

Article

Development of Indicator of Data Sufficiency for Feature-based Early Time Series Classification with Applications of Bearing Fault Diagnosis

Gilseung Ahn , Hwanchul Lee, Jisu Park and Sun Hur * 

Department of Industrial and Management Engineering, Hanyang University, Ansan 15588, Korea; ahn.kilseung@gmail.com (G.A.); hwanchul@hanyang.ac.kr (H.L.); jisu6922@naver.com (J.P.)

* Correspondence: hursun@hanyang.ac.kr; Tel.: +82-31-400-5265

Received: 11 May 2020; Accepted: 3 July 2020; Published: 6 July 2020



Abstract: Diagnosis of bearing faults is crucial in various industries. Time series classification (TSC) assigns each time series to one of a set of pre-defined classes, such as normal and fault, and has been regarded as an appropriate approach for bearing fault diagnosis. Considering late and inaccurate fault diagnosis may have a significant impact on maintenance costs, it is important to classify bearing signals as early and accurately as possible. TSC, however, has a major limitation, which is that a time series cannot be classified until the entire series is collected, implying that a fault cannot be diagnosed using TSC in advance. Therefore, it is important to classify a partially collected time series for early time series classification (ESTC), which is a TSC that considers both accuracy and earliness. Feature-based TSCs can handle this, but the problem is to determine whether a partially collected time series is enough for a decision that is still unsolved. Motivated by this, we propose an indicator of data sufficiency to determine whether a feature-based fault detection classifier can start classifying partially collected signals in order to diagnose bearing faults as early and accurately as possible. The indicator is trained based on the cosine similarity between signals that were collected fully and partially as input to the classifier. In addition, a parameter setting method for efficiently training the indicator is also proposed. The results of experiments using four benchmark datasets verified that the proposed indicator increased both accuracy and earliness compared with the previous time series classification method and general time series classification.

Keywords: early time series classification; data sufficiency; bearing fault diagnosis; feature-based classification

1. Introduction

Bearings are one of the important components in rotary machines such as motors, wind turbines, helicopters, automobiles, and gearboxes [1]. The fault diagnosis of bearings is a crucial task because they are among the most important components of rotation machines; faulty bearings are one of the main causes of machine failure [2]. Consequently, predictive maintenance methods for bearings have attracted interest from both academia and industry. Jin et al. [3] developed a health index based on the bearing vibration signal and designed a method to detect bearing faults by selecting the appropriate threshold with a Box-Cox transformation. Singleton et al. [4] introduced a data-driven methodology, which relies on both time and time–frequency domain features to track the evolution of bearing faults. Kumar et al. [5] developed a health index using singular value decomposition, the average value of the cumulative feature, and Mahalanobis distance to evaluate and compared the four conditions of the bearing. Caesarendra and Tjahjowidodo [6] confirmed the change of the low-speed slew bearing condition from normal to failure using impulse factor, margin factor, approximate entropy, and largest

Lyapunov exponent (LLE). Li et al. [7] proposed an approach for motor rolling bearing fault diagnosis using neural networks and time–frequency-domain bearing vibration analysis.

Time series classification (TSC) is a supervised learning task that assigns each time series instance to a predefined class, such as fault and normal [8]. In other words, TSC aims to train and use the classifier f to diagnose with time series x_i for the bearing i as $\hat{y}_i = f(x_i)$ where \hat{y}_i is predicted fault status by the time series classifier f . Because each instance may have a different length, feature extraction is regarded as an essential step for the task when a typical classifier, other than RNN (recurrent neural network)-based ones such as long short-term memory (LSTM), is employed. Of course, RNN-based models can classify unequal time series but they may be improper for the early time series classification (ETSC) task due to their expensive computational costs.

TSC that includes feature extraction is called feature-based TSC and has been frequently used in various area including biomedical [9,10], manufacturing [11–13], and so forth.

Many studies have been conducted on feature-based TSCs for bearing fault diagnosis based on vibration signals. For example, Wu et al. [14] extracted features from vibration signals using multiscale permutation entropy and trained a support vector machine (SVM) for fault diagnosis. Goyal et al. [15] extracted statistical features, such as mean, standard deviation, root mean square, and skewness from a vibration signal collected by a noncontact sensor. An accelerometer then selected several features based on the Mahalanobis distance for training the SVM. They observed that a noncontact sensor can be applied to identify bearing faults, and that a linear SVM outperformed other SVMs. Gunerkar et al. [16] employed wavelet transform to extract time domain features from vibration signals and trained supervised models, including an artificial neural network (ANN) and a k-nearest neighbor algorithm. From their experiment, they observed that the ANN outperformed other models in terms of accuracy. In recent years, convolutional neural networks (CNNs) have frequently been applied to solve bearing fault diagnosis problems, because they have filters to extract features from images. For example, Zhao et al. [17] proposed a planet-bearing fault classification method based on synchrosqueezing transform and CNN. In this method, the vibration signal was converted into a time–frequency color map using a synchrosqueezing transform. Then, the map was input into the CNN with six convolution layers and four max pooling layers, which assigned the map to one of three predefined classes: inner race fault, outer race fault, and health.

In time-sensitive applications, such as fault detection, earliness is as important as accuracy because late fault diagnosis leads to delayed maintenance and can cause the bearing fault to become permanent despite the accurate diagnosis. Even a few seconds of delay can lead to a critical situation such as an engineering system breakdown. Earliness is a measure to determine how early a classifier begins the classification job and is computed by the average ratio of the time until the classifier starts the classification to the time to collect the full time series.

