 \in \mathbb{N})$

**Minimize**

$z = -\sum_{i=1} \sum_{j=1} A_{ij} / x_1$  (maximize the normalized outstanding area )

where  $A_{ij} = \begin{cases} 1 \text{ (outstanding)}, & \text{if } R_c = \left( \frac{N_{a-r,ij}}{N_a} - C_p \frac{N_{n-r,ij}}{N_n} \right) \geq 0.95 \text{ (or } 0.99) \\ 0, & \text{otherwise} \end{cases}$

**Figure 5.** Problem formulation of CIM optimization.

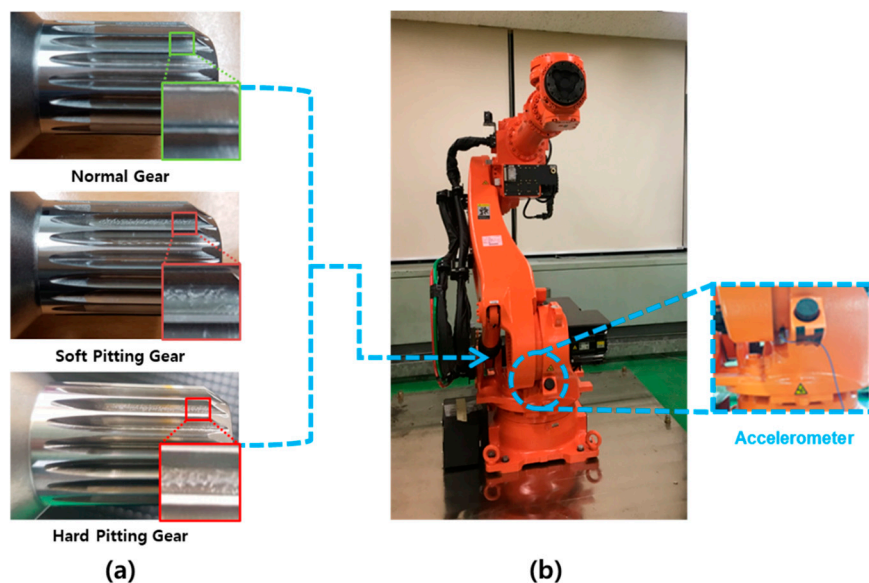
When deciding on the status of a window, if the number of N-R datasets (i.e.,  $N_{n-r,ij}$ ) that make the window outstanding is significantly greater than the number of A-R datasets (i.e.,  $N_{a-r,ij}$ ) that make the window outstanding, then the window can be set as an outstanding window. We can obtain

the normalized outstanding area by the number of outstanding windows divided by the number of divisions in the time domain,  $x_1$ . By achieving the optimal solution of this problem, we can obtain the optimal resolution of the CIM as well as the list of outstanding windows. For the diagnosis of a new observation based on the CIM, we merely need to calculate the number of outstanding windows from the observed signal and compare it with the outstanding window list of the CIM.

#### 4. Case Study 1: Diagnosis of Input Gear Problems of Six-Degree-of-Freedom Manipulator

##### 4.1. Data Acquisition Process

We apply the CIM-based diagnosis method proposed in Sections 2 and 3 to the diagnosis of a six-degree-of-freedom manipulator. This industrial robust system is a suitable example for validating the proposed method as it comprises numerous components, including arms, motors, belts, bearings, and gear sets. The signal observed from this system is, therefore, typically complex and non-stationary. The manipulator employed in this case study is the HS-180S, manufactured by Hyundai Robotics Inc in Daegu, Korea. (<https://www.hyundai-robotics.com/>), as shown in Figure 6b.



**Figure 6.** (a) Three different types of input gear at second joint, and (b) HS-180 manipulator and position of accelerometer.

The analysis target is the second joint of the manipulator shown in Figure 6b. As can be seen in the figure, an accelerometer is attached on top of the joint housing to measure the radial vibration [32]. In this case study, we assumed that the input gear of the second joint could have pitting defects during the operation, as shown in Figure 6a. The top photo is the normal input gear, the middle one is the gear with soft pitting, and the bottom one is the gear with hard pitting. The pitting defects were intentionally created with a grinder for the purposes of testing. Soft and hard pitting defects depend on the amount of scratching on the surface of the gear tooth.

The motion of the manipulator for acquiring signal data is rotating the upper and lower rams three times in  $60^\circ$  around the section joint, as shown in Figure 7. This action was repeated 300 times to collect 900 sets of accelerometer data (300 normal dataset, 300 hard-pitting dataset, 300 soft-pitting dataset), which are sufficient for developing and validating our approach. In other words, we could have 300 normal datasets and 600 abnormal datasets, which include hard pitting and soft pitting faults. The time duration is 20 s for the three rotations, and the sampling rate is 12.8 kHz. The primary data acquisition specifications are presented in Table 1.



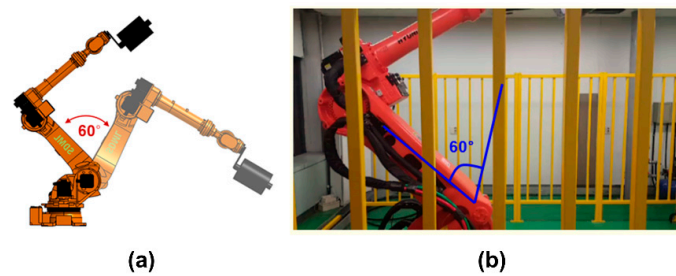


Figure 7. Data acquiring motion of robot shown in (a) a graphical image and (b) its photograph.

Table 1. Data acquisition specifications.

Features	Properties
Sensitivity of Sensor	10 mV/g
Sampling Frequency	12,800 sample/s
Acquired duration	20 s
Data Length	256,000 sample points
Normal data	300 datasets
Hard pitting data	300 datasets
Soft pitting data	300 datasets

#### 4.2. Data Synchronization

Figure 8a is the raw signal data acquired from the accelerometer with the normal input gear shown in Figure 6a. The signal with the larger amplitude and the shorter time duration is acquired during the downward movement, and the signal with the smaller amplitude and the longer time duration during the upward movement. In order to determine the frequency bandwidth useful for data synchronization, we draw a continuous spectrogram of the raw signal, as shown in Figure 8b, and extract the signal of the bandwidth in the rectangular area, which is the most distinct after the band-pass filter, as shown in Figure 8c. We determine the time delay from the filtered signals and apply it to the raw signal for time synchronization. An example of the data synchronization results is shown in Figure 9.

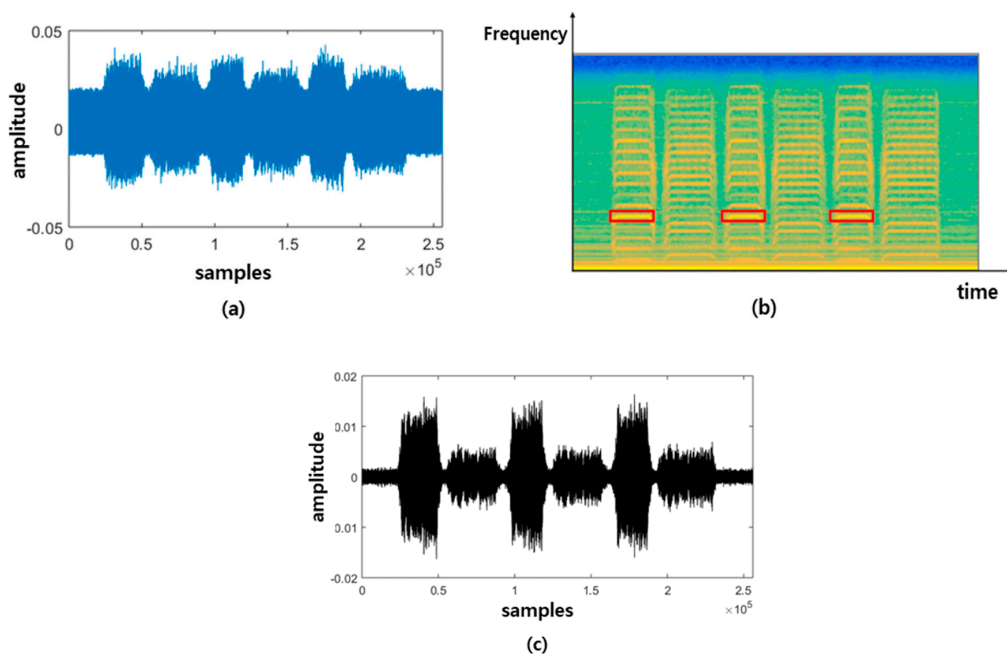


Figure 8. (a) Set of acquired raw signal data, (b) spectrogram of raw data, and (c) filtered data.

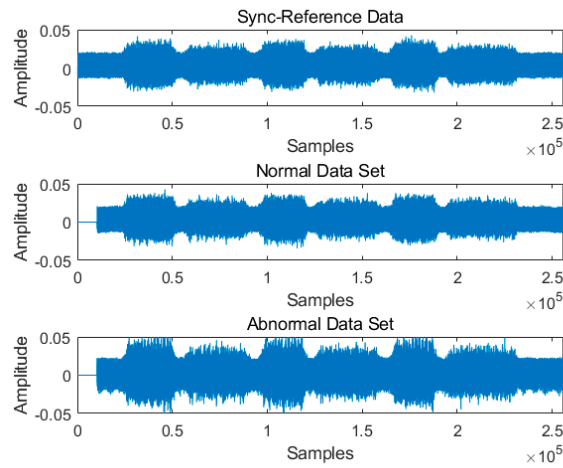


Figure 9. Time synchronization of raw signal data.

4.3. Wavelet Packet Decomposition (WPD) Spectrogram and Spectral Subtraction

The synchronized signals are transformed into spectrograms by WPD. Each image of the spectrograms includes 6656 parameters in the time domain (i.e., horizontal direction) and 32 parameters in the frequency domain (i.e., vertical direction). Figure 10a–c shows the wavelet packet spectrogram of normal signal, soft-pitting signal and hard-pitting signal, respectively. The distinction between these spectrograms are not clear. As shown in Figure 10b, the area with purple color bars in the N-R spectrogram indicate the regions of significant differences in the parameter values between the normal and reference spectrograms, where the reference spectrogram is an instance of normal spectrograms. Compared with Figure 10d,e, we can observe that the size of the purple area of the S-R spectrogram is greater than that of the N-R spectrogram. This indicates that the difference between the abnormal (soft pitting) and reference spectrograms is greater than that between the normal and reference spectrograms. To distinguish the abnormal signals from the normal signal, we need only focus on the purple area highlighted in Figure 10e and not in Figure 10d. A similar phenomenon might be observed when we compare the H-R spectrogram in Figure 10f with other subtracted spectrograms. It is a reasonable assumption that the purple area highlighted in Figure 10d is from the random noise as we randomly select a normal signal as the reference signal. This screening process is essential for making a robust decision insensitive to the random errors in measured signals for the diagnosis of abnormal conditions. This process will be performed in the CIM optimization process discussed in Section 4.4.

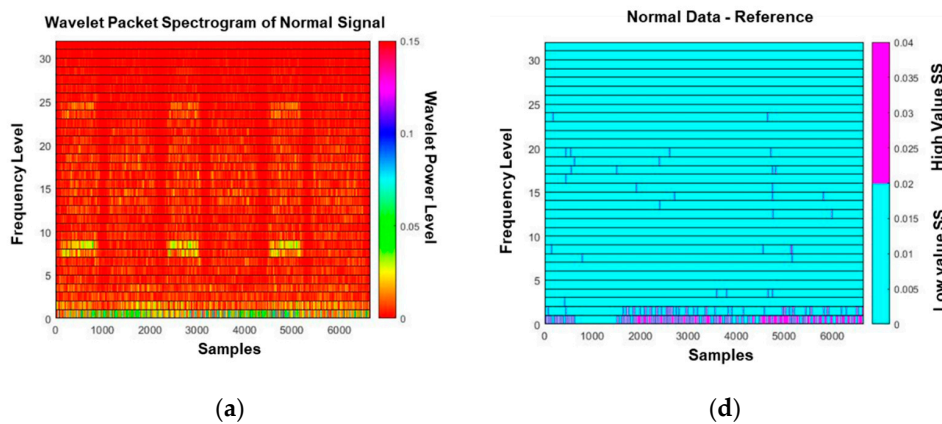
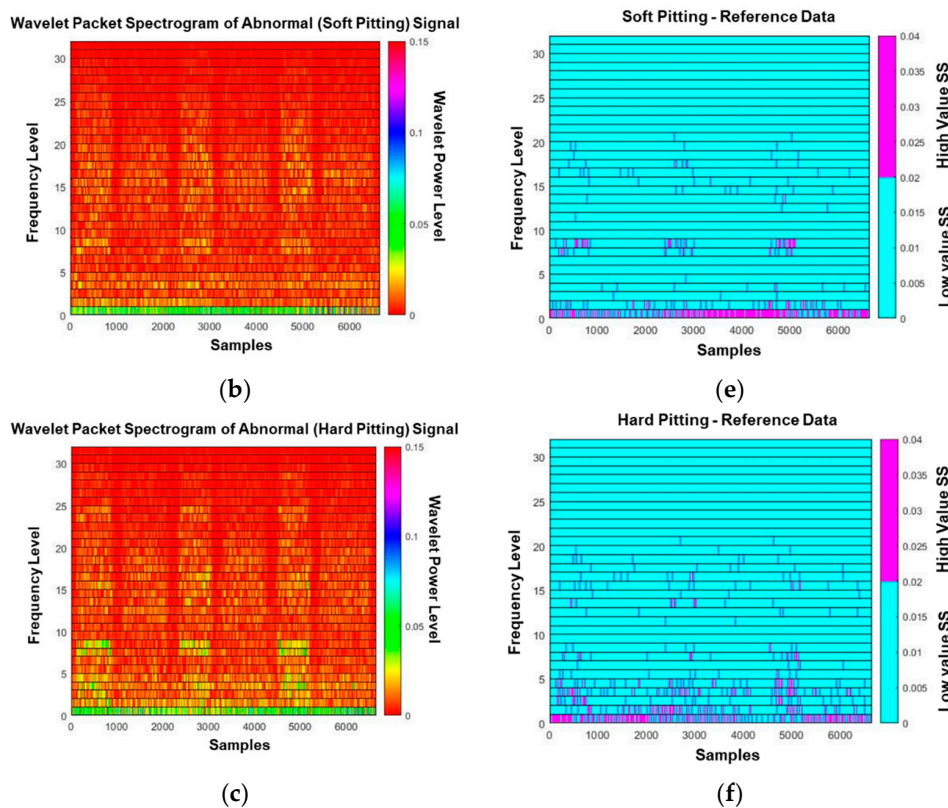


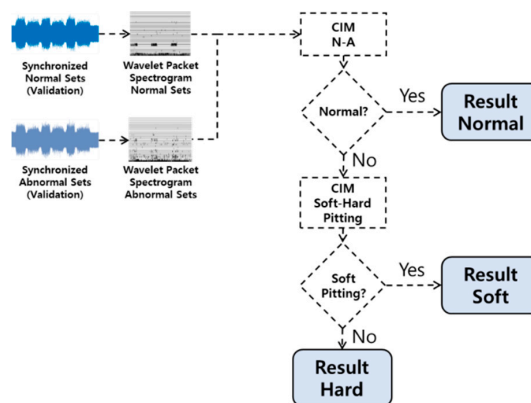
Figure 10. Cont.



**Figure 10.** Wavelet packet spectrograms of (a) a normal signal, (b) an abnormal (soft-pitting) signal, and (c) another abnormal (hard-pitting) signal; (d) N-R spectrogram, (e) A-R (S-R) spectrogram, (f) A-R (H-R) spectrogram.

#### 4.4. Optimization for Creating CIMs

In this case study, we classify the defect conditions (i.e., soft and hard pitting) of the gear, as shown in Figure 11. We first classify an input gear into a normal or abnormal condition, and then classify the abnormal gears into soft pitting or hard pitting defects. Therefore, we develop two CIMs that are for classifying normal, or abnormal conditions with soft pitting or hard pitting. Of the 900 datasets measured in Section 4.1, we use only 5% of them for developing the two CIMs, which are 15 datasets of normal, soft pitting, and hard pitting. The rest of the datasets (i.e., 855 sets) are reserved for validation in Section 4.5. The datasets for developing the CIM were pre-processed following the steps of time synchronization, transformation into WPD spectrograms, and spectrogram subtraction, as described in Sections 4.2 and 4.3.



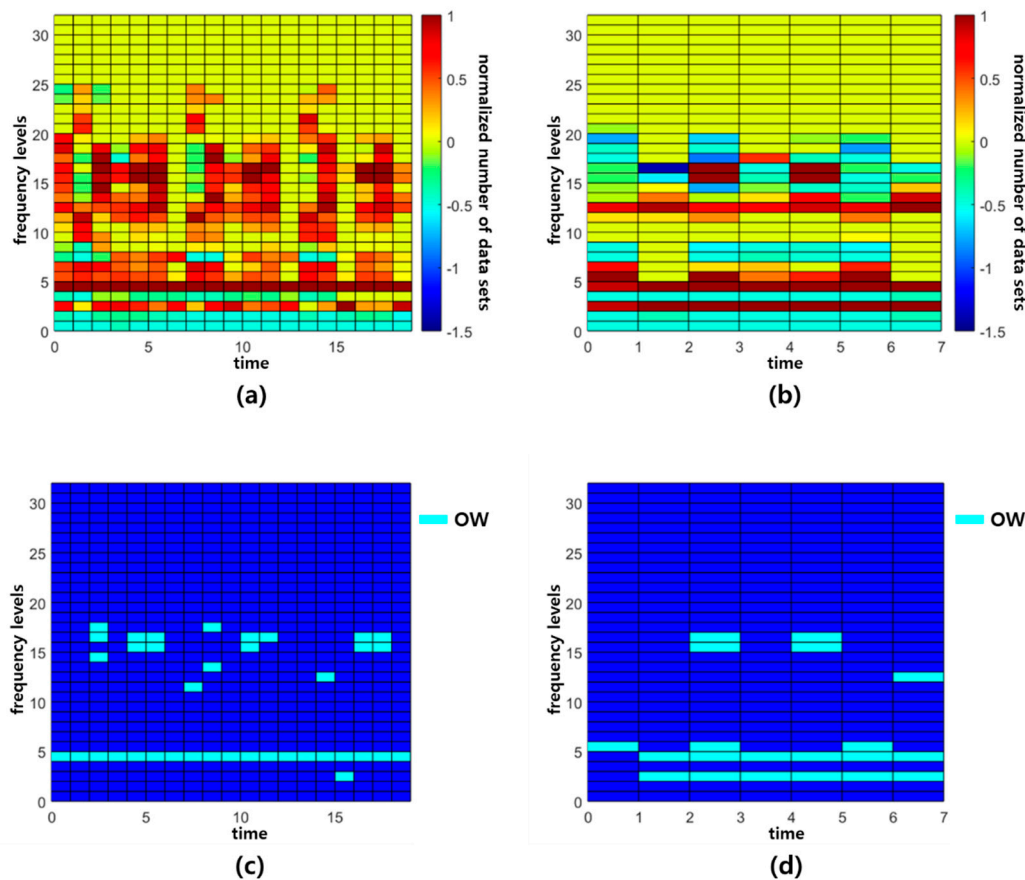
**Figure 11.** Flow chart of classifier.

For the CIM optimization shown in Figure 5, parameter  $C_p$  needs to be specified by users. As the values of those parameters increase, outstanding windows (OWs) are selected in a conservative manner. Typically,  $C_p$  is set as more than one for reducing the sensitivity random errors. In this formulation, we conservatively set  $C_p = 1.5$  and select the OWs of which  $R_c \geq 0.95$ . The optimal designs found for the two CIMs is presented in Table 2. Hard-reference (H-R) is the spectrogram obtained by subtracting the reference signal from the hard-pitting signal. S-R (soft-reference) stands for the spectrogram when we do the same with the soft-pitting signal.

**Table 2.** Achieved optimal CIMs.

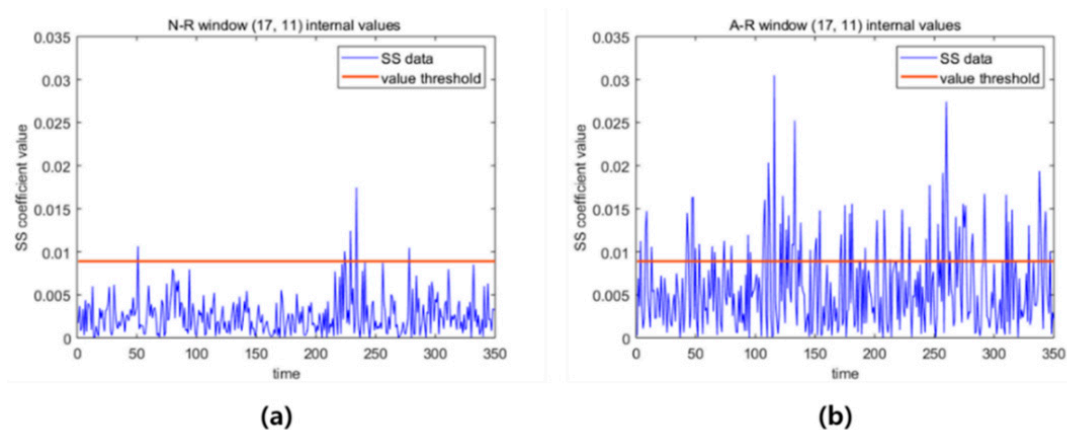
Design Variables of CIMs	CIM (N-R vs. A-R)	CIM (S-R vs. H-R)
$x_1$ : the number of divisions in time domain	19	7
$x_2$ : threshold value for a parameter to be outstanding	0.0089	0.0224
$x_3$ : the minimum number of outstanding parameters for a window to be outstanding	49	9

Figure 12a,b shows the two CIMs colored with the outstanding degree  $R_c$  of each window highlighted. If the color of a window is red (i.e., close to one), then the window could be an OW. The windows in the figures could include a large number of SS (spectral subtraction) parameters. In these CIMs, 350 and 951 SS parameters are included in each window of Figure 12a,b, respectively. Figure 12c,d shows the OWs (i.e., highlighted windows) of which  $R_c \geq 0.95$ . The number of OWs in the CIM for N-R vs. A-R shown in Figure 12c is 38 and that of the CIM for S-R vs. H-R shown in Figure 12d is 20.



**Figure 12.** (a) CIM for N-R vs. A-R, (b) CIM for S-R vs. H-R, (c) selected outstanding windows (OWs) ( $R_c \geq 0.95$ ) in CIM for N-R vs. A-R, and (d) selected OWs ( $R_c \geq 0.95$ ) of CIM for S-R vs. H-R.

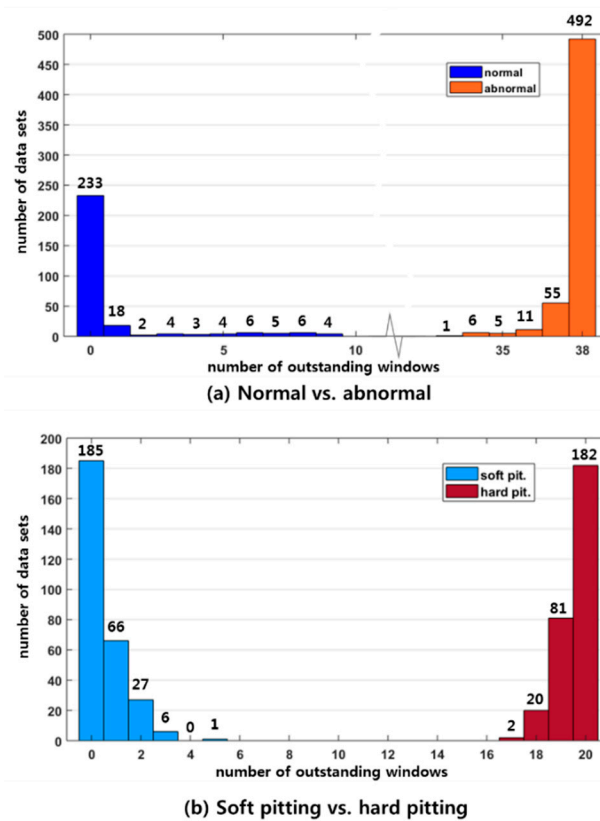
We focus on a specific window as an example to further elaborate details of the classification based on the CIM. Figure 13 shows the SS coefficients included in the window at the 17-th frequency level and 11-th time domain. Figure 13a shows the values of 350 coefficients in the window of an N-R subtracted spectrogram obtained from a normal signal, and Figure 13b shows the same values in the same window of an A-R subtracted spectrogram from an abnormal signal. In Figure 13a, only a few SS coefficients (fewer than 10 coefficients in this example) are above the threshold based on the optimized threshold value for a parameter to be outstanding (0.0089). However, approximately half of the SS coefficients in Figure 13b are above the threshold, which is the reason why the window is selected as OW. We may classify a new observed signal data by investigating the number of windows in the O-R subtracted spectrogram that are outstanding among the OWs in the CIM. We discuss these validation results in Section 4.5.



**Figure 13.** (a) N-R window (17, 11) SS coefficients of a training dataset, (b) A-R window (17, 11) SS coefficients of a training dataset.

#### 4.5. Validation Results

We validated the usefulness of our CIM-based classification based on the 855 datasets that were not used for developing the CIMs. Of these, one third (i.e., 285 sets) are normal, one third soft pitting, and one third hard pitting. The classification results using the CIM for N-R vs. A-R are shown in Figure 14a. The bulk of the normal datasets include no OWs among the 38 OWs in the CIM. The maximum number of OWs of normal data is nine. On the other hand, the bulk of the abnormal data (i.e., soft or hard pitting) highlights all OWs of the CIM. The minimum number of OWs of abnormal data is 33. With the reference OW number as 19, we can classify normal vs. abnormal gears with 100% accuracy. Similarly, we further classify the types of abnormal gears using the second CIM for S-R vs. H-R. The two groups of OW numbers can be clearly identified, as shown in Figure 14b. By setting the reference OW number as 10, we can also classify soft or hard pitting with 100% accuracy. This result is obtained with only 5% (15 samples of each type) of the total samples. Moreover, the computation time for the preprocessing of signals and making a decision is less than one second using a personal computer with an Intel® i7 processor manufactured by LENOVO in Hong Kong, China. We believe that this is a promising result.



**Figure 14.** Number of OWs for classification: (a) normal vs. abnormal gears, and (b) soft pitting vs. hard pitting.

## 5. Case Study 2: Bearing Data Set from National Aeronautics and Space Administration (NASA) Repository

### 5.1. Bearing Data Set

The data set used in this case study is obtained from the shared website of the Prognostic Data Repository of NASA (<https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>) and was often employed in many studies for the purpose of method validation. Four bearings installed in a rotating system were investigated in this study. The shaft was rotating in a constant speed (2000 revolution per minute) by alternating current motor. A radial load of 6000 pound were applied to the shaft and bearings by a spring mechanism [33]. The data used in this case study were acquired from an accelerometer installed in one of the bearings, which had been found to be defective after the continuous run for 34 days. Each data set has 20,480 sampling points with a sampling rate of 20 kHz. The vibration signals were recorded every 10 min. Signal data were continuously collected for 34 days. The further details regarding experiment setup and data acquisition conditions might be found in [34]. In order to validate the performance of CIM, we employed the same case study that Eren did in [35], where diagnosis performances of four different techniques were compared using this NASA bearing data.

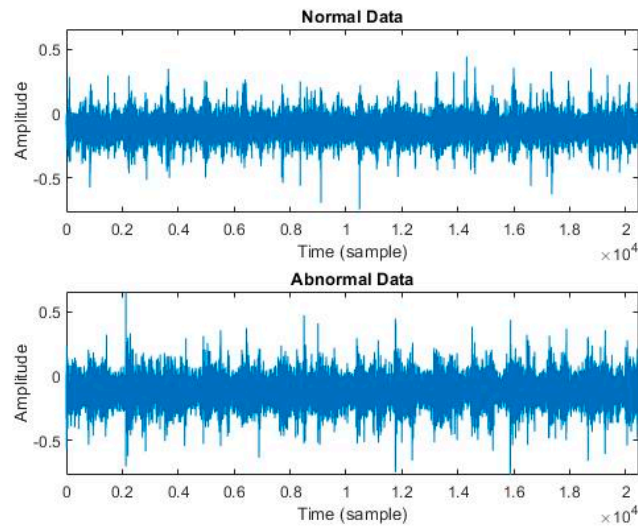
### 5.2. CIM and Comparison Results

Table 3 shows the optimal design variables achieved after optimizing (i.e., training CIM). For the training, we used only 60 normal and 60 abnormal data sets in this case study; whereas Eren used a total of 260 runs are used in both normal and abnormal cases [34]. Figure 15a shows the raw signals of normal and abnormal conditions. Figure 15b shows the CIM obtained after the training process and Figure 15c the nine OWs selected for classifying normal vs. abnormal conditions. Since the number of

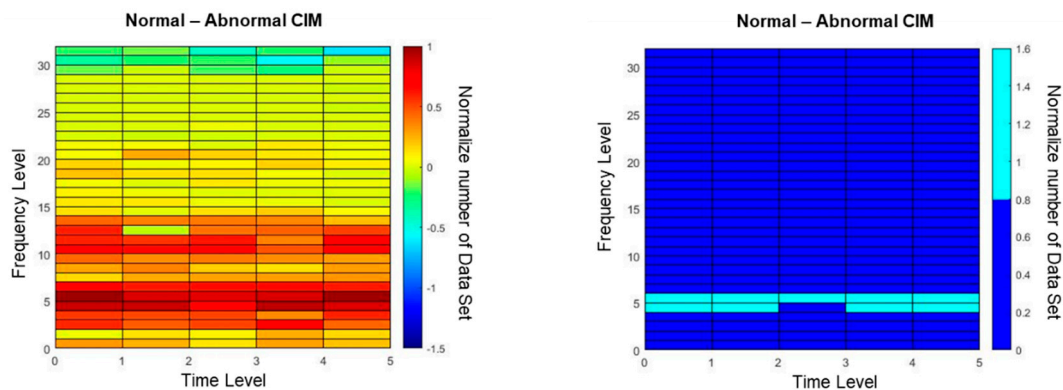
OWs of the CIM is 9, we can divide normal vs. abnormal by half of 9, which is 4.5. Figure 16 shows that 500 validation data sets can be clearly classified with borderline.

**Table 3.** Optimum design variables-bearing case.

CIMs Design Variable	CIM (N-R vs. A-R), Bearing Case
$x_1$	5
$x_2$	0.2658
$x_3$	4



(a)



(b)

(c)

**Figure 15.** (a) Normal and abnormal data in time series, (b) CIM for N-R vs. A-R, (c) selected OWs ( $R_c \geq 0.8$ ) in CIM for N-R vs. A-R.

Table 4 is a comparison of confusion matrixes between the CIM approach and 1D CNN approach [35], where both of approaches used new 500 test data sets for each condition. The CIM approach achieved better classification with zero false negative and zero false positive cases while 1D CNN results include 28 false positives and one false negative.

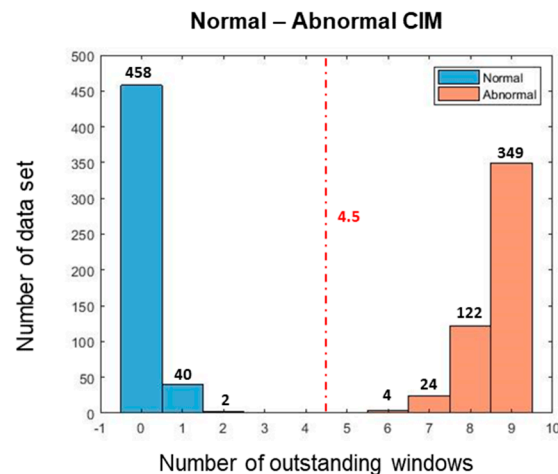


Figure 16. Number of OWs for classification in bearing case: normal vs. abnormal.

Table 4. Comparison of confusion matrices between CIM approach and 1D CNN approach.

Method		Ground Truth	
		Normal	Abnormal
Classification result of CIM	Normal	500 (TN)	0 (FP)
	Abnormal	0 (FN)	500 (TP)
Classification result of 1D CNN	Normal	472 (TN)	28 (FP)
	Abnormal	1 (FN)	499 (TP)

The standard performances, accuracy, sensitivity, specificity, and precision, of the five different methods are presented in Table 5. Accuracy, the percentage of patients correctly qualified, is  $(TP + TN)/(TP + TN + FP + FN)$ . Sensitivity, the probability of a positive test result among those having the target condition, is  $TP/(TP + FN)$ . Specificity, the probability of a negative test result among those without the target condition, is  $TN/(TN + FP)$ . Precision, the so called positive prediction value or probability of having the target condition given a positive test result, is  $TP/(TP + FP)$ . The details of the definition can be found in [36].

Table 5. Bearing fault detection performance of the proposed method.

Method	Fault Detection			
	Accuracy	Sensitivity	Specificity	Precision
Proposed CIM	100	100	100	100
1D CNN	97.1	99.8	94.4	94.7
FFT-MLP	95.0	100	90.0	90.9
FFT-RBFN	96.0	100	92.0	92.6
FFT-SVM	94.5	99.0	90.0	90.8

In Table 5, we compare the performance of the CIM with those of other approaches, one dimensional convolution neural network (1D CNN), multilayer perceptron (MLP), radial basis function networks (RBFN), and support vector machines (SVM). The performance data of these methods can be found in [35]. The proposed CIM approach has outstanding performance with 100% for all criteria with the smaller amount of training sets; whereas others' performance ranged from 90% to 100%. The comparisons demonstrated that the proposed method is a promising solution when applied to the complex rotary system.



## 6. Discussion and Closure

We identified the following observations from the two examples for the validation of the proposed method. Firstly, the proposed CIM approach requires the smaller amount of training data. In the first case study, we used only 45 datasets out of 900 samples, which is only 5% of the data, for developing the CIM. The classification results with the two developed (or trained) CIMs exhibited 100% prediction accuracy, which is an outstanding result. In the second case study, the number of training data sets are 60 out of 560 in total. The results of the case study are also 100% of prediction accuracy, which cannot be achieved by other methods with much large training data.

In addition, the CIM approach for identifying the mechanical faults could be fully automated without expert knowledge of the system being required, which is similar to the current methods, however, they need a large amount of training data. Once a signal is observed, we could accurately identify the status of a system by pre-processing (i.e., time-synchronization, TFR transformation, and spectral subtraction) and make a decision based on optimized CIM in an automated and timely manner.

The validation examples are both for non-stationary and stationary systems. As the robot arm moves upward and downward, two different types of signals, with a time gap, are obtained from the sensor. Moreover, the operating frequencies of the two movements are different as their velocities differ. In this case, the extracted signals of a frequency would be mixed with redundant data, which will increase the signal-to-noise ratio. However, the CIM approach uses TFRs and searches for OWs, which is uniquely beneficial among many different signal frequencies and time domains. This method will efficiently capture the critical information of the two completely different robot arm movements. The second case study, the rotating machine supported by the NASA repository, is a stationary system, of which classification result using the CIM approach is also very accurate with 100%.

From the above observations, we believe that the CIM-based diagnosis approach proposed in this study is functional with a small amount of training data for the diagnosis of non-stationary and stationary systems in an automated manner without expert knowledge for extracting important signal features. However, the proposed CIM approach is designed for the diagnosis of fast dynamic systems. It may not suit slow dynamic systems, such as reactors. In addition, we need further research, such as a hybrid approach using CIM and dynamic analyses, to apply this technique to prognostics.

**Author Contributions:** Conceptualization, H.-J.C.; Data curation, H.P.V. and S.H.; Methodology, J.H.; Writing—original draft, J.H. and H.-J.C.; Writing—review and editing, H.P.V., H.-J.C. and S.-K.C.

**Funding:** This research was supported by the Chung-Ang University Research Scholarship Grants in 2017. This work was partially supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government-Ministry of Trade Industry and Energy (MOTIE). (No. N0001075).

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

### Nomenclature

$t$	real time in raw signal
$\tau$	time delay of signal
$W_{j,k}^n(t)$	wavelet packet kernel function
$j$	wavelet packet index scale
$k$	wavelet packet translation operation
$n$	wavelet packet modulation parameter
$\phi$	wavelet packet first mother function
$\psi$	wavelet packet second mother function
$h(k)$	wavelet packet quadrature mirror filter
$g(k)$	wavelet packet quadrature mirror filter
$\omega_{j,m,k}(t)$	wavelet packet coefficient
$P'_s(w)$	signal spectrum denoised by spectral subtraction
$P_s(w)$	raw signal spectrum
$P_n(w)$	noise signal spectrum

$\hat{X}(b, a)$	wavelet spectrum denoised by spectral subtraction
$Y_s(b, a)$	wavelet spectrum of raw signal
$N_n(b, a)$	wavelet spectrum of noise signal
$\alpha$	reduction factor of wavelet spectral subtraction
$a$	wavelet filter location parameter
$b$	wavelet filter scale parameter
$x_1$	number of divisions in time domain
$x_2$	threshold value for parameter to be outstanding
$x_3$	minimum number of outstanding parameters for window to be outstanding
$A$	number of outstanding windows
$A_{norm}$	normalized number of outstanding windows
$D_f$	number of divisions in frequency domain
$C_p$	penalty weight for N-R subtraction
$R_c$	outstanding degree of window
$L_S$	length of synchronized datasets
$N_n$	number of normal datasets
$N_a$	number of abnormal datasets
$N_{n-r,ij}$	number of normal datasets that make window at $i$ -th row and $j$ -th col outstanding in N-R spectrogram
$N_{a-r,ij}$	number of normal datasets that make window at $i$ -th row and $j$ -th col outstanding in A-R spectrogram
$B_{ij}$	counting number of $i$ -th row and $j$ -th col

## References

- Vichare, N.M.; Pecht, M.G. Prognostics and health management of electronics. *IEEE Trans. Compon. Packag. Technol.* **2006**, *29*, 222–229. [[CrossRef](#)]
- Qian, C.; Kim, D.-S.; Lee, D.-W. On-vehicle triboelectric nanogenerator enabled self-powered sensor for tire pressure monitoring. *Nano Energy* **2018**, *49*, 126–136. [[CrossRef](#)]
- Elforjani, M.; Mba, D. Accelerated natural fault diagnosis in slow speed bearings with acoustic emission. *Eng. Fract. Mech.* **2010**, *77*, 112–127. [[CrossRef](#)]
- Al-Ghamd, A.M.; Mba, D. A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mech. Syst. Signal Process.* **2006**, *20*, 1537–1571. [[CrossRef](#)]
- Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mech. Syst. Signal Process.* **2014**, *42*, 314–334. [[CrossRef](#)]
- Li, Y.; Li, G.; Yang, Y.; Liang, X.; Xu, M. A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy. *Mech. Syst. Signal Process.* **2018**, *105*, 319–337. [[CrossRef](#)]
- Santos, P.; Villa, L.F.; Reñones, A.; Bustillo, A.; Maudes, J. An SVM-based solution for fault detection in wind turbines. *Sensors* **2015**, *15*, 5627–5648. [[CrossRef](#)] [[PubMed](#)]
- Caesarendra, W.; Pratama, M.; Kosasih, P.B.; Tjahjowidodo, T.; Glowacz, A. Parsimonious Network based on Fuzzy Inference System (PANFIS) for Time Series Feature Prediction of Low Speed Slew Bearing Prognosis. *Appl. Sci.* **2018**, *8*, 2656. [[CrossRef](#)]
- Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [[CrossRef](#)]
- Benmoussa, S.; Djezir, M.A. Remaining useful life estimation without needing for prior knowledge of the degradation features. *IEEE IET Sci. Meas. Technol.* **2017**, *11*, 1071–1078. [[CrossRef](#)]
- Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [[CrossRef](#)]

12. Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3768–3774. [[CrossRef](#)]
13. Luo, Y.; Baddour, N.; Liang, M. A shape-independent approach to modelling gear tooth spalls for time varying mesh stiffness evaluation of a spur gear pair. *Mech. Syst. Signal Process.* **2018**, *120*, 836–852. [[CrossRef](#)]
14. Endo, H.; Randall, R.B.; Gosselin, C. Differential diagnosis of spall vs. cracks in the gear tooth fillet region: Experimental validation. *Mech. Syst. Signal Process.* **2009**, *23*, 636–651. [[CrossRef](#)]
15. Park, J.; Ha, J.M.; Oh, H.; Youn, B.D.; Choi, J.; Kim, N.H. Model-based fault diagnosis of a planetary gear: A novel approach using transmission error. *IEEE Trans. Reliab.* **2016**, *65*, 1830–1841. [[CrossRef](#)]
16. Lebold, M.; McClintic, K.; Campbell, R.; Byington, C.; Maynard, K. Review of vibration analysis methods for gearbox diagnostics and prognostics. In Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, USA, 1–4 May 2000.
17. He, W.; Miao, Q.; Azarian, M.; Pecht, M. Health monitoring of cooling fan bearings based on wavelet filter. *Mech. Syst. Signal Process.* **2015**, *64*, 149–161. [[CrossRef](#)]
18. Henriquez, P.; Alonso, J.B.; Ferrer, M.A.; Travieso, C.M. Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 642–652. [[CrossRef](#)]
19. Cerrada, M.; Zurita, G.; Cabrera, D.; Sánchez, R.-V.; Artés, M.; Li, C. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mech. Syst. Signal Process.* **2016**, *70–71*, 87–103. [[CrossRef](#)]
20. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [[CrossRef](#)]
21. Razavi-Far, R.; Zanjani, M.; Saif, M. An integrated class-imbalance learning scheme for diagnosing bearing defects in induction motors. *IEEE Trans. Ind. Inf.* **2017**, *13*, 2758–2769. [[CrossRef](#)]
22. Kwon, D.; Hodkiewicz, M.R.; Fan, J.; Shibutani, T.; Pecht, M.G. IoT-Based Prognostics and Systems Health Management for Industrial Applications. *IEEE Access* **2016**, *27*, 3659–3670. [[CrossRef](#)]
23. Papoulis, A. *The Fourier Integral and Its Applications*, 1st ed.; McGraw-Hill: New York, NY, USA, 1962; pp. 244–253, ISBN 0070484473.
24. Tabrizi, A.; Garibaldi, L.; Fasana, A.; Marchesiello, S. Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. *Adv. Dyn. Stab. Control Mech. Syst.* **2014**, *50*, 865–874. [[CrossRef](#)]
25. Peng, Z.; Chu, F. Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. *Mech. Syst. Signal Process.* **2004**, *18*, 199–221. [[CrossRef](#)]
26. Verstraete, D.; Ferrada, A.; Droguett, E.L.; Meruane, V.; Modarres, M. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock Vib.* **2017**, *2017*, 5067651. [[CrossRef](#)]
27. Yen, G.G.; Lin, K.-C. Wavelet packet feature extraction for vibration monitoring. *IEEE Trans. Ind. Electron.* **2000**, *47*, 650–667. [[CrossRef](#)]
28. Berouti, M.; Schwartz, R.; Makhoul, J. Enhancement of speech corrupted by acoustic noise. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Washington, DC, USA, 2–4 April 1979.
29. Denda, Y.; Nishiura, T.; Kawahara, H.; Irino, T. Speech recognition with wavelet spectral subtraction in real noisy environment. In Proceedings of the 7th International Conference on Signal Processing, Beijing, China, 31 August–4 September 2004.
30. El Bouchikhi, E.; Choqueuse, V.; Benbouzid, M.E. Current Frequency Spectral Subtraction and Its Contribution to Induction Machines’ Bearings Condition Monitoring. *IEEE Trans. Energy Convers.* **2013**, *28*, 135–144. [[CrossRef](#)]
31. Lee, Y.; Han, S.; Jang, S.; Kim, W.; Choi, H.-J.; Choi, S.-K. Multidisciplinary materials and geometry optimization of superheater tubes for advanced ultra-supercritical power boilers. *J. Mech. Sci. Technol.* **2018**, *32*, 3359–3369. [[CrossRef](#)]
32. Staufer, P.; Gattringer, H. State estimation on flexible robots using accelerometers and angular rate sensors. *Mechatronics* **2012**, *22*, 1043–1049. [[CrossRef](#)]
33. Yu, J. Health Condition Monitoring of Machines Based on Hidden Markov Model and Contribution Analysis. *IEEE Trans. Instrum. Meas.* **2012**, *8*, 2200–2211. [[CrossRef](#)]

34. Qiu, H.; Lee, J.; Lin, J.; Yu, G. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Adv. Eng. Inform.* **2003**, *17*, 127–140. [[CrossRef](#)]
35. Eren, L. Bearing Fault Detection by One-Dimensional Convolutional Neural Networks. *Math. Probl. Eng.* **2017**, *2017*, 8617315. [[CrossRef](#)]
36. Stralen, K.J.V.; Stel, V.S.; Reisma, J.B.; Dekker, F.W.; Zoccali, C.; Jager, K.J. Diagnostic methods I: Sensitivity, specificity, and other measures of accuracy. *Kidney Int.* **2009**, *75*, 1257–1263. [[CrossRef](#)] [[PubMed](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).