

Article

# Detecting and Learning Unknown Fault States by Automatically Finding the Optimal Number of Clusters for Online Bearing Fault Diagnosis

Md Rashedul Islam , Young-Hun Kim, Jae-Young Kim  and Jong-Myon Kim \*

Department of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan 44610, Korea; rashed.cse@gmail.com (M.R.I.); dudgns09081@gmail.com (Y.-H.K.); kgy7097@gmail.com (J.-Y.K.)

\* Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Received: 1 May 2019; Accepted: 31 May 2019; Published: 6 June 2019



**Featured Application:** The proposed model of this paper is for the bearing fault diagnosis of industrial rotating machinery. Specifically, the general fault diagnosis model only can predict the bearing fault based on the predefined number of stored fault information. The proposed approach provides an online fault diagnosis process, where unknown faults are detected and updated with knowledge of the proposed diagnosis system.

**Abstract:** This paper proposes an online fault diagnosis system for bearings that detect emerging fault modes and then updates the diagnostic system knowledge (DSK) to incorporate information about the newly detected fault modes. New fault modes are detected using *k*-means clustering along with a new cluster evaluation method, i.e., multivariate probability density function's cluster distribution factor (MPDFCDF). In this proposed model, a heterogeneous pool of features is constructed from the signal. A hybrid feature selection model is adopted for selecting optimal feature for learning the model with existing fault mode. The proposed online fault diagnosis system detects new fault modes from unknown signals using *k*-means clustering with the help of proposed MPDFCDF cluster evaluation method. The DSK is updated whenever new fault modes are detected and updated DSK is used to classify faults using the *k*-nearest neighbor (*k*-NN) classifier. The proposed model is evaluated using acoustic emission signals acquired from low-speed rolling element bearings with different fault modes and severities under different rotational speeds. Experimental results present that the MPDFCDF cluster evaluation method can detect the optimal number of fault clusters, and the proposed online diagnosis model can detect newly emerged faults and update the DSK effectively, which improves the diagnosis performance in terms of the average classification performance.

**Keywords:** online fault diagnosis; *k*-means clustering; hybrid feature selection; cluster evaluation; envelope analysis; and acoustic emission

## 1. Introduction

The industrial motor is vital equipment in modern industrial application such as in pumps, air compressors, steel mills, paper mills, wind turbines, and others. It is perhaps the most commonly used rotating machine with the primary goal of providing manufacturing machinery rotation under different load conditions [1]. Thus, the bearing is the most critical and important piece of equipment in the industrial motor. However, under the continuous and heavy workload, rapid raising of the voltage pulse, and contamination of the operating environment, the metal fatigues of different parts of the bearing appear in the form of surface cracks [2]. The gradual increase of the crack severity becomes a major cause of failure of the bearing [2,3]. Finally, the unattended bearing-related failures account

for a significant portion of functional breakdowns in induction motors as well as shutdown of the industrial process.

To overcome the situation and decrease the tremendous loss of industrial production, several types of bearing condition monitoring and early fault detection models have been invented. Conventional model-based condition monitoring is most popular among them. However, model-based condition monitoring demands a deep knowledge of the whole industrial process [4,5]. Alternatively, the modern development of sensor technology, information processing techniques, and machine learning models help to invent an efficient data-driven bearing fault diagnosis model by considering a large amount of signal data during continuous industrial operations [4].

In the data-driven approach, data are collected from the actual system environment using different types of sensors for analyzing and distinguishing normal and abnormal conditions without developing the complex model of the system. To collect the data from the bearing, several types of sensor are used, i.e., current sensor, thermal sensor, vibration acceleration sensor, and acoustic emission sensor. Motor current signature analysis (MCSA) is adopted for fault diagnosis [6,7] because of the fact that it is sensitive to different failures of industrial motor and provides nonintrusive monitoring. Vibration signal analysis-based fault diagnosis is widely used for this purpose, which may provide the most information about bearing failures [8–10]. The vibration analysis and MCSA show good performance for the early detection of bearing defects. However, these approaches can only show suboptimal performance in the modern complex industrial system because of their difficulty in capturing dynamic activities at weak vibration and current levels, respectively. Acoustic emission (AE) is a popular alternative method for fault diagnosis due to its high operating frequency range and low-energy signals capturing capability [4,11–13]. AE can identify bearing abnormalities before they appear on the surface at the vibration acceleration range. Thus, AE is used to monitor for incipient bearing defects in this paper.

In order to develop an efficient fault diagnosis technique, many researchers proposed the data-driven bearing fault diagnosis model. Thomas et al. proposes a heterogeneous feature extraction process using envelop power spectrum, time and frequency domain statistical analysis, and wavelet packet decomposition analysis. In this paper, the authors adopt a simple greedy search model, which is computational time expensive. The  $k$ -NN, feedforward net, and SVM (support vector machine) are used in a supervised manner in this research [14]. Islam et al. proposed a hybrid feature selection algorithm and discriminant heterogeneous feature analysis for bearing fault diagnosis. In this paper, acoustic emission sensor is used for acquiring vibration data of the fault. A novel hybrid feature selection model is introduced for selecting optimal features from extracted heterogeneous feature vector. Finally,  $k$ -NN is used for identifying bearing fault [15]. In [16], Kun et al. proposed an adaptive frequency band selection for an empirical wavelet decomposition model for bearing fault diagnosis. In this paper, the harmonic significance index and particle swarm optimization is used for selecting an optimal sub band. There are many kinds of research where the traditional fault diagnosis models are introduced and have achieved high accuracy for detecting faults based on trained knowledge.

The traditional fault diagnosis methods use prior knowledge of failure modes to diagnose bearing condition. However, in complex modern industrial processes, bearing often operates continuously under heavy loads, which causes material fatigue and leads to faults (i.e., cracks or spalls). Thus, it is difficult to have complete failure knowledge in advance, which means traditional diagnostic systems make decisions based on incomplete knowledge. Due to the progressive nature of bearing faults, new fault modes can appear during continuous operation. Discounting information about new faults degrades the effectiveness of fault diagnosis systems. In contrast, an online fault diagnosis system considers new faults continually for enhancing the fault knowledge in reliable fault diagnosis [17]. Jiang et al. proposed a progressive fault diagnosis model, where a new hybrid ensemble auto-encoder (HEAE) is used to increase the feature quality for detecting a fault [18]. In this research, several classification algorithms are used for validating the performance. However, this work still considers a

background trained model for identifying the new fault. Also, the fault classification model is not updated by incorporating the new fault information.

To address this issue, an online fault diagnosis model is proposed in this paper that can detect new faults as they appear using unsupervised clustering technique, update its knowledge in real time, and use the updated knowledge for a reliable diagnosis. The number of faults can be determined from a set of condition monitoring data through clustering. The  $k$ -means clustering algorithm is a simple and popular way to solve the unsupervised clustering problem [19] without the need for training on existing knowledge. It can, therefore, be used to cluster unknown bearing signals and detect new faults. The  $k$ -means algorithm first selects  $k$  initial centroids as the centers of  $k$  clusters. Each sample is then assigned to the nearest cluster through an iterative process. The algorithm is sensitive to the initial centroids, which are generally selected randomly [19,20]. Several methods have been proposed to select optimal centroids. Bradley et al. proposed a cluster center initialization algorithm, which divides the overall dataset into subsamples using  $k$ -means and then selects final centroids based on minimum clustering error [21]. Likas et al. introduced global  $k$ -means, a tree-based center initialization algorithm to select initial centroids independently of initial position and empirical parameters [22]. David et al. proposed the  $k$ -means ++, which yield better clustering performance using random seeds [23]. Those algorithms, however, suffer from randomness, are computationally expensive, and introduce new parameters. In [20], Khan proposed a simple and effective initial centroid selection method that reduces randomness, improves clustering performance, and does not introduce any empirical variables. Thus, this method proposed in [20] is used as a baseline for initial centroid selection.

In  $k$ -means clustering, the number of possible clusters  $k$  is an important parameter. Since the number of fault modes in a bearing cannot be known in advance, an efficient cluster evaluation technique is required to determine  $k$ . The silhouette coefficient is a popular cluster evaluation metric [24,25] that quantifies the proximity of points in a cluster to one another, which helps in identifying the number of visible clusters. A silhouette value is calculated for each point by measuring its average dissimilarity  $a_i$  to other points in the same cluster and its minimum average dissimilarity  $b_i$  to points in other clusters. The silhouette coefficient, calculated by averaging all silhouette values, gives the number of clusters in the data. In [26], Rahman et al. proposed a computationally simpler cluster evaluation algorithm called the compactness and separation measure of clusters (COSES). It calculates cluster separability by measuring the minimum distance between cluster seeds and calculates compactness by averaging the sum of the distances of all samples of a cluster to its seed.

However, cluster evaluation techniques based on average distances are unsuitable for evaluating the feature distribution of bearings, because samples at larger distances from the cluster center tend to dominate those lying closer to it, and the uneven distribution of densities across the feature space can misrepresent the actual number of clusters in the data. To address these issues, a new cluster evaluation algorithm, i.e., multivariate probability density function's cluster distribution factor (MPDFCDF), is proposed here. The MPDFCDF identifies the optimal number of clusters and uses that information to detect new faults. Once a new fault mode is detected, the online diagnostic system knowledge (DSK) is updated through heterogeneous feature pool extraction and optimal fault signature selection based on existing knowledge and newly detected fault signals. Subsequently, the online fault diagnosis module uses  $k$ -NN for classification.

Key contributions of this paper are:

- Since it is difficult to know in advance what types of faults a healthy bearing can experience, traditional offline fault diagnosis systems classify faults on the basis of incomplete knowledge. To address this issue, we propose an online bearing fault diagnosis model that first detects unknown fault modes in real-time using  $k$ -means clustering and continually updates the DSK for more reliable fault diagnosis.
- It is difficult to determine the number of discernible fault modes through  $k$ -means clustering alone without knowing how to define  $k_{kmean}$ . To address this issue, we propose a new cluster evaluation method, the MPDFCDF, to identify the optimal number of clusters ( $k_{opt}$ ) in the fault signatures.
- To evaluate our proposed model, we recorded bearing signals for different shaft speeds and constructed a heterogeneous fault signatures pool to extract the maximum possible fault information. We used a hybrid fault signature selection to create discriminative fault signatures and build a DSK. Finally, we used the  $k$ -NN classification algorithm to estimate the classification performance.

The rest of this paper is organized as follows. Section 2 describes the data acquisition from a test rig. Section 3 provides details of the proposed online fault diagnosis model, including the DSK updating process, unknown fault mode detection using the proposed MPDFCDF cluster evaluation method, and bearing fault diagnosis. Section 4 presents the experimental results and analysis of the proposed model, and Section 5 concludes the paper.

## 2. Data Acquisition

To conduct the experiment, AE signals of different faults are collected from the collaborative experimental setup at Intelligence Dynamic Lab (Gyeonsang National University, Korea). Figure 1 depicts the block diagram of the experimental setup for collecting acoustic emission (AE) signals from bearings. In this experimental setup, an induction motor is connected to a drive-end shaft, and a non-drive-end shaft is attached to the drive-end shaft through a gearbox with a 1.52:1 gear reduction ratio. Both the drive-end and non-drive-end shafts use cylindrical rolling element bearings (model FAG NJ206-E-TVP2), where the number of rolling elements is 13, the contact angle ( $\alpha_{cangle}$ ) is  $0^\circ$ , and the roller diameter ( $B_d$ ) and pitch diameters ( $P_d$ ) is 9 mm and 46.5 mm, respectively [27]. Figure 2 presents the parts and measurements of cylindrical rolling element bearing. In this setup, the general-purpose wide-band frequency AE sensor ( $WS\alpha$ ) from Physical Acoustics Corporation (PAC) is used [28]. The operating frequency is 100–1000 kHz, peak sensitivity is 55 dB, and resonant frequency 125 kHz of the AE sensor. The AE sensor is attached at the top of the non-drive-end bearing house and the collected signals are sampled at 250 kHz using a PCI-2 based system [28]. Figure 3 presents the picture of the experimental setup and data acquisition system. Bearing faults are seeded defects or cracks of different sizes (small size crack: 3 mm and big size crack: 12 mm) as shown in Figure 4, where fault names are Bearing Crack Outer (BCO), Bearing Crack Inner (BCI), and Bearing Crack Roller (BCR).

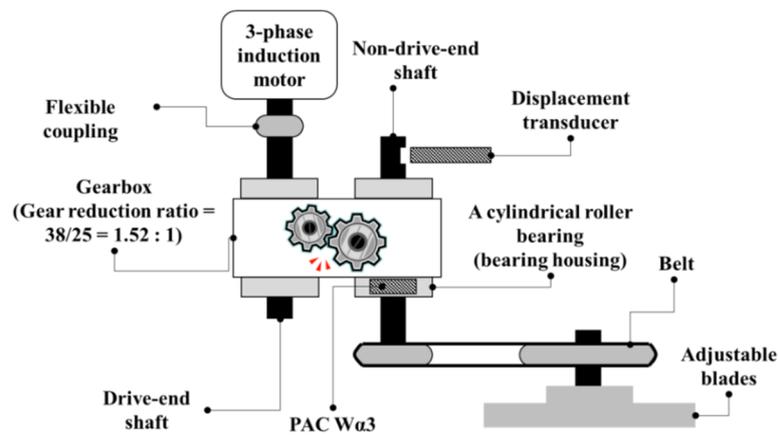


Figure 1. Block diagram of the experimental setup.

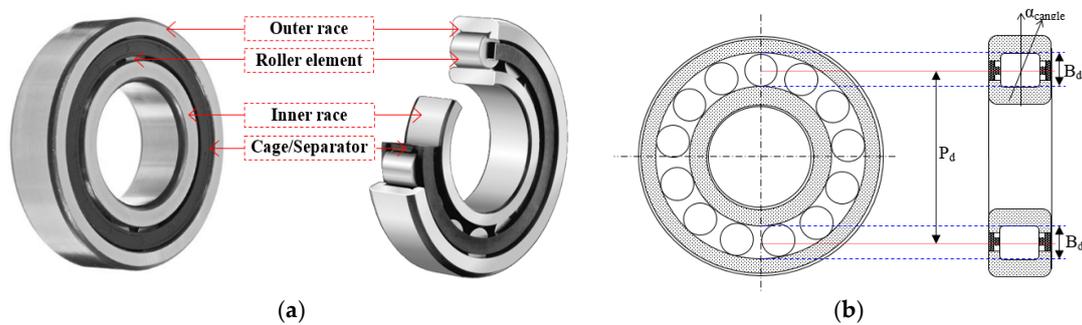


Figure 2. Cylindrical rolling element bearings, (a) parts of the bearing, (b) measurement of bearing.

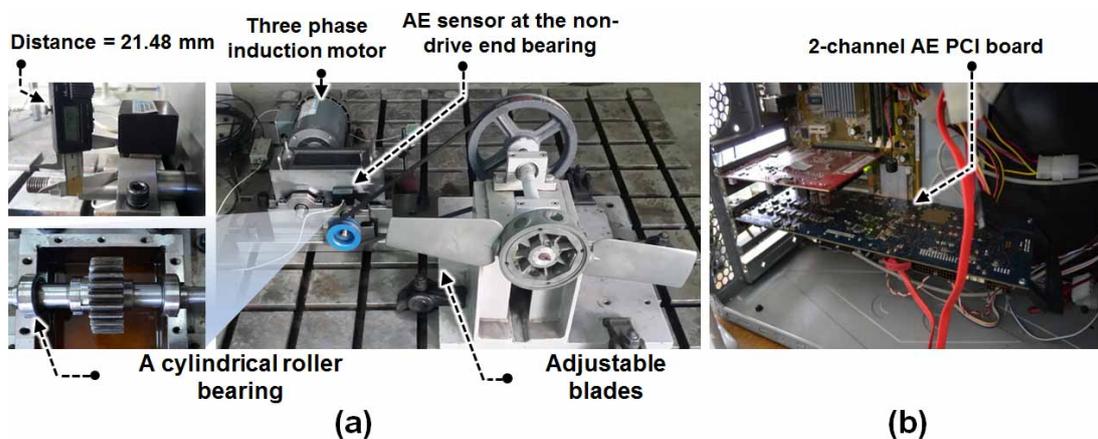


Figure 3. (a) Bearing fault signal collection test rig and (b) data acquisition system.

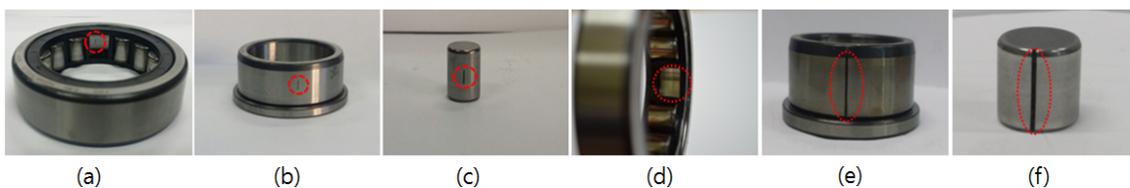
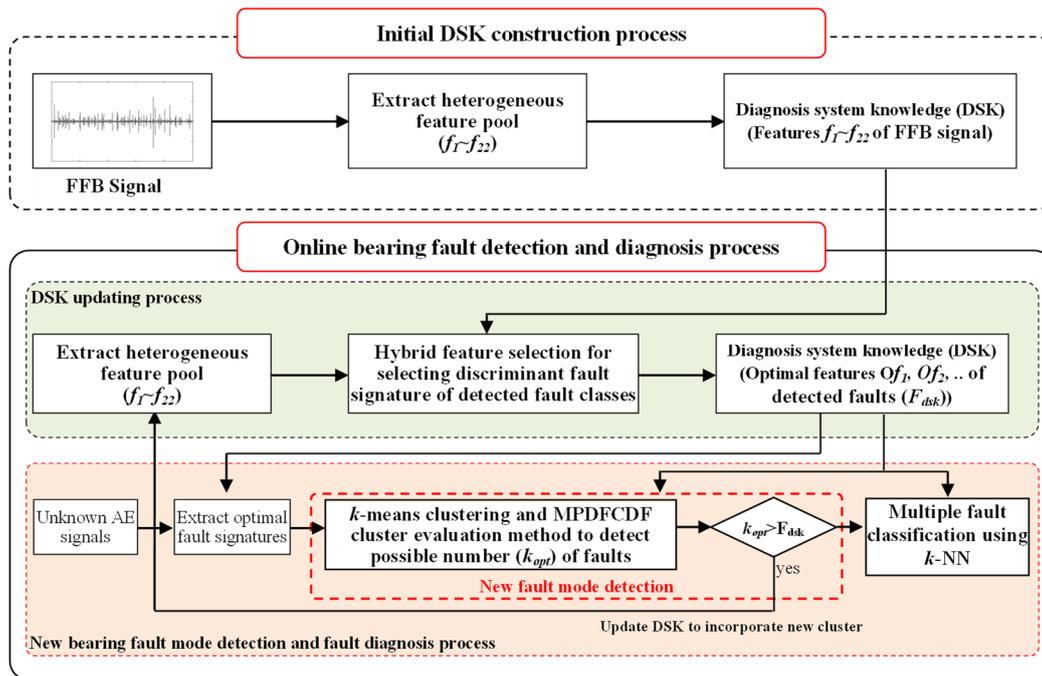


Figure 4. Faulty bearing parts with different crack severities. (a) BCO (bearing crack on the outer surface) 3 mm, (b) BCI (bearing crack on the inner surface) 3 mm, (c) BCR (bearing crack on the roller surface) 3 mm, (d) BCO 12 mm, (e) BCI 12 mm, (f) BCR 10 mm.

### 3. Proposed Online Fault Diagnosis Model

The proposed model for online fault diagnosis consists of three processes, as shown in Figure 5. An initialization process builds the initial DSK with a heterogeneous feature vector corresponding to a healthy bearing signal (FFB). An online bearing fault detection and diagnosis process detect new fault modes in unknown signals using  $k$ -means clustering and the proposed MPDFCDF cluster evaluation method. If new fault modes are detected, an updating process incorporates the newly detected fault class into the DSK, and the updated DSK is then used for fault diagnosis



**Figure 5.** Flow diagram of the proposed online bearing fault diagnosis model, where  $F_{dsk}$  is the number of fault modes in the diagnostic system knowledge (DSK).

#### 3.1. Heterogeneous Feature Pool Configuration

To discriminate fault signatures, we use maximum possible fault features extraction [14,15]. Different paradigms exist for the extraction of heterogeneous fault features, in which statistical features from time-domain and frequency-domain signals are used to uncover meaningful information for fault diagnosis. We use statistical features from the time-domain AE signal:  $f_1$ , root-mean-square (RMS);  $f_2$ , square root of the amplitude (SRA);  $f_3$ , kurtosis (KV);  $f_4$ , skewness (KV);  $f_5$ , peak-to-peak (PPV);  $f_6$ , crest factor (CF);  $f_7$ , impulse factor (IF);  $f_8$ , margin factor (MF);  $f_9$ , shape factor (SF);  $f_{10}$ , kurtosis factor (KF); and frequency-domain features of the signal, including  $f_{11}$ , frequency center (FC);  $f_{12}$ , RMS frequency (RMSF); and  $f_{13}$ , root variance frequency (RVF) [4]. Equations (1) and (2) present mathematical equations describing the extracted statistical features.

$$\begin{aligned}
 RMS &= \left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}, \\
 SRA &= \left( \frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right)^2, \\
 KV &= \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^4, \\
 SV &= \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^3, \\
 PPV &= \max(x_i) - \min(x_i), \\
 CF &= \frac{\max(|x_i|)}{\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}}, \\
 IF &= \frac{\max(|x_i|)}{\frac{1}{N} \sum_{i=1}^N |x_i|}, \\
 MF &= \frac{\max(|x_i|)}{\left( \frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right)^2}, \\
 SF &= \frac{\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}}{\frac{1}{N} \sum_{i=1}^N |x_i|}, \\
 KF &= \frac{\frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^4}{\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^2}
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 FC &= \frac{1}{N} \sum_{i=1}^N f_i, \\
 RMSF &= \left( \frac{1}{N} \sum_{i=1}^N f_i^2 \right)^{1/2}, \\
 RVF &= \left( \frac{1}{N} \sum_{i=1}^N (f_i - FC)^2 \right)^{1/2}
 \end{aligned} \tag{2}$$

To identify particular failure modes, bearing defect properties must be observed at different signal frequencies. Therefore, we extract *RMSF* values from frequency ranges that include harmonics of the following bearing defect frequencies [29]: Ball pass frequency of the outer raceway (*BPFO*), ball pass frequency of the inner raceway (*BPMI*), ball spin frequency (*BSF*), and fundamental train frequency (*FTF*). They are calculated as follows:

$$\begin{aligned}
 BPFO &= \frac{N_{rollers} \times F_{shaft}}{2} \left( 1 - \frac{B_d}{P_d} \cos \alpha_{cangle} \right), \\
 BPMI &= \frac{N_{rollers} \times F_{shaft}}{2} \left( 1 + \frac{B_d}{P_d} \cos \alpha_{cangle} \right), \\
 BSF &= \frac{P_d \times F_{shaft}}{2 \times B_d} \left( 1 - \left( \frac{B_d}{P_d} \cos \alpha_{cangle} \right)^2 \right), \\
 FTF &= \frac{F_{shaft}}{2} \left( 1 - \frac{B_d}{P_d} \cos \alpha_{cangle} \right),
 \end{aligned} \tag{3}$$

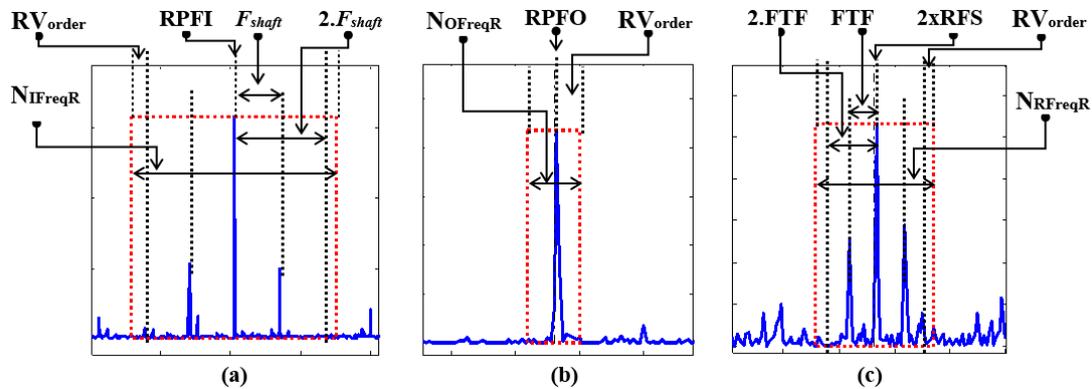
where  $N_{rollers}$  is the number of rollers,  $F_{shaft}$  is the shaft speed in hertz,  $\alpha_{cangle}$  is the contact angle,  $B_d$  is the roller diameter, and  $P_d$  is the pitch diameter.

*RMSF* features are calculated by first determining the power spectral envelope of the AE signal and then identifying the defect frequency ranges of *BPFO*, *BPMI*, and  $2 \times BSF$  up to their third harmonics. Except for *BPFO*, which is caused by the force of the roller on the stationary outer raceway fault, these frequencies can experience some modulation. *BPMI* and its harmonics are amplitude-modulated by  $F_{shaft}$ , and  $2 \times BSF$  is modulated by the rotational speed of the cage (*FTF*). Hence, amplitude modulation produces sidebands as an explicit symptom of inner-raceway and roller faults in bearings. Furthermore,

random variations in the analytic bearing defect frequencies on the order of 1–2% are observed [30], which are accounted for by the factor  $RV_{order} = 2\%$  [27]. Frequency ranges for outer-related defects ( $N_{OFreqR,i}$ ), inner-related defects ( $N_{IFreqR,i}$ ), and roller-related defects ( $N_{RFreqR,i}$ ) are computed using Equation (4) and depicted in Figure 6.

$$\begin{aligned}
 N_{OFreqR,i} &= (1 - RV_{order}) \cdot (BPFO_i) \sim (1 + RV_{order}) \cdot (BPFO_i) \\
 N_{IFreqR,i} &= (1 - RV_{order}) \cdot (BPFI_i - 2 \cdot F_{shaft}) \sim (1 + RV_{order}) \cdot (BPFI_i + 2 \cdot F_{shaft}) \\
 N_{RFreqR,i} &= (1 - RV_{order}) \cdot (2 \times BSF_i - 2 \cdot FTF) \sim (1 + RV_{order}) \cdot (2 \times BSF_i + 2 \cdot FTF)
 \end{aligned}
 \tag{4}$$

where  $i$  represents the harmonic number. Using these frequency ranges, we compute nine RMSF values for three types of defect frequencies up to their third harmonic:  $RMSF_{BPFI1}(f_{14})$ ,  $RMSF_{BPFI2}(f_{15})$ ,  $RMSF_{BPFI3}(f_{16})$ ,  $RMSF_{BPFO1}(f_{17})$ ,  $RMSF_{BPFO2}(f_{18})$ ,  $RMSF_{BPFO3}(f_{19})$ ,  $RMSF_{2 \times BSF1}(f_{20})$ ,  $RMSF_{2 \times BSF2}(f_{21})$ ,  $RMSF_{2 \times BSF3}(f_{22})$ . In total, we extract 22 features from each signal to create its heterogeneous feature pool.



**Figure 6.** Calculation of  $N_{OFreqR,i}$ ,  $N_{IFreqR,i}$ , and  $N_{RFreqR,i}$ . (a) Defect frequencies of BCI; (b) defect frequencies of BCO; (c) defect frequencies of BCR.

### 3.2. DSK Construction by Selecting Discriminant Fault Signatures of Detected Fault Modes

Discriminant features of candidate data classes are important to effectively distribute class samples and improve classification accuracy. However, the classifier knowledge varies depending on the number of emerging fault classes. In a high-dimensional feature space, features differ in their ability to effectively discriminate a given fault condition. Irrelevant or redundant features degrade the performance of predictive models (e.g., supervised and unsupervised classifiers). To address this issue, we use a discriminant fault signature selection process to build a DSK of detected fault modes.

Feature selection involves the evaluation of feature subsets based on processes that are generally either wrapper or filter based [31]. However, diagnostic performance can be improved by combining the advantages of both the wrapper-and filter-based approaches [15,32,33]. In this study, we introduce a hybrid feature selection algorithm that uses an effective combination of filter and wrapper methods to select the most discriminant signature of emerging faults. In the selection process, extracted features of detected fault classes ( $N_{detected\_class} \times N_{sample} \times N_{feature}$ ) are divided into two halves, where one half is used for the filter selection process and the other half is used for the wrapper selection process ( $N_{feature}$  is the size of the original feature vector). In the filter selection process, ( $N_{detected\_class} \times N_{sample}/2 \times N_{feature}$ ) are randomly divided into a  $k$ -fold cross validation  $k_{cv}$ , where  $k = 2$ . Optimal feature sets are selected using sequential forward selection (SFS) with feature subset evaluation. The filter selection process selects 10 suboptimal feature subsets from  $N_{iteration} \times k_{cv}$ , where  $N_{iteration} = 5$ .

The 10 suboptimal subsets are combined to create a feature occurrence histogram. Subsequently, in the wrapper selection process, the algorithm selects feature combinations from different levels of the feature occurrence histogram and estimates the average classification accuracy of the  $k$ -NN classifier. In the classification process, 50% of the ( $N_{detected\_class} \times N_{sample}/2 \times N_{feature}$ ) are used as a training set and

the rest are used as a resting set. The wrapper selection process picks the best feature combination depending on the maximum average classification accuracy. Figure 7 illustrates the overall flow of the DSK construction by selecting discriminant fault signatures of detected fault modes.

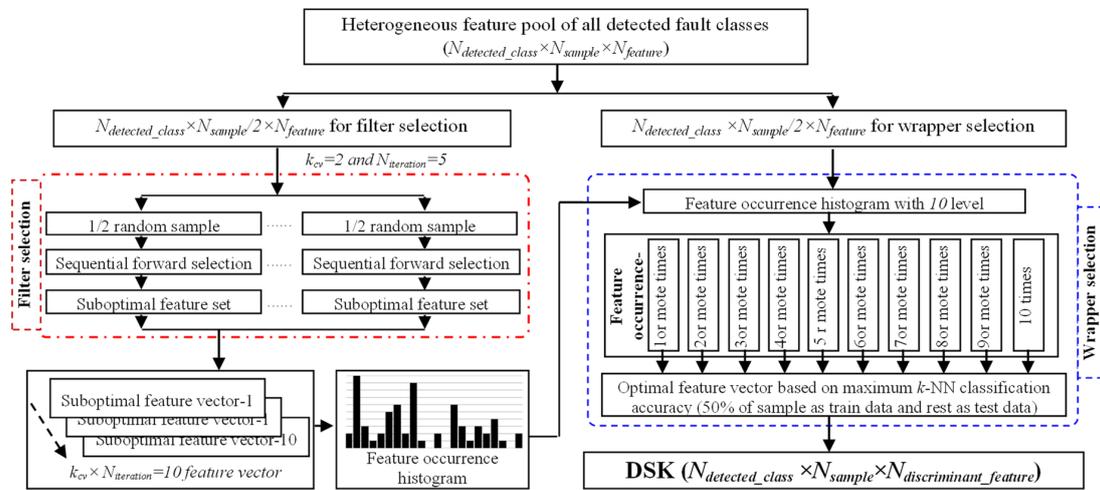


Figure 7. The flow of DSK construction by selecting discriminant fault signatures of detected fault modes.

Usually, the SFS algorithm evaluates different subsets and selects an optimal subset based on an objective value. In this study, we adopt the maximum class separability (MCS) feature subset evaluation algorithm proposed by Islam et al., which is illustrated in Figure 8 [34]. According to this algorithm, the final objective value is calculated as Equation (5).

$$objective\_value = \frac{between\_class\_separability}{within\_class\_compactness} \tag{5}$$

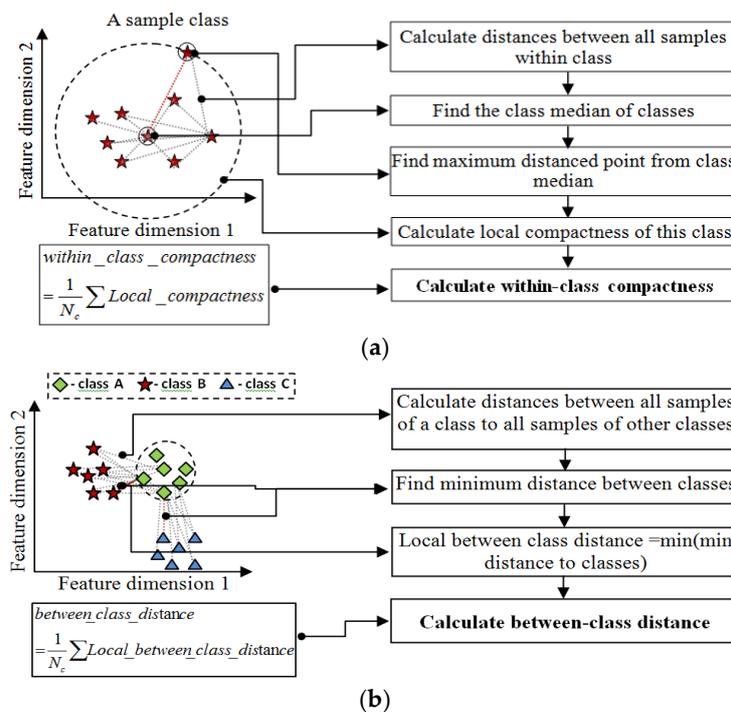
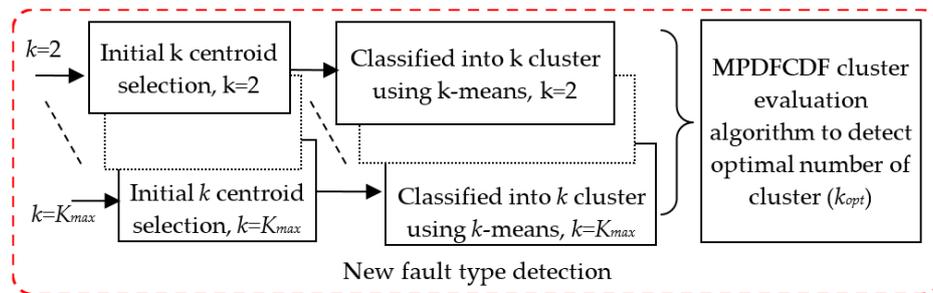


Figure 8. Feature subset evaluation algorithm: (a) Within-class compactness value, and (b) between-class distance value.

### 3.3. Proposed New Fault Mode Detection

As mentioned in Section 1, a new fault can appear or the severity of an existing crack can increase during the diagnosis process. Unsupervised clustering algorithms are appropriate for detecting an unknown number of faults in a fault signal. The process of our proposed online fault diagnosis system is illustrated in Figure 9. This process compares the optimal features extracted from unknown signals with the existing DSK to identify the number of fault modes.



**Figure 9.** Proposed fault mode detection using  $k$ -means clustering and a novel cluster evaluation algorithm to detect the optimal number of clusters  $k$ .

#### 3.3.1. k-Means Clustering

$k$ -means clustering is the simplest unsupervised clustering algorithm and has been adopted in several application domains [35,36]. In  $k$ -means clustering,  $k$  centroids are initially selected at random, and data samples  $\{x_1, \dots, x_n\}$  are assigned to disjoint clusters corresponding to the nearest centroid based on a clustering criterion function  $F$  [19], which is generally the squared distance between each sample  $x_i$  and the centroid  $c_j$  of cluster  $C_j$ , as given in Equation (6)

$$F(c_1, \dots, c_k) = \sum_{j=1}^k \sum_{i=1}^{N_j} \|x_{i,j} - c_j\|^2 \tag{6}$$

where  $c_1 \dots c_k$  are the centroids of clusters  $C_1 \dots C_k$ , respectively,  $k$  is the total number of clusters, and  $N_j$  is the total number of samples assigned to the  $i$ th cluster. After data samples are assigned to clusters, centroids are relocated based on their membership information.

This process is repeated until there is no change in the value of  $F$  [37]. Since the  $k$ -means clustering algorithm is sensitive to the  $k$  initial centroids, it may converge to a suboptimal solution [20,38]. Many algorithms have been devised to reduce the sensitivity of the process to  $k$ , but they inevitably increase the computational cost or introduce new variables. In this study, we use the initial centroid selection method of Khan [20], which considers the deepest valleys and highest gaps between clusters in a dataset. This algorithm sorts the data samples in ascending order of their magnitude  $d_1, \dots, d_n$ , then calculates the Euclidean distance between each pair of consecutive points,  $D_i = \text{dist}(d_i - d_{i+1})$ , where  $i = 1 \dots n - 1$ . Next, the distances  $D_i$  are sorted in descending order while keeping their original indices  $i$  for identifying the position numbers after further processing. The algorithm selects the first  $k - 1$  distances, where  $k$  is the user-defined number of clusters, and points corresponding to the selected distances are combined into an upper bound set of  $k$  data groups  $\{i_1 \dots i_{k-1}, n\}$ , and a lower bound set of  $k$  data groups  $\{1, i_1 + 1, \dots, i_{k-1} + 1\}$ . Finally, center medians are calculated using the data points between the lower and upper bound positions. These center medians are the initial  $k$  centroids and are used as the initial seeds for the  $k$ -means clustering process. This process is illustrated in Figure 10.

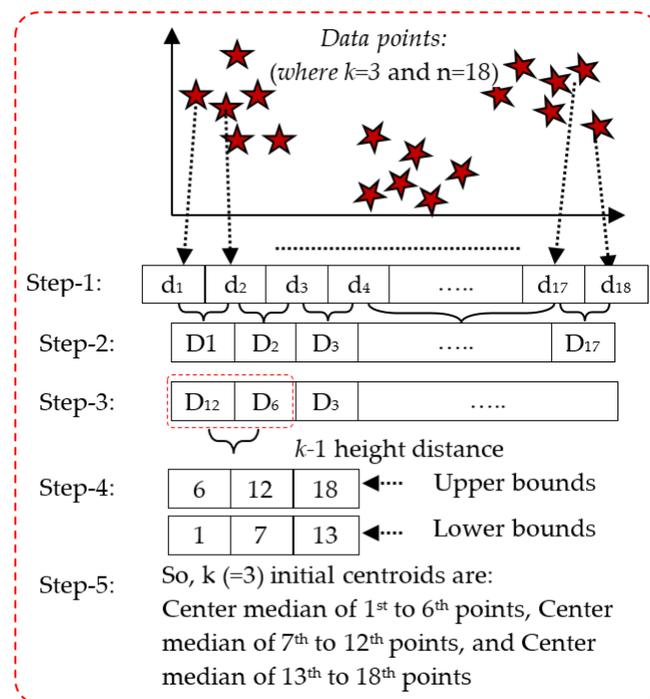
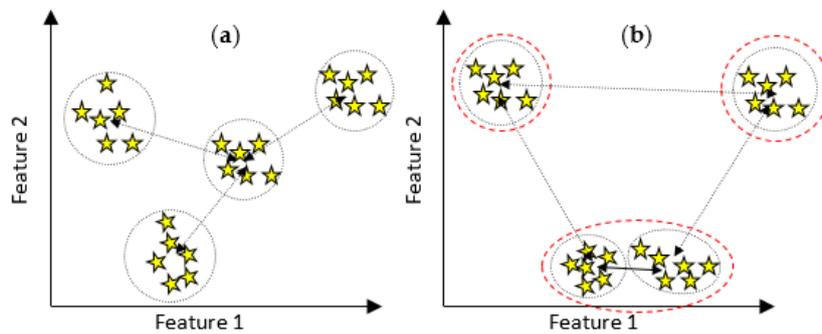


Figure 10. Initial centroids selection process with an example.

### 3.3.2. The Proposed MPDFCDF Cluster Evaluation Method

As mentioned in Section 1,  $k$ -means clustering requires  $k_{kmeans}$  to be specified in advance. However, it is difficult to know the number of clusters (fault modes) of unknown signals in online fault detection and diagnosis system. This problem can be resolved by repeating the clustering process for different values of  $k_{kmeans}$  (e.g.,  $k_{kmeans} = 2 \dots 7$ ) and then determining  $k_{opt}$  through an efficient cluster evaluation method. To do this, elbow is a well-known unsupervised cluster evaluation technique [39,40]. The elbow is a visual graph evaluation technique, where the number of clusters is increased one by one and evaluate a cost function of the clusters in every steps. The cost function is basically the calculation of the sum of squared errors of clusters. During the time of increasing the number of clusters, the value of cost function dramatically decreased and then became steady. Based on the curve of the graph, it determines the optimal number of cluster. However, the visual determination is often ambiguous. Also, the evaluation of the cost function of clusters is challenging.

The silhouette coefficient [24] and COSES [26] methods work as a cluster evaluation for well-separated data. However, data from faulty bearings have an unusual and complex distribution, which is not handled well by average-distance-based cluster evaluation methods. In Figure 11a, data are well separated into easily identifiable clusters, but in Figure 11b, two data groups are in close proximity to each other but far away from the other groups. In such cases, the larger distance dominates over the smaller distance, leading to incorrect identification of the number of clusters. Furthermore, unequal densities of data groups could bias the cluster evaluation. To address these issues, the MPDFCDF is proposed, which considers the multivariate Gaussian distribution and probability density functions (PDFs) of the data in feature space.



**Figure 11.** Two cluster distribution examples, (a) well separated data into easily identifiable clusters and (b) two data groups are in close proximity to each other but far away from the other groups.

The PDF  $\rho(x)$  can be used to express the statistical distribution of data in clusters. The univariate PDF of a normal distribution is given by Equation (7), where  $\mu$  is the mean, and  $\sigma^2$  is the variance

$$\rho(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \tag{7}$$

However, real-world problem sets exhibit behavior that is more suitably represented by multivariate distributions, such as the multivariate normal distribution given by Equation (8)

$$\rho(X) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} \exp\left(-\frac{1}{2}(X - \mu)' \Sigma^{-1} (X - \mu)\right) \tag{8}$$

where  $d$  is the dimension of data,  $X$  and  $\mu$  are  $1 \times d$  vectors, and  $\Sigma$  is the  $d \times d$  symmetric positive definite covariance matrix.

The properties of these distributions and the distance between their means represent key information that can help in identifying  $k_{opt}$ . If samples are distributed in a scattered way, the probability that data exist in dense areas is low, and the deviation of samples from the mean is high. On the other hand, if samples are distributed in a compact way, the probability that data exist in dense areas is high and the deviation of samples from the mean is low. Thus, the ratio of the highest PDF value to the variance of the distribution is a good measure of the cluster density, whereas distances between means of clusters measure the cluster separation.  $k_{opt}$  is identified by evaluating clustered samples on these criteria, using different values of  $k$ . The MPDFCDF cluster evaluation method, thus, consists of the following steps, which are illustrated in Figure 12:

- Step 1: Classify samples belonging to  $k$  clusters by  $k$ -means clustering;
- Step 2: Calculate the mean  $\mu$  and covariance  $\Sigma$  of all clusters;
- Step 3: Calculate the PDFs for all clusters using Equation (8);
- Step 4: Calculate the local distribution factor for each cluster;

$$Local\_Distribution\_Factor_c = \frac{\max(\rho_c(X))}{2 \times \max(\Sigma_c)}$$

the 2-sigma rule [41] is used here, which states that 95% of samples exist within 2-sigma of the mean and the remaining 5% can be regarded as outliers.

- Step 5: Calculate the global density factor for the distribution

$$Global\_Density\_Factor = \min(Local\_density\_factor);$$

- Step 6: Calculate the global separability factor

$$Global\ Separability\ Factor = \min(Inter\_Cluster\_Dist)$$

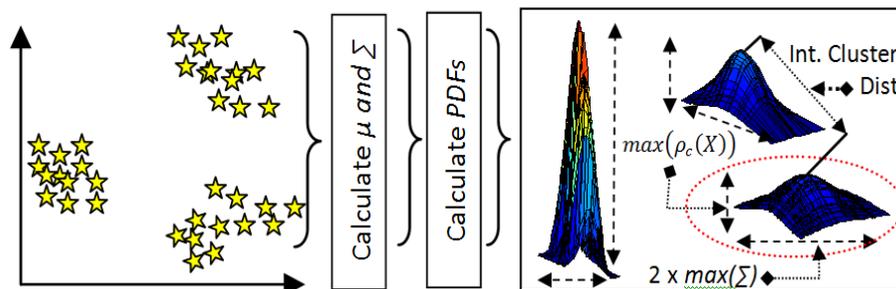
where

$$Inter\_Cluster\_Dist(c_i - c_j) = \sqrt{(c_i - c_j)^2};$$

- Step 7: Calculate the MPDFCDF

$$MPDFCDF = abs(Global\ Density\ Factor - Global\ Separability\ Factor)$$

where the minimum MPDFCDF value identifies  $k_{opt}$ .



**Figure 12.** Steps of the multivariate probability density function’s cluster distribution factor (MPDFCDF) cluster evaluation method.

### 3.4. System Update

In the proposed online fault detection and diagnosis model, known fault data from the DSK are combined with the optimal fault signature of unknown signals to detect new fault modes.  $k$ -means clustering and the proposed MPDFCDF cluster evaluation method are used to determine the number of fault modes ( $k_{opt}$ ) in the combined data. If  $k_{opt}$  is larger than the number of known faults  $F_{dsk}$ , the model automatically updates the DSK to incorporate the newly detected fault class.

### 3.5. Fault Classification Using $k$ -NN

The proposed online fault diagnosis model classifies unknown samples using  $k$ -NN [11,42]. We validate the model by estimating the generalized classification accuracy through  $k$ -fold cross-validation. The  $k$ -NN classifier classifies a sample based on the majority class among its  $k_{knn}$  nearest neighbors, which are determined using distance criteria. Several distance criteria have been used in previous studies, such as the Euclidean distance, correlation between samples, city block distance, cosine distance, and Hamming distance. In this study, we use the Euclidean distance, which is the most widely used distance criterion and can be formulated as Equation (9).

$$dist(x_i, x_j) = \sqrt{\sum_{d=1}^D (x_{i,d} - x_{j,d})^2} \tag{9}$$

where  $dist(x_i - x_j)$  represents the Euclidean distance between two data points  $x_i$  and  $x_j$ , and samples are represented by  $d$  feature dimensions.

## 4. Experimental Results and Analysis

### 4.1. Experimental Datasets

For the experiment of the proposed model, signals were collected from a fault-free bearing (FFB) and faulty bearings at various shaft speeds. In this study, six fault modes are considered: A 3-mm crack on the outer raceway (BCO 3 mm), a 3-mm crack on the inner raceway (BCI 3 mm), a 3-mm crack on the roller (BCR 3 mm), a 12-mm crack on the outer raceway (BCO 12 mm), a 12-mm crack on the inner raceway (BCI 12 mm), and a 10-mm crack on the roller (BCR 12 mm). Signals are collected for different rotational speeds (300 rpm to 500 rpm). Finally, AE data are divided into five datasets based on the shaft rotational speed, where each dataset contains a total of 630 AE signals of 5 s duration each (i.e., 90 AE signals for FFB and 90 for each of the six bearing faults). Table 1 presents the details of the experimental dataset.

**Table 1.** Various datasets with the crack severity specifications considered in this study.

		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
Average RPM		300 rpm	350 rpm	400 rpm	450 rpm	500 rpm
Fault Severity	Small crack	Crack length: 3 mm, width: 0.35 mm, depth: 0.3 mm on outer raceway, inner raceway, and roller				
	Big crack	Crack length: 12 mm on outer and inner raceways and 10 mm on roller, width: 0.49 mm, depth: 0.5 mm				

To validate the proposed model, five test cases are defined to progressively introduce new faults (small or big cracks) into the system, as shown in Table 2. Actually, to diagnosis the continues growth of different faults, the data acquisition process should run for a long time (several years) during operation of a bearing, which is very difficult. Thus, to validate the proposed online bearing fault and identify new faults, faulty signals of different faults are incorporate in progressive test cases. Finally, the proposed model is validated by five considered datasets. Each dataset is divided into two halves, one for analysis and the other for evaluation. The analysis dataset, containing 45 of 90 signals from each class, is used to detect fault modes and update the DSK. The evaluation dataset, also containing 45 samples from each class, is used for diagnosis. It is considered that the system only starts with the diagnostic knowledge of a healthy machine, originating from 30 FFB signals, where each signal is represented by a 22-dimensional feature vector. Each test case uses 30 randomly-selected samples from the corresponding fault class in the analysis dataset to evaluate the unknown fault mode detection process and update the system. For evaluation, we randomly select 30 samples from each candidate fault class in the evaluation dataset.

**Table 2.** Organization of different test cases.

	<b>Initial</b>	<b>Test Case 1</b>	Update System knowledge	<b>Test Case 2</b>	Update System knowledge	<b>Test Case 3</b>	No Update	<b>Test Case 4</b>	Update System knowledge	<b>Test Case 5</b>	Update System knowledge	<b>Final</b>
System Condition	FFB	FFB		FFB BCI 3 mm		FFB BCI 3 mm BCO 3 mm BCR 3 mm		FFB BCI 3 mm BCO 3 mm BCR 3 mm		FFB BCI 3 mm BCO 3 mm BCR 3 mm BCI 12 mm		FFB BCI 3 mm BCO 3 mm BCR 3 mm BCI 12 mm BCO 12 mm BCR 12 mm
Unknown Signals		BCI 3 mm		BCO 3 mm BCR 3 mm		BCI 3 mm BCR 3 mm		BCI 12 mm		BCO 12 mm BCR 12 mm		

4.2. Identification of the Optimal Number of Clusters Kopt Using the MPDFCDF Cluster Evaluation Method

The proposed MPDFCDF cluster evaluation method determines the optimal number of clusters by considering cluster density and inter-cluster separation based on the multivariate normal PDF. In Figure 13, the MPDFCDF is compared with two conventional state-of-the-art methods for five sample distributions containing multiple fault signals.

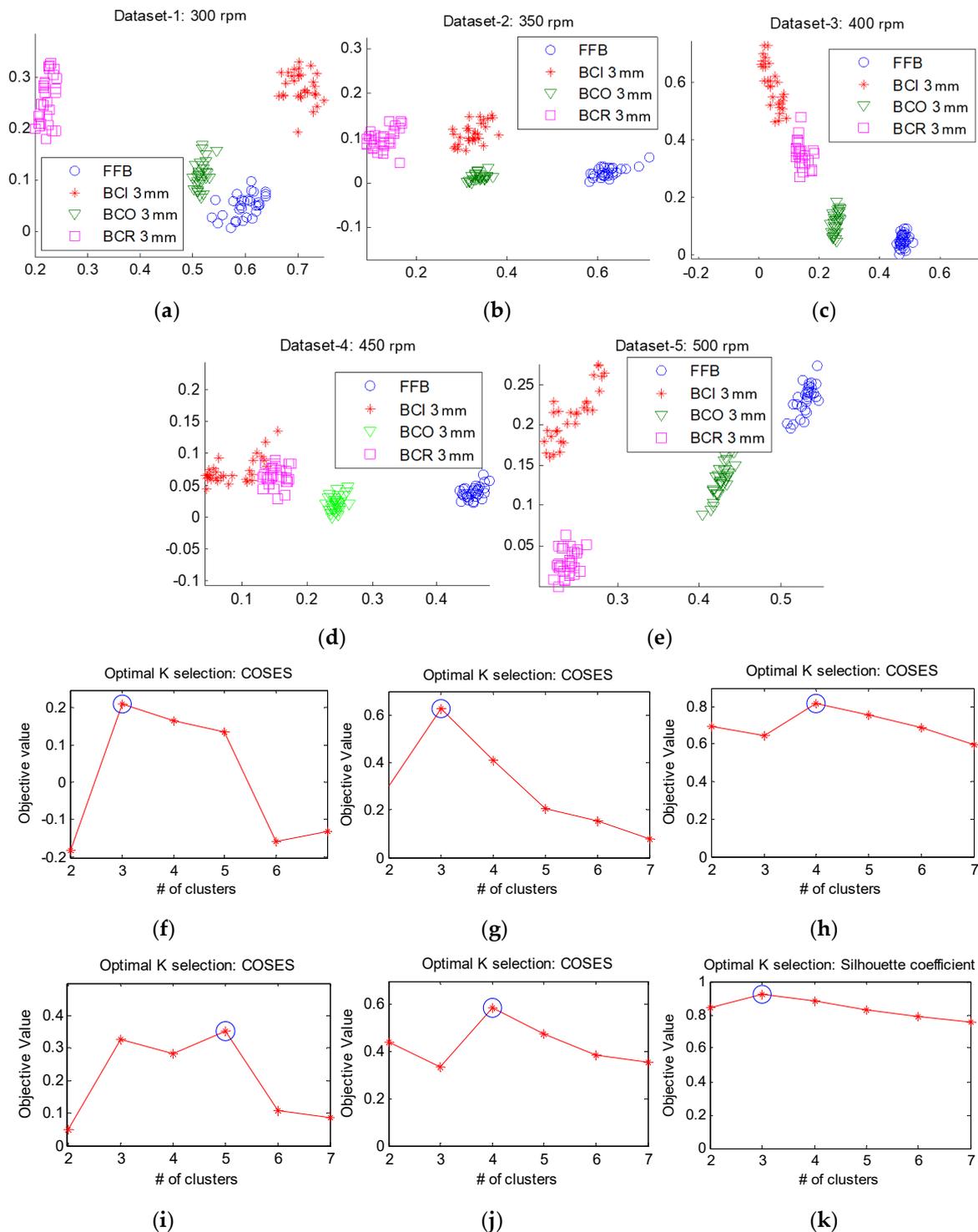
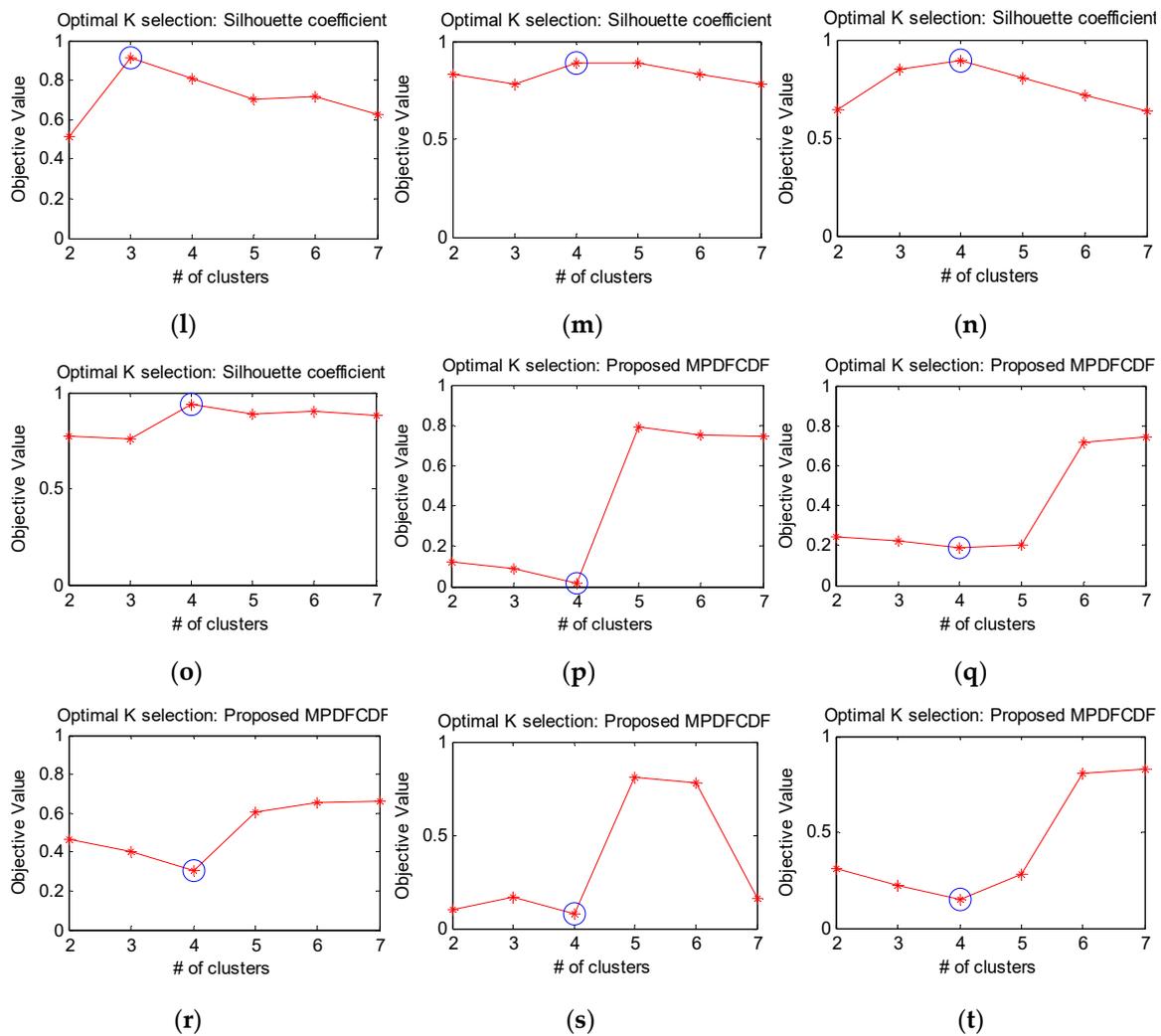


Figure 13. Cont.



**Figure 13.** (a–e) Sample distribution of different datasets and optimal  $k_{opt}$  (blue big circle) selection using the compactness and separation measure of (f–j) clusters (COSES) method, (k–o) silhouette coefficient, and (p–t) the proposed MPDFCDF cluster evaluation method.

### 4.3. New Fault Mode Detection and System Update for Online Fault Diagnosis

Initially, we assume that the system only has knowledge of a healthy machine. Thus, the DSK includes 22 features of each of the 30 FFB signals. For each test case, the system extracts an optimal number of features from unknown signals and adds those features to the DSK to form the analysis set. The new fault mode detection module detects  $k_{opt}$ , which represents the number of fault modes in the analysis set. If  $k_{opt}$  is greater than  $F_{dsk}$ , the number of features in the current DSK, the new fault signatures are added to the DSK.

Here, system knowledge is represented as  $C_{dsk} \times N_{dsk} \times F_{dsk}$

As a central part of the online diagnosis system, the DSK updating process extracts a heterogeneous feature pool from the signals of all detected fault modes. The hybrid feature selection algorithm selects the discriminant fault signatures and constructs a DSK, which is used in the evaluation process. Table 3 summarizes the optimal feature subsets for different test cases with different datasets. After completing each test case, the DSK is represented as  $C_{dsk} \times N_{dsk} \times F_{dsk}$ , where  $C_{dsk}$  is the number of emerging fault classes,  $N_{dsk}$  is the number of samples in each class, and  $F_{dsk}$  is the number of discriminant fault signature variables.

**Table 3.** Optimal feature vector for different test cases of different datasets.

Datasets	Test Cases											
	Initial		Test Case 1		Test Case 2		Test Case 3		Test Case 4		Test Case 5	
	Features	System Knowledge	Features	System Knowledge	Features	System Knowledge	Features	System Knowledge	Features	System Knowledge	Features	System Knowledge
1	f1~f22	1 × 30 × 22	f1, f12, f13, f15	2 × 30 × 4	f2, f9, f10	4 × 30 × 3	f2, f9, f10	4 × 30 × 3	f2, f9	5 × 30 × 2	f2, f9, f13	7 × 30 × 3
2	f1~f22	1 × 30 × 22	f2, f9, f14	2 × 30 × 3	f2, f9, f22	4 × 30 × 3	f2, f9, f22	4 × 30 × 3	f2, f9, f13	5 × 30 × 3	f2, f9, f11, f22	7 × 30 × 4
3	f1~f22	1 × 30 × 22	f2, f15, f16, f20	2 × 30 × 4	f2, f16	4 × 30 × 2	f2, f16	4 × 30 × 2	f2, f11, f15, f16	5 × 30 × 4	f2, f9, f11	7 × 30 × 3
4	f1~f22	1 × 30 × 22	f2, f9, f13	2 × 30 × 3	f2, f9, f20, f21	4 × 30 × 4	f2, f9, f20, f21	4 × 30 × 4	f2, f9	5 × 30 × 2	f2, f9, f13	7 × 30 × 3
5	f1~f22	1 × 30 × 22	f2, f9	2 × 30 × 2	f2, f9, f20, f21	4 × 30 × 4	f2, f9, f20, f21	4 × 30 × 4	f2, f9, f22	5 × 30 × 3	f2, f11	7 × 30 × 2

#### 4.4. Effectiveness of Online Fault Diagnosis

The effectiveness of the proposed online fault diagnosis system is evaluated by considering its classification performance using the  $k$ -NN classifier. For each test case, the DSK is used as a training set for the  $k$ -NN classifier, where  $k_{knn}$  is set to 5 (the number of nearest neighbors  $k$  is equal to the nearest integer to  $\sqrt{N_{dsk}}$  [25]). The evaluation process includes five progressive test cases  $t$ , each containing an evaluation set  $C_{eval, t} \times N_{eval, t}$ . To calculate the classification accuracy before updating the system, the DSK of the previous test case  $C_{dsk, t-1} \times N_{dsk, t-1} \times F_{dsk, t-1}$  is used as a training set and  $C_{eval, t} \times N_{eval, t} \times F_{dsk, t-1}$  is used as a test set. To calculate the classification accuracy after updating the system, the DSK of the current test case  $C_{dsk, t} \times N_{dsk, t} \times F_{dsk, t}$  is used as a training set and  $C_{eval, t} \times N_{eval, t} \times F_{dsk, t}$  is used as a test set. In the classification process, we use  $k$ -cv (cross-validation), where  $k_{cv}$  is set to 3. Thus,  $N_{eval, t}$  is randomly divided into thirds, and each third is used for testing during each  $k$ -cv iteration. The cross-validation process is executed  $N_{iteration}$  times ( $N_{iteration} = 10$  in this study), and the generalized classification performance is calculated by averaging the classification accuracy. Table 4 presents the diagnosis performance for the five test cases in terms of the average sensitivity per class and the average classification accuracy (ACA), defined in Equations (10) and (11), respectively.

$$\text{Sensitivity} = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100(\%), \quad (10)$$

$$\text{ACA} = \frac{\sum N_{TP}}{N_{samples}} \times 100(\%) \quad (11)$$

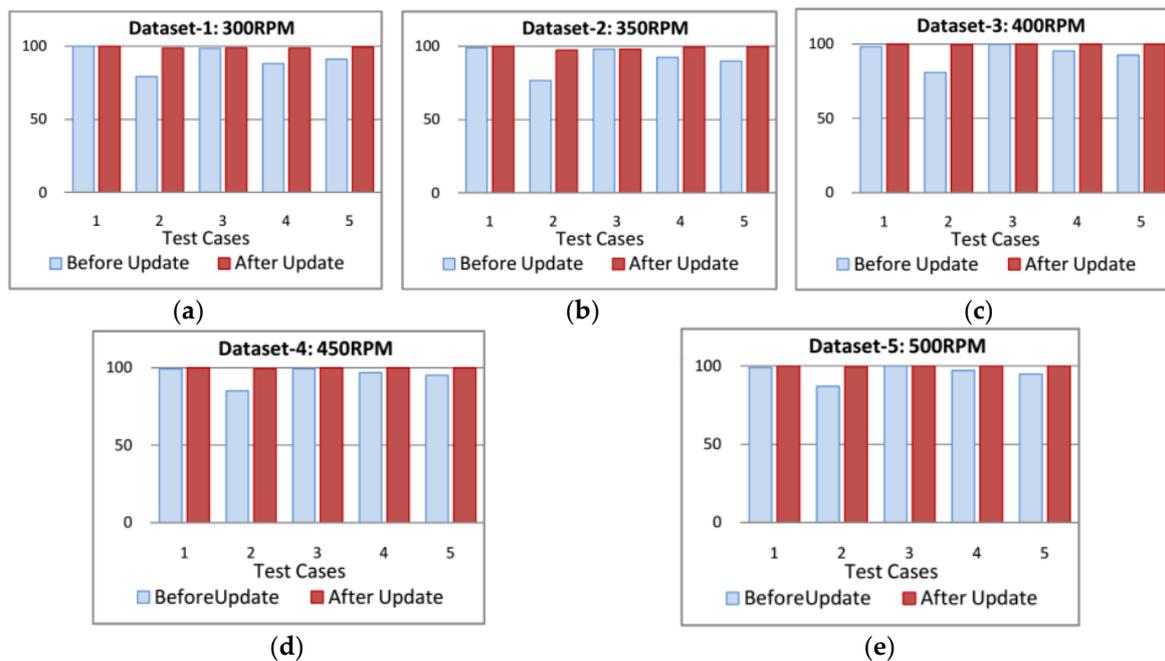
where  $N_{TP}$  is the number of correctly classified samples in a fault class,  $N_{FN}$  is the number of incorrectly classified samples in a fault class, and  $N_{samples}$  is the total number of test samples.

Figure 14 shows the average classification performance for different datasets. The proposed method detects the unknown faults well and updates the DSK (i.e., fault modes and optimum fault signatures). The updated DSK then provides improved classification performance. For each test case, when new faults are introduced, the classification performance initially degrades because the system does not have enough information about the new faults. However, the performance improves once the DSK is updated with information on the new fault modes. The classification accuracy using the proposed method is high for test case 1 because it involves only two well-distributed classes. For test case 3, the classification accuracy of the proposed method is almost 100% because the unknown data contains no new faults. Otherwise, the classification performance decreases with an increase in new faults (small or big cracks). However, after updating the DSK, the classification performance improves by about 21%. Furthermore, classification accuracies for more severe faults (e.g., bigger cracks) are better than those for less severe faults (e.g., smaller cracks) because more severe faults contain more separable features. According to the experiment, the proposed model can detect the newly accounted fault in data. However, in a real monitoring environment like industry, the quality of bearing degrades very slowly. According to the suggestion of the American Bearing Manufacturing Association (ABMA) and the American Rolling Bearing Company, the minimum L10 life of a continuous 24 h operated bearing is 60,000 h, which is around seven years. The proposed model might be deployed in the real industry environment, where a chunk of data will be captured a certain interval, i.e., once a day, once a week, once a month, or even more. The proposed automatic diagnosis model will analyze the collected data and identify the new fault or any abnormality of the bearing if it exists. The newly detected fault mode may incorporate the database of the diagnosis model. The system may run up to a tolerant level of quality of bearing.









**Figure 14.** Diagnosis performance improvement for different test cases with different conditional datasets, i.e., (a) Dataset-1: 300 RPM, (b) Dataset-2: 350 RPM, (c) Dataset-3: 400 RPM, (d) Dataset-4: 450 RPM, and (e) Dataset-5: 500 RPM. The y-axis represents average classification accuracy (ACA) (%).

## 5. Conclusions

Reliable and early fault diagnosis of the bearing may reduce the loss of manufacturing production by decreasing the possibility of an unwanted breakdown of the industrial motor. The traditional fault diagnosis model worked on the predefined and incomplete knowledge of faults of a bearing. However, due to the progressive nature of the growing fault during the operation under heavy load, new faults may appear on the surface of the bearing. To address these issues, this paper proposed a reliable and online fault diagnosis system for detecting unknown fault modes in the operating life of the bearing and update the diagnosis system knowledge (DSK) in real time. The heterogeneous feature vector is extracted from the collected AE signal from the bearing for training the diagnosis model. To detect the unknown number of faults, the  $k$ -means unsupervised clustering algorithm is used by finding the optimal number of clusters. However, determining the correct number of clusters  $k_{opt}$  is challenging due to the dynamic behavior of faults. For this reason, a new MPDFCDF cluster evaluation method is introduced to calculate  $k_{opt}$  and establish the existence of new fault modes. When a new fault mode is detected, the system automatically updates the DSK. Finally, the optimal features selection method is used for selecting optimal features from the updated fault database and the  $k$ -NN classification model is adopted for identifying faults in the unknown signals. For evaluating the proposed model, sensor signals with multiple faults, different fault severities, and different rotation speeds were considered. According to the experimental results, the proposed model showed much better diagnosis performance after incorporating the detected new fault. In addition, the proposed MPDFCDF algorithm outperformed existing cluster evaluation methods in terms of finding the optimal number of fault modes in data.

**Author Contributions:** All authors contributed equally to the conception of the idea, as well as implementing and analyzing the experimental results, and writing the manuscript.

**Funding:** This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of Korea (No. 20181510102160), in part by the Ministry of Trade, Industry & Energy (MOTIE) of Korea and Korea Institute for Advancement of Technology (KIAT) through the Encouragement Program for The Industries of Economic Cooperation Region. (P0006123).

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

## References

1. Zhang, P.; Du, Y.; Habetler, T.G.; Lu, B. A survey of condition monitoring and protection methods for medium-voltage induction motors. *IEEE Trans. Ind. Appl.* **2011**, *47*, 34–46. [[CrossRef](#)]
2. Kang, M.; Kim, J.; Kim, J.-M.; Tan, A.C.C.; Kim, E.Y.; Choi, B.-K. Reliable Fault Diagnosis for Low-Speed Bearings Using Individually Trained Support Vector Machines with Kernel Discriminative Feature Analysis. *IEEE Trans. Power Electron.* **2015**, *30*, 2786–2797. [[CrossRef](#)]
3. Xu, Y.; Zhang, K.; Ma, C.; Cui, L.; Tian, W. Adaptive Kurtogram and its applications in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *130*, 87–107. [[CrossRef](#)]
4. Kang, M.; Islam, M.R.; Kim, J.; Kim, J.; Pecht, M. A Hybrid Feature Selection Scheme for Reducing Diagnostic Performance Deterioration Caused by Outliers in Data-Driven Diagnostics. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3299–3310. [[CrossRef](#)]
5. Seshadrinath, J.; Singh, B.; Panigrahi, B.K. Vibration analysis based interturn fault diagnosis in induction machines. *IEEE Trans. Ind. Inform.* **2014**, *10*, 340–350. [[CrossRef](#)]
6. Zhou, W.; Lu, B.; Habetler, T.G.; Harley, R.G. Incipient Bearing Fault Detection via Motor Stator Current Noise Cancellation Using Wiener Filter. *IEEE Trans. Ind. Appl.* **2009**, *45*, 1309–1317. [[CrossRef](#)]
7. Zhou, W.; Habetler, T.G.; Harley, R.G. Bearing Fault Detection via Stator Current Noise Cancellation and Statistical Control. *IEEE Trans. Ind. Electron.* **2008**, *55*, 4260–4269. [[CrossRef](#)]
8. Berry, J.E. How to Track Rolling Element Bearing Health with Vibration Signature Analysis. *Sound Vib.* **1991**, *11*, 24–35.
9. Jiang, F.; Zhu, Z.; Li, W.; Ren, Y.; Zhou, G.; Chang, Y. A Fusion Feature Extraction Method Using EEMD and Correlation Coefficient Analysis for Bearing Fault Diagnosis. *Appl. Sci.* **2018**, *8*, 1621. [[CrossRef](#)]
10. Seshadrinath, J.; Singh, B.; Panigrahi, B.K. Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets. *IEEE Trans. Power Electron.* **2014**, *29*, 936–945. [[CrossRef](#)]
11. Pandya, D.H.; Upadhyay, S.H.; Harsha, S.P. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. *Expert Syst. Appl.* **2013**, *40*, 4137–4145. [[CrossRef](#)]
12. Niknam, S.A.; Songmene, V.; Au, Y.H.J. The Use of Acoustic Emission Information to Distinguish Between Dry and Lubricated Rolling Element Bearings in Low-Speed Rotating Machines. *Int. J. Adv. Manuf. Technol.* **2013**, *69*, 2679–2689. [[CrossRef](#)]
13. Eftekharijad, B.; Carrasco, M.R.; Charnley, B.; Mba, D. The Application of Spectral Kurtosis on Acoustic Emission and Vibrations from a Defective Bearing. *Mech. Syst. Signal Process.* **2011**, *25*, 266–284. [[CrossRef](#)]
14. Rauber, T.W.; Boldt, F.A.; Varejao, F.M. Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 637–646. [[CrossRef](#)]
15. Islam, R.; Khan, S.A.; Kim, J.-M. Discriminant Feature Distribution Analysis-Based Hybrid Feature Selection for Online Bearing Fault Diagnosis in Induction Motors. *J. Sens.* **2016**, *2016*, 1–16. [[CrossRef](#)]
16. Yu, K.; Lin, T.R.; Tan, J.; Ma, H. An adaptive sensitive frequency band selection method for empirical wavelet transform and its application in bearing fault diagnosis. *Measurement* **2019**, *134*, 375–384. [[CrossRef](#)]
17. Yin, G.; Zhang, Y.-T.; Li, Z.-N.; Ren, G.-Q.; Fan, H.-B. Online fault diagnosis method based on Incremental Support Vector Data Description and Extreme Learning Machine with incremental output structure. *Neurocomputing* **2014**, *128*, 224–231. [[CrossRef](#)]
18. Jiang, W.; Zhou, J.; Liu, H.; Shan, Y. A multi-step progressive fault diagnosis method for rolling element bearing based on energy entropy theory and hybrid ensemble auto-encoder. *ISA Trans.* **2019**, *87*, 235–250. [[CrossRef](#)] [[PubMed](#)]
19. Yiakopoulos, C.T.; Gryllias, K.C.; Antoniadis, I.A. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Syst. Appl.* **2011**, *38*, 2888–2911. [[CrossRef](#)]
20. Khan, F. An initial seed selection algorithm for k-means clustering of geo referenced data to improve replicability of cluster assignments for mapping application. *Appl. Soft Comput.* **2012**, *12*, 3698–3700. [[CrossRef](#)]
21. Bradley, P.S.; Fayyad, U.M. Refining initial points for K-means clustering. In Proceedings of the Fifteenth International Conference on Machine Learning, San Francisco, CA, USA, 24–27 July 1998; pp. 91–99.

22. Likas, A.; Vlassis, N.; Verbeek, J.J. The global K-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461. [[CrossRef](#)]
23. Arthur, D.; Vassilvitskii, S. K-means ++: The Advantages of Careful Seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.
24. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005.
25. Tan, P.-N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Pearson Addison Wesley: Boston, MA, USA, 2005.
26. Rahman, M.A.; Islam, M.Z. A hybrid clustering technique combining a novel genetic algorithm with K-Means. *Knowl.-Based Syst.* **2014**, *71*, 345–365. [[CrossRef](#)]
27. Kang, M.; Kim, J.; Wills, L.M.; Kim, J.-M. Time-Varying and Multi resolution Envelope Analysis and Discriminative Feature Analysis for Bearing Fault Diagnosis. *IEEE Trans. Power Electron.* **2015**, *62*, 7749–7761.
28. WS Sensor, General Purpose Wideband Sensor. Available online: [http://www.physicalacoustics.com/content/literature/sensors/Model\\_WSa.pdf](http://www.physicalacoustics.com/content/literature/sensors/Model_WSa.pdf) (accessed on 20 May 2019).
29. Bediaga, I.; Mendizabal, X.; Arnaiz, A.; Munoa, J. Ball Bearing Damage Detection Using Traditional Signal Processing Algorithms. *IEEE Instrum. Meas. Mag.* **2013**, *16*, 20–25. [[CrossRef](#)]
30. Randall, R.B.; Antoni, J. Rolling Element Bearing Diagnostics—A Tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [[CrossRef](#)]
31. Li, B.; Zhang, P.-L.; Tian, H.; Mi, S.-S.; Liu, D.-S.; Ren, G.-Q. A New Feature Extraction and Selection Scheme for Hybrid Fault Diagnosis of Gearbox. *Expert Syst. Appl.* **2011**, *38*, 10000–10009. [[CrossRef](#)]
32. Liu, C.; Jiang, D.; Yang, W. Global Geometric Similarity Scheme for Feature Selection in Fault Diagnosis. *Expert Syst. Appl.* **2014**, *41*, 3585–3595. [[CrossRef](#)]
33. Li, Z.; Yan, X.; Tian, Z.; Yuan, C.; Peng, Z.; Li, L. Blind Vibration Component Separation and Nonlinear Feature Extraction Applied to the Non stationary Vibration Signals for the Gearbox Multi-Fault Diagnosis. *Measurement* **2013**, *46*, 259–271. [[CrossRef](#)]
34. Islam, R.; Khan, S.A.; Kim, J.-M. Maximum class separability-based discriminant feature selection using a GA for reliable fault diagnosis of induction motors. *Lect. Notes Artif. Intell. (LNAI)* **2015**, *9227*, 526–537.
35. Steinley, D. K-means clustering: A half-century synthesis. *Br. J. Math. Stat. Psychol.* **2006**, *9*, 1–34. [[CrossRef](#)]
36. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
37. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *8*, 129–137. [[CrossRef](#)]
38. Naldi, M.C.; Campello, R.J.G.B. Comparison of distributed evolutionary k-means clustering algorithms. *Neurocomputing* **2015**, *163*, 78–93. [[CrossRef](#)]
39. Zhang, Y.; Mańdziuk, J.; Quek, C.H.; Goh, B.W. Curvature-based method for determining the number of clusters. *Inf. Sci.* **2017**, *415–416*, 414–428. [[CrossRef](#)]
40. Yahyaoui, H.; Own, H.S. Unsupervised clustering of service performance behaviors. *Inf. Sci.* **2018**, *422*, 558–571. [[CrossRef](#)]
41. Kazmier, L.J. *Schaum's Outline of Business Statistics*; McGraw Hill Professional: New York, NY, USA, 2009; p. 359.
42. Yigit, H. A weighting approach for KNN classifier. In Proceedings of the 2013 International Conference on Electronics, Computer and Computation (ICECCO), Ankara, Turkey, 7–9 November 2013; pp. 228–231.