TSC that considers both accuracy and earliness is called early TSC (ETSC) [18]. A few studies have proposed ETSC methods. For example, Hatami and Chira [19] developed an ensemble for early classification consisting of two classifiers with reject option (CWRO), which determines whether it can classify a (partially) collected instance. A CWRO does not classify an instance if its maximum posterior probability (i.e., $\max_k \Pr(y = c_k | x)$) or maximum decision function value is below a certain threshold. An instance is classified by the ensemble when every CWRO in the ensemble does not reject classification. The major limitations of the ensemble are as follows: First, it is difficult to determine the threshold of the reject option, especially for decision function values that are not probabilities. Second, the ensemble can reject an instance when it is hard to classify, even though it is fully collected, resulting in lower earliness. He et al. [20] proposed a shapelet-based early classification method for multivariate time series. The method extracts a set of shapelet candidates, conducts clustering of the candidates, and selects a core shapelet from each cluster based on the weighted mean of the accuracy and earliness for each class. A new time series instance is assigned to a class as long as the number of core shapelets for the class becomes a threshold or is randomly labeled. This method is very expensive in terms of computational complexity, and its performance, including accuracy and earliness, is highly dependent

on the set of shapelet candidates. That is, if the candidate set is inappropriately constructed owing to missing values and ill-defined parameters, the classification performance may be poor. Xing et al. [21] proposed a method to extract interpretable features from time series for interpretability of ETSC in medical and health informatics, industry production management, safety, and security management. Ghalwash and Obradovic [22] presented multivariate shapelet detection (MSD) that extracts time series patterns from all dimensions of the time series that distinctly manifest the target class locally, and the time series were classified by searching for the earliest closest patterns. Mori et al. [23] presented a method for early classification based on combining a set of probabilistic classifiers together with a stopping rule (SR), which acts as a trigger to indicate when to output a prediction or when to wait for more data.

Even though both starting time and processing time impact the earliness, advancing the starting time is a more reasonable option because processing time is very difficult to be reduced and it is usually small and almost the same for every time series. In order to advance the starting time, it is important to decide whether the (partially) collected time series is enough for classification or not. In other words, decisions on data sufficiency should be made periodically, and one can start classification once data is decided as enough for classification. However, to the best of our knowledge, there is no previous research that has addressed the question, “is the time series long enough for classification?” Only a few studies addressed questions such as “is the financial time series long enough for clustering?” [24] and “is the time series long enough for identifying the qualitative changes?” [25].

This paper proposes a feature-based early classification method for bearing fault diagnosis with a data sufficiency indicator. This indicator determines whether a given partially collected signal is sufficiently long to be classified by a fault diagnosis classifier based on its similarity to a fully collected signal. If the indicator determines that it is sufficiently long, the classifier begins classifying the signal without further collection. The indicator does not have any risk of reject classifying an instance, which is a common problem in previous methods because it is not based on the shapelet or distance but on statistical features, and it predicts the bearing fault classifier decision rather than the actual status of the bearing. The remainder of this paper is organized as follows: Section 2 formalizes the early bearing fault diagnosis problem and develops a solution to the problem. Section 3 proposes a data sufficiency indicator for a time series classifier and explains how to use and train the proposed indicator in detail. Section 4 conducts experiments to demonstrate that the proposed indicator can increase both accuracy and earliness compared with previous methods. Section 5 concludes this paper and suggests future research directions.

2. Early Bearing Fault Diagnosis

Bearing fault diagnosis using TSC is based on assigning time series from a bearing (e.g., vibration signal) to one of several pre-defined statuses (e.g., normal, inner race fault, outer race fault, or ball fault). More formally, let $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,T_i})$ or $x_{i,1:T_i}$ be the time series and y_i be the fault status of bearing i ($i = 1, 2, \dots, n$), where $x_{i,t}$ is the signal collected at time t for the bearing. Then, the problem aims to train the classifier f to diagnose the status of bearing i as $\hat{y}_i = f(x_i)$. However, it is difficult to develop a classifier with a raw signal (i.e., x_i), because the length of each bearing signal can differ from one another and be too long to train the classifier efficiently, and it even may not have significant features. For this reason, feature-based TSCs have been employed in many studies [9–13,26–28]. The usual process of developing a feature-based time series classifier is depicted in Figure 1.

A time series classifier with features can be expressed as follows:

$$\hat{y}_i = f(\Phi(x_i)) = f(\varphi_1(x_i), \varphi_2(x_i), \dots, \varphi_m(x_i)) \quad (1)$$

where φ_k denotes the feature function k ($k = 1, 2, \dots, m$). The feature functions used in this study were adopted from [26] and are listed in Table 1. These features have been frequently used for bearing diagnosis problems, because they summarize bearing signals very well. For instance, crest factor

indicates wear or cavitation, and root mean square shows the severity of bearing faults [29]. Readers can refer to [26,29] for more information on the nature of each feature function.

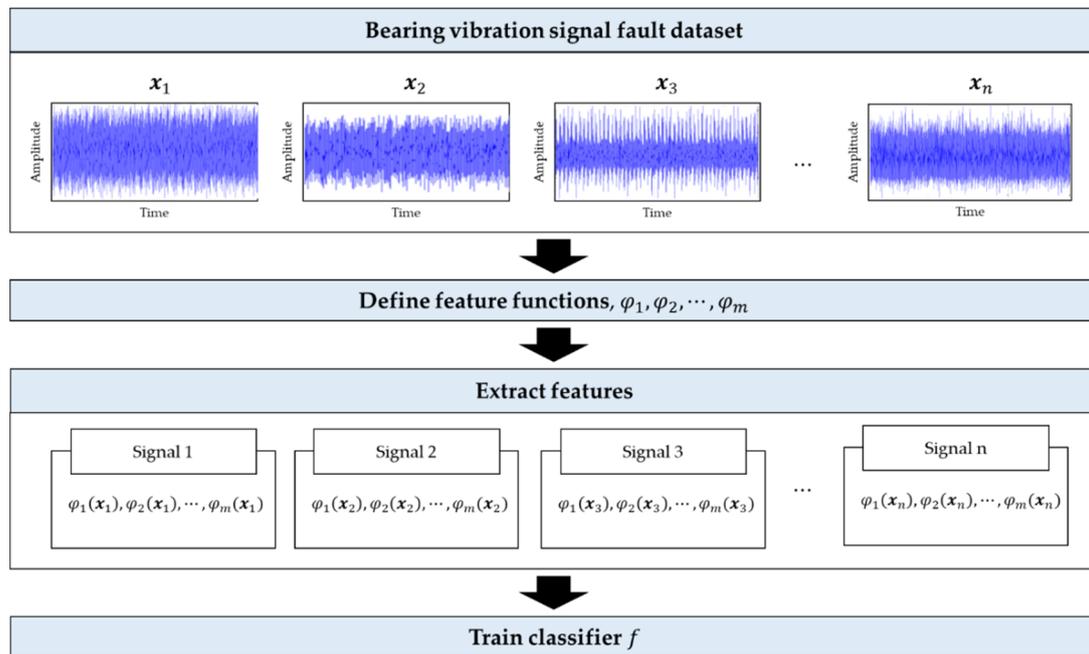


Figure 1. Development process of feature-based time series classifier.

Table 1. Feature functions used in this study.

Domain	Feature Function	Formula
Time domain	Mean	$\varphi_{mean} = \frac{\sum_{t=1}^{T_i} x_{i,t}}{T_i}$
	Standard deviation	$\varphi_{std} = \sqrt{\frac{\sum_{t=1}^{T_i} (x_{i,t} - \varphi_{mean})^2}{T_i}}$
	Root mean square	$\varphi_{rms} = \sqrt{\frac{\sum_{t=1}^{T_i} x_{i,t}^2}{T_i}}$
	Peak	$\varphi_{peak} = \max(x_i)$
	Shape factor	$\varphi_{sf} = \frac{\varphi_{rms}}{ \varphi_{mean} }$
	Crest factor	$\varphi_{cf} = \frac{\varphi_{peak}}{\varphi_{rms}}$
	Impulse factor	$\varphi_{if} = \frac{\varphi_{peak}}{ \varphi_{mean} }$
	Clearance factor	$\varphi_{cf} = \frac{\sum_{t=1}^{T_i} \sqrt{ x_{i,t} } / T_i}{\varphi_{peak}}$
	Skewness	$\varphi_{sk} = \frac{\sum_{t=1}^{T_i} x_{i,t}^3 / T_i}{\varphi_{rms}^3}$
	Kurtosis	$\varphi_{ku} = \frac{\sum_{t=1}^{T_i} x_{i,t}^4 / T_i}{\varphi_{rms}^4}$
Frequency domain	Mean frequency	$\varphi_{mf} = \frac{\sum_{r=1}^l s_r}{r}$
	Center frequency	$\varphi_{cf} = \frac{\sum_{r=1}^l e_r \times s_r}{\sum_{r=1}^l s_r}$
	Root mean square frequency	$\varphi_{rmsf} = \sqrt{\frac{\sum_{r=1}^l e_r^2 \times s_r^2}{\sum_{r=1}^l s_r}}$
	Standard deviation frequency	$\varphi_{sdf} = \sqrt{\frac{\sum_{r=1}^l (e_r - \varphi_{cf})^2 \times s_r}{\sum_{r=1}^l s_r}}$

In Table 1, e_r and s_r indicate the frequency and power spectrum, respectively, of the r^{th} spectrum line resulting from the estimation of the power spectral density of a signal.

As mentioned before, a bearing fault diagnosis task requires not only accuracy but also earliness. The accuracy and earliness of a classifier indicate how well it classifies instances and how early it can start and complete the classification, respectively. In other words, the signal should be classified as accurately and as early as possible. TSC considering earliness additionally is called early time series classification (ETSC). In order to classify a time series as soon as possible (i.e., for ETSC), it is important to start and complete the accurate classification early [8], as illustrated in Figure 2. As seen, ETSC starts (at τ) and finishes the classification earlier than general time series classification (GTSC) starts (at T) and finishes. In order to classify a time series as soon as possible (i.e., for ETSC), it is important to start and complete the accurate classification early [8], as illustrated in Figure 2. As seen, ETSC starts (at τ) and finishes the classification earlier than GTSC starts (at T) and finishes. In order to start earlier, a classifier should decide whether a bearing is a fault or not with partially collected time series $x_{i, 1:\tau} = (x_{i,1}, x_{i,2}, \dots, x_{i,\tau})$. A classifier with a shorter classification time should be used to reduce the processing time. A feature-based classifier requires relatively smaller classification time and the feature values (e.g., mean, standard deviation, center frequency, etc.) do not change significantly once a sufficient amount of signal is collected. However, the problem to determine whether the partially collected time series is enough for decision on the fault still needs to be solved.

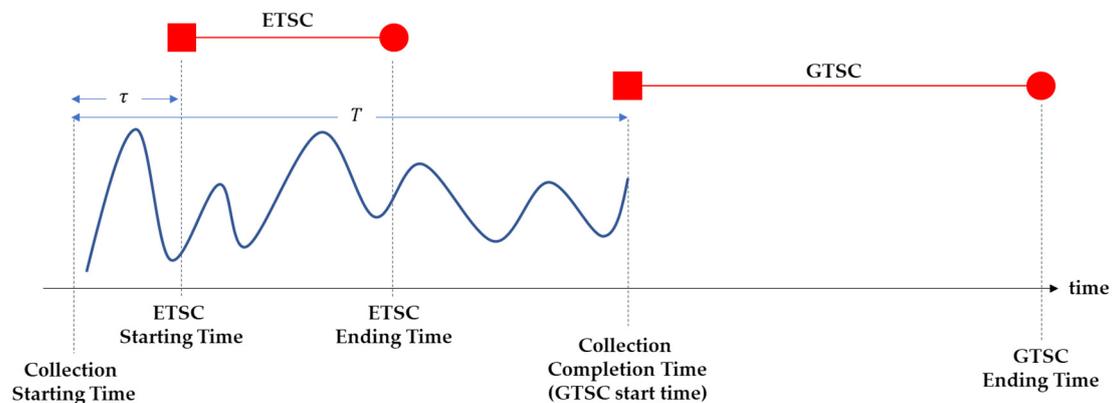


Figure 2. Comparison of general time series classification and early time series classification.

The main problems considered in this study are whether the collected signal is sufficiently long to be classified by an early classifier, and when the classifier can start the classification. In other words, the problems are to determine whether $f(\Phi(x_{i, 1:\tau_i})) = f(\Phi(x_{i,1}, x_{i,2}, \dots, x_{i,\tau_i}))$ and $f(\Phi(x_i)) = f(\Phi(x_{i,1}, x_{i,2}, \dots, x_{i,T_i}))$ are sufficiently similar that the classifier can begin classification, and to estimate the minimum time $\hat{\tau}_i$ such that $\text{SIM}_f(x_{i, 1:\hat{\tau}_i}, x_i) \geq \alpha$, where $\text{SIM}_f(A, B)$ is the similarity between A and B as input for f , and α is a threshold.

3. Proposed Indicator

As explained in Section 2, it is important to decide whether the partially collected time series is enough for a classifier. In this study, we propose an indicator for the decision problem. The indicator is also a classifier and trained based on a bearing fault dataset, which is also used to train a classifier f , and thus it makes a decision quickly and accurately.

This section describes the proposed indicator in detail, focusing on its application to ETSC. Then, we explain how the indicator determines whether the partially collected signal is sufficiently long for classification based on the similarity between the partially and fully collected signals. Finally, we explain how the indicator is trained.

Let I_f be an indicator to determine whether the collected signal $x_{i,1:t}$ is sufficiently long for classification by f until t for bearing i , expressed as follows:

$$I_f(x_{i,1:t}) = \begin{cases} 1, & \text{if } x_{i,1:t} \text{ is sufficiently long for classification by } f, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where “ $x_{i,1:t}$ is sufficiently long for classification by f ” implies $\hat{y}_i = f(x_{i,1:t}) = f(x_{i,1:T})$, that is, the decisions of f for $x_{i,1:t}$ and fully collected signal $x_{i,1:T} = x_i$ are the same. Thus, one can start classifying the signal of bearing i with $x_{i,1:t}$ when $I_f(x_{i,1:t}) = 1$. Note that the decisions of f for $x_{i,1:t}$ and fully collected signal $x_{i,1:T} = x_i$ being the same does not guarantee a correct classification result (i.e., y_i and $f(x_i) = f(x_{i,1:t})$ may be different). The specific process is presented in Figure 3, where τ_0 and τ are the start time and period of the indicator, respectively. That is, $I_f(x_{i,1:\tau_0+z\times\tau})$ is calculated for $z = 0, 1, 2, \dots$ until it becomes 1, and the partially collected signal starts to be classified.

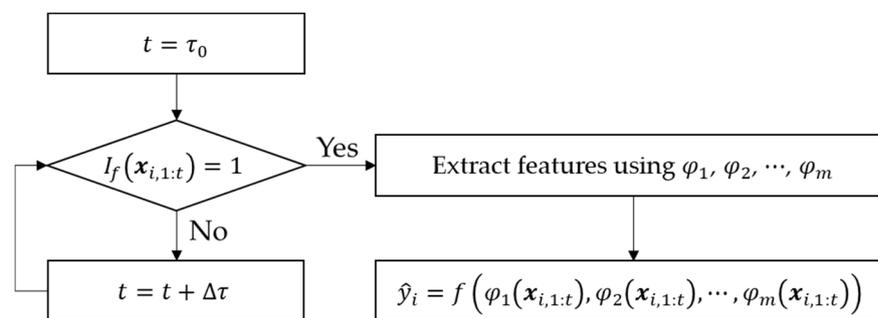


Figure 3. Early classification process using the proposed indicator and a classifier.

As mentioned above, the indicator $I_f(x_{i,1:t}) = 1$ when $x_{i,1:t}$ and $x_i (= x_{i,1:T_i})$ are similar to each other as input for f . In other words, $x_{i,1:t}$ is considered sufficiently long when the classification results of $x_{i,1:t}$ and $x_{i,1:T_i}$ by f are similar to each other. The similarity between $x_{i,1:\tau_i}$ and x_i as input of f , $SIM_f(x_{i,1:\tau_i}, x_i)$ is defined as the cosine similarity between two vectors $\delta(x_i) = (\delta_1(x_i), \dots, \delta_C(x_i))$ and $\delta(x_{i,1:\tau_i}) = (\delta_1(x_{i,1:\tau_i}), \dots, \delta_C(x_{i,1:\tau_i}))$ as follows:

$$SIM_f(x_{i,1:\tau_i}, x_i) = \frac{\sum_{c=1}^C \delta_c(x_i) \times \delta_c(x_{i,1:\tau_i})}{\sqrt{\sum_{c=1}^C \delta_c(x_i)^2} \times \sqrt{\sum_{c=1}^C \delta_c(x_{i,1:\tau_i})^2}} \quad (3)$$

where $\delta_c(x_i)$ is the decision function value of x_i for class c ($c \in \{1, 2, \dots, C\}$). The reason for using cosine similarity is that it is proper to express similarity between two vectors not based on their scales but on their directions [30], and direction is more important to measure the similarity between $\delta(x_i)$ and $\delta(x_{i,1:\tau_i})$.

For the decision function $\delta_c(x_i)$, one can use the hyperplane, $w_c x_i + b_c$, for class c if an SVM is adopted as the classifier. If the classifier is an ANN, then the output node c can play the role of a decision function, and $\Pr(y = c) \times \Pr(x_i | y = c)$ can be used if the naïve Bayes classifier is used. $\delta(x_i)$ and $\delta(x_{i,1:\tau_i})$ are used instead of the predicted classes $f(x_i)$ and $f(x_{i,1:\tau_i})$ to prevent the case where $f(x_i)$ and $f(x_{i,1:\tau_i})$ are coincidentally the same. Cosine similarity is adopted because it is appropriate to calculate directional similarity, and setting the similarity threshold is easy because its value is in $[-1, 1]$. It should be noted that the indicator, which is also regarded as a classifier, does not directly calculate $SIM_f(x_{i,1:\tau_i}, x_i)$ but predicts whether the similarity is greater than a threshold because it is used at time $\tau_i < T_i$ when $x_{i, \tau_i+1:T_i}$ is unknown.

Figure 4 shows the training process of I_f , which consists of four steps. It should be noted that the classifier f is also trained in parallel in the process.

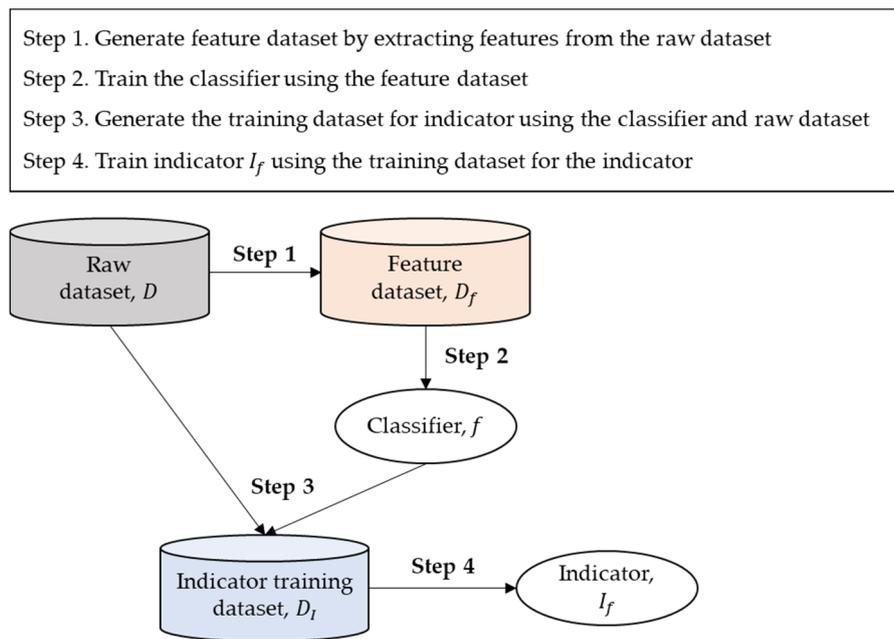


Figure 4. Learning framework for the proposed indicator.

As shown in Figure 4, the feature dataset $D_f = \{(\Phi(x_i), y_i) | i = 1, 2, \dots, n\}$ is generated by extracting features from the raw dataset $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$. The classifier f is trained with the feature dataset D_f . The indicator training dataset $D_I = \cup_{i=1}^n \{(x_{i,1:\tau}, \psi_{i,1:\tau}) | \tau = \tau_0, \tau_0 + \Delta\tau, \tau_0 + 2 \times \Delta\tau, \dots, T_i\}$, where $\psi_{i,1:\tau} = 1$ if $SIM_f(x_{i,1:t}, x_i)$ is equal to or greater than the threshold α ; otherwise, $\psi_{i,1:\tau} = 0$ and generated using the following algorithm. In this algorithm, τ_0 and $\Delta\tau$ denote the first time and period, respectively, to check if partially collected signals are enough for classification by the indicator. α , τ_0 , and $\Delta\tau$ are user-defined parameters, which impact on both training time and processing time of the proposed indicator. Specifically, the bigger α is, and the smaller τ_0 and $\Delta\tau$ are, the greater the number of iterations to train the indicator is possible, and, thus, the more accurate the indicator is expected to be.

Algorithm 1. Generation of the indicator training dataset.

Input	f, x_i , for $i = 1, 2, \dots, n, \alpha, \tau_0, \Delta\tau$
Procedure	<p>Step 1. Initialize $i = 1$ and $D_I = \emptyset$</p> <p>Step 2. Initialize $t = \tau_0$</p> <p>Step 3. Calculate $SIM_f(x_{i,1:t}, x_i)$ using Equation (3)</p> <p>Step 4. $\psi_{i,1:\tau} = 1$ if $SIM_f(x_{i,1:t}, x_i) \geq \alpha$, and $\psi_{i,1:\tau} = 0$ otherwise</p> <p>Step 5. $D_I = D_I \cup \{(x_{i,1:t}, \psi_{i,1:t})\}$</p> <p>Step 6. Increase t by $\Delta\tau$</p> <p>Step 7. If $t \geq T_i$ or $SIM_f(x_{i,1:t}, x_i) = 1$, increase i by 1 and go back to step 2.</p> <p>Step 8. If $i > n$, terminate the algorithm. Otherwise, go back to step 2</p>
Output	D_I

Finally, the indicator is trained with D_I . It should be noted that D_I is usually class-imbalanced (i.e., $I_f(x_{i,1:t}) = 0$ for most i and t); therefore, oversampling or undersampling may be necessary to solve the problem and train an unbiased indicator.

Clearly, the parameters τ_0 and $\Delta\tau$ have an impact on the effectiveness and efficiency of the proposed indicator. There are, however, no ground rules in setting the parameters. When training the indicator, there is no information about τ_0 and $\Delta\tau$. In this case, a tenth of the sampling frequency may be a good choice. When using the indicator, we can consider the first time of x_i , $\tau_{0,i}$ as “sufficiently long for classification by f ” (i.e., the smallest t satisfying $I_f(x_{i,1:t}) = 1$ in Step 3 of Algorithm 1). We suggest that τ_0 be set as $\min\{\tau_{0,1}, \tau_{0,2}, \dots, \tau_{0,n}\}$ because the indicator can find $\tau_{0,i}$ with the highest efficiency when $\tau_0 = \min\{\tau_{0,1}, \tau_{0,2}, \dots, \tau_{0,n}\}$. It should be noted that one cannot guarantee $\min\{\tau_{0,1}, \tau_{0,2}, \dots, \tau_{0,n}\} < \tau_{0,n'}$ for every $n' > n$ (i.e., the instance that is not in the training dataset), and $\min\{\tau_{0,1}, \tau_{0,2}, \dots, \tau_{0,n}\} - \tau_{0,n'}$ is the loss of decision time when $\min\{\tau_{0,1}, \tau_{0,2}, \dots, \tau_{0,n}\} > \tau_{0,n'}$. Similarly, we suggest that $\Delta\tau$ be set as $\min\{\tau_{0,i_1} - \tau_{0,i_2} | \tau_{0,i_1} \geq \tau_{0,i_2}\}$.

4. Experiment

4.1. Objective and Process

The objective of the experiment is to verify whether the proposed indicator is better than CWRO in increasing earliness without loss of accuracy. The specific processes using a dataset are as follows:

- Step 1. The dataset is randomly split into a training and test dataset for objective evaluation of the proposed indicator. Specifically, the set of indices $I = \{1, 2, \dots, n\}$ of time series instances is randomly separated to I_{Train} and I_{Test} with a ratio of 7:3. That is, 70% of samples is randomly selected whose indices are in I_{Train} and is used to train the model, and the remaining 30% of samples in I_{Test} is to test it.
- Step 2. A classifier and an indicator are trained using the training dataset $\{(x_i, y_i) | i \in I_{Train}\}$, as depicted in Figure 4. We selected ANN and SVM as a classifier, because they have been most frequently used as a feature-based time series classifiers in previous research, e.g., in [14–17]. Each classifier is trained by means of all features presented in Table 1, as was done in [26].
- Step 3. The trained classifier is tested using the test dataset, $\{(x_i, y_i) | i \in I_{Test}\}$, in terms of the micro f1-score. It is employed as an accuracy measure because it is a proper measure of multiclass classification, which may have a class imbalance problem. The micro f1-score, which is the harmonic mean of micro precision and recall, is calculated as follows:

$$\text{micro } F_1 = 2 \times \frac{\text{micro precision} \times \text{micro recall}}{\text{micro precision} + \text{micro recall}} \quad (4)$$

where micro precision and recall are calculated as follows:

$$\text{micro precision} = \frac{\sum_{c=1}^C TP_c}{\sum_{c=1}^C TP_c + FP_c} \quad (5)$$

$$\text{micro recall} = \frac{\sum_{c=1}^C TP_c}{\sum_{c=1}^C TP_c + FN_c} \quad (6)$$

where TP_c , FP_c , and FN_c indicate true positive, false positive, and false negative, respectively, when class c is regarded as positive.

- Step 4. The accuracy and earliness of the classifier trained with the proposed indicator are calculated using $\{(x_{i, 1:\tau_i}, y_i) | i \in I_{Test}\}$, where τ_i indicates the minimum value among $t \in \{\tau_0 + z \times \tau | z = 1, 2, \dots\}$ satisfying $I_f(x_{i,1:t}) = 1$ (i.e., $SIM(x_{i,1:t}, x_i) \geq \alpha$). Earliness is calculated as follows:

$$\text{earliness} = \frac{\sum_{i \in I_{Test}} \left(1 - \frac{\tau_i}{T_i}\right)}{|I_{Test}|} \quad (7)$$

- where τ_i denotes the classification start time for time series instance i' . For accurate and efficient decision for the indicator, our computational experience shows that α should be equal to or bigger than 0.9. τ_0 and $\Delta\tau$ can be determined as proposed in the final paragraph of Section 3.
- Step 5. The accuracy and earliness of the classifier trained with the CWRO approach are calculated using $\{(x_{i, 1:v_i}, y_i) | i \in I_{Test}\}$, where v_i indicates the maximum value among $t \in \{\tau_0 + z \times \tau\}$ satisfying the standard deviation of $\delta(x_i) \geq \varepsilon$.
- Step 6. Accuracy and earliness obtained from Steps 4 and 5 are compared.

4.2. Datasets

We collected four benchmark datasets of bearing vibration from an accelerometer from the existing literature. The dataset information is presented in Table 2 and refers to [31,32] for detail setting to collect the datasets. Note that because the sampling durations of the datasets are 10 or 40 s, one may think that ETSC is not necessary. However, the signal is continuously collected when bearing fault detection is applied in the real world; thus, the real sampling duration could be in hours, days, and even weeks.

Table 2. Experimental dataset information.

Dataset	Data Type	Sampling Frequency (Hz)	Sampling Duration (s)	Class Variable Distribution	Reference
Dataset #1	Vibration	200,000	10	Healthy: 12 Inner race fault: 12 Outer race fault: 12	
Dataset #2	Vibration	200,000	10	Healthy: 12 Inner race fault: 12 Outer race fault: 12 Ball fault: 12	[31]
Dataset #3	Rotational speed	200,000	10	Healthy: 12 Inner race fault: 12 Outer race fault: 12 Ball fault: 12	
Dataset #4	Vibration	12,000	40	Healthy: 5 Inner race fault: 16 Outer race fault: 16 Ball fault: 16	[32]

4.3. Results

Table 3 compares GTSC, CWRO, and the proposed model (a classifier with the proposed indicator) in terms of accuracy and earliness. As explained in Section 4.1, accuracy and earliness were measured using (4) and (7), respectively. In Table 3, each line denotes the accuracy and earliness of a classifier (SVM or ANN) for the dataset (#1–#4) when a preprocessing model for ETSC (GTSC, CWRO, or the proposed model) and its parameter (none for GTSC, ε for CWRO, and α for the proposed model) is applied. Numbers in boldface represent the best among the results of the given dataset and classifier.

Since GTSC does not consider earliness and uses the entire signal, earliness is zero. For dataset #1, the earliness of CWRO is zero except for the case where the classifier is SVM and ε is 0.5, implying that there is no clear difference between partially collected time series under normal and fault status, and thus CWRO is not appropriate for this kind of dataset. However, the proposed model shows not only high earliness but also higher accuracy than that of GTSC. For dataset #2–#4, CWRO shows non-zero earliness but small earliness and low accuracy. From the results, we observe the following: First, a general time series classifier does not always yield the best classification performance in terms of accuracy. This implies that using the fully collected time series does not guarantee higher accuracy; instead, dimension reduction techniques, including early classification, may increase accuracy. Second,

the proposed model outperforms CWRO and general time series classifiers in terms of both accuracy and earliness for all cases. In addition, the range of ϵ is $[0, \infty)$, but the range of α is $[-1, 1]$, implying that it is easier for the user to set the value of α than that of ϵ . Third, the accuracy and earliness highly depend on the dataset and classifier used. It is obvious that classification performance depends on a classifier and dataset for every classification problem, and the proposed indicator depends on the used classifier.

Table 3. Performance analysis result.

Dataset	Classifier	Model	Parameter	Accuracy	Earliness
Dataset #1	SVM	GTSC	None	0.6667	0.0000
		CWRO	$\epsilon = 0.5$	0.6667	0.9950
			$\epsilon = 1.0$	0.6667	0.0000
			$\epsilon = 1.5$	0.6667	0.0000
			$\epsilon = 2.0$	0.6667	0.0000
		Proposed model	$\alpha = 0.90$	0.6667	0.9950
	$\alpha = 0.95$		0.7778	0.9750	
	$\alpha = 0.99$		0.8889	0.9050	
	ANN	GTSC	None	0.6667	0.0000
		CWRO	$\epsilon = 0.5$	0.6667	0.0000
			$\epsilon = 1.0$	0.6667	0.0000
			$\epsilon = 1.5$	0.6667	0.0000
$\epsilon = 2.0$			0.6667	0.0000	
Proposed model		$\alpha = 0.90$	0.7778	0.9750	
	$\alpha = 0.95$	0.7778	0.9750		
	$\alpha = 0.99$	0.7778	0.9750		
Dataset #2	SVM	GTSC	None	0.5786	0.0000
		CWRO	$\epsilon = 0.5$	0.3333	0.9950
			$\epsilon = 1.0$	0.3333	0.9950
			$\epsilon = 1.5$	0.3333	0.9000
			$\epsilon = 2.0$	0.3333	0.9000
		Proposed model	$\alpha = 0.90$	0.5786	0.9950
	$\alpha = 0.95$		0.5786	0.9850	
	$\alpha = 0.99$		0.5786	0.9000	
	ANN	GTSC	None	0.5786	0.0000
		CWRO	$\epsilon = 0.5$	0.3333	0.9000
			$\epsilon = 1.0$	0.3333	0.9000
			$\epsilon = 1.5$	0.3333	0.9000
$\epsilon = 2.0$			0.3333	0.9000	
Proposed model		$\alpha = 0.90$	0.5786	0.9950	
	$\alpha = 0.95$	0.5786	0.9850		
	$\alpha = 0.99$	0.5786	0.9000		

Table 3. Cont.

Dataset	Classifier	Model	Parameter	Accuracy	Earliness
Dataset #3	SVM	GTSC	None	0.2500	0.0000
		CWRO	$\varepsilon = 0.5$	0.1333	0.9950
			$\varepsilon = 1.0$	0.1333	0.9950
			$\varepsilon = 1.5$	0.1333	0.9000
			$\varepsilon = 2.0$	0.1333	0.9000
		Proposed model	$\alpha = 0.90$	0.2500	0.9950
			$\alpha = 0.95$	0.2500	0.9950
	$\alpha = 0.99$		0.2500	0.9000	
	ANN	GTSC	None	0.2500	0.0000
		CWRO	$\varepsilon = 0.5$	0.1333	0.9000
			$\varepsilon = 1.0$	0.1333	0.9000
			$\varepsilon = 1.5$	0.1333	0.9000
			$\varepsilon = 2.0$	0.1333	0.9000
		Proposed model	$\alpha = 0.90$	0.2500	0.9000
$\alpha = 0.95$			0.2500	0.9000	
$\alpha = 0.99$	0.2500		0.9000		
Dataset #4	SVM	GTSC	None	0.5294	0.0000
		CWRO	$\varepsilon = 0.5$	0.7059	0.9792
			$\varepsilon = 1.0$	0.7059	0.9792
			$\varepsilon = 1.5$	0.4705	0.7015
			$\varepsilon = 2.0$	0.4705	0.7015
		Proposed model	$\alpha = 0.90$	0.7692	0.9413
			$\alpha = 0.95$	0.7692	0.9413
	$\alpha = 0.99$		0.7892	0.8544	
	ANN	GTSC	None	0.8235	0.0000
		CWRO	$\varepsilon = 0.5$	0.5294	0.7000
			$\varepsilon = 1.0$	0.5294	0.7000
			$\varepsilon = 1.5$	0.5294	0.7000
			$\varepsilon = 2.0$	0.5294	0.7000
		Proposed model	$\alpha = 0.90$	0.6956	0.8544
$\alpha = 0.95$			0.5714	0.8131	
$\alpha = 0.99$	0.5294		0.7000		

5. Conclusions

Bearing fault detection is one of the most important tasks in the manufacturing industry, which is often accomplished by TSC. Most previous studies focused on accuracy but failed to consider earliness. In time-sensitive applications such as bearing fault detection, earliness is a very important measure for a time series classifier because it is highly related to cost and safety. Although a few ETSC methods have been proposed, they are unsuitable for applications in fault detection problems because of reasons such as the difficulty of parameter setting, improper features for fault detection, and low accuracy.

In this paper, we proposed an early bearing fault diagnosis method based on a data sufficiency indicator. The indicator determines whether a signal collected within a specific period is sufficiently long to be classified by the fault diagnosis classifier. The experiment with benchmark datasets confirmed that the proposed method outperforms previous methods in terms of accuracy and earliness. Although this study focused on bearing fault diagnosis, the proposed indicator can also be applied to any type of ETSC problem.

There are two future research directions based on the limitations of the present study. First, we employed feature functions from previous studies. Although these are frequently used in research,

it remains uncertain whether they are effective for ETSC. Therefore, it is necessary to develop feature functions for early bearing fault diagnosis. Second, the proposed indicator was specially designed for the given classifier. In other words, the indicator is highly dependent on the classifier for bearing fault diagnosis, implying that it may not show good performance when another classifier is used. Therefore, the second research direction is to develop a robust indicator that is almost independent of a specific classifier and shows good results regardless of the classifier used. Third, we will modify and apply the proposed indicator for ETSC in a different area whose sampling rate is much longer than seconds. Fourth, selecting the τ_0 and τ_1 impact on the processing time of the proposed indicator, and thus we will develop a method to select the best values of them for each time series to reduce the processing time in future research. Finally, we will develop a hybrid model of the proposed model and CWRO for more efficient and effective ESTC.

Author Contributions: Conceptualization, G.A. and S.H.; methodology, G.A.; software, G.A. and J.P.; data curation, J.P.; original draft preparation, G.A. and H.L.; review and editing, G.A. and S.H.; funding acquisition, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (2019R1A2C1088255).

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

References

- Jin, X.; Zhao, M.; Chow, T.W.; Pecht, M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Trans. Ind. Electron.* **2013**, *61*, 2441–2451. [[CrossRef](#)]
- Gupta, P.; Pradhan, M.K. Fault detection analysis in rolling element bearing: A review. *Mater. Today* **2017**, *4*, 2085–2094. [[CrossRef](#)]
- Jin, X.; Sun, Y.; Que, Z.; Wang, Y.; Chow, T.W. Anomaly detection and fault prognosis for bearings. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 2046–2054. [[CrossRef](#)]
- Singleton, R.K.; Strangas, E.G.; Aviyente, S. Extended Kalman filtering for remaining-useful-life estimation of bearings. *IEEE Trans. Ind. Electron.* **2014**, *62*, 1781–1790. [[CrossRef](#)]
- Caesarendra, W.; Tjahjowidodo, T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines* **2017**, *5*, 21. [[CrossRef](#)]
- Li, B.; Chow, M.Y.; Tipsuwan, Y.; Hung, J.C. Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2000**, *47*, 1060–1069. [[CrossRef](#)]
- Kankar, P.K.; Sharma, S.C.; Harsha, S.P. Fault diagnosis of ball bearings using machine learning methods. *Expert. Syst. Appl.* **2011**, *38*, 1876–1886. [[CrossRef](#)]
- Baydogan, M.G.; Runger, G.; Tuv, E. A bag-of-features framework to classify time series. *IEEE Trans. Pattern. Anal.* **2013**, *35*, 2796–2802. [[CrossRef](#)]
- Povinelli, R.J.; Johnson, M.T.; Lindgren, A.C.; Ye, J. Time series classification using Gaussian mixture models of reconstructed phase spaces. *IEEE Trans. Knowl. Data. Eng.* **2004**, *16*, 779–783. [[CrossRef](#)]
- Pourbabae, B.; Roshtkhari, M.J.; Khorasani, K. Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Trans. Syst. Man. Cybern. Syst.* **2018**, *48*, 2095–2104. [[CrossRef](#)]
- Van Wyk, B.J.; Van Wyk, M.A.; Qi, G. Difference histograms: A new tool for time series analysis applied to bearing fault diagnosis. *Pattern. Recognit. Lett.* **2009**, *30*, 595–599. [[CrossRef](#)]
- Liu, C.L.; Hsaio, W.H.; Tu, Y.C. Time series classification with multivariate convolutional neural network. *IEEE Trans. Ind. Electron.* **2018**, *66*, 4788–4797. [[CrossRef](#)]
- Jeong, Y.S.; Jayaraman, R. Support vector-based algorithms with weighted dynamic time warping kernel function for time series classification. *Knowl Based. Syst.* **2015**, *75*, 184–191. [[CrossRef](#)]
- Wu, S.D.; Wu, P.H.; Wu, C.W.; Ding, J.J.; Wang, C.C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **2012**, *14*, 1343–1356. [[CrossRef](#)]
- Goyal, D.; Choudhary, A.; Pabla, B.; Dhama, S.S. Support vector machines based non-contact fault diagnosis system for bearings. *J. Intell. Manuf.* **2019**, *30*, 1–15. [[CrossRef](#)]

16. Gunerkar, R.S.; Jalan, A.K.; Belgamwar, S.U. Fault diagnosis of rolling element bearing based on artificial neural network. *J. Mech. Sci. Technol.* **2019**, *33*, 505–511. [[CrossRef](#)]
17. Zhao, D.; Wang, T.; Chu, F. Deep convolutional neural network based planet bearing fault classification. *Comput. Ind.* **2019**, *107*, 59–66. [[CrossRef](#)]
18. Ahn, G.; Hur, S. Efficient genetic algorithm for feature selection for early time series classification. *Comput. Ind. Eng.* **2020**, *142*, 106345. [[CrossRef](#)]
19. Hatami, N.; Chira, C. Classifiers with a reject option for early time-series classification. In Proceedings of the IEEE Symposium on Computational Intelligence and Ensemble Learning, Singapore, 16–19 April 2013.
20. He, G.; Duan, Y.; Peng, R.; Jing, X.; Qian, T.; Wang, L. Early classification on multivariate time series. *Neurocomputing* **2015**, *149*, 777–787. [[CrossRef](#)]
21. Xing, Z.; Pei, J.; Yu, P.S.; Wang, K. Extracting interpretable features for early classification on time series. In Proceedings of the SIAM International Conference on Data Mining, Mesa, AZ, USA, 28–30 April 2011.
22. Ghalwash, M.F.; Obradovic, Z. Early classification of multivariate temporal observations by extraction of interpretable shapelets. *BMC Bioinform.* **2012**, *13*, 195. [[CrossRef](#)]
23. Mori, U.; Mendiburu, A.; Dasgupta, S.; Lozano, J.A. Early classification of time series by simultaneously optimizing the accuracy and earliness. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 4569–4578. [[CrossRef](#)]
24. Marti, G.; Andler, S.; Nielsen, F.; Donnat, P. Clustering Financial Time Series: How Long Is Enough? In Proceedings of the Twenty-Fifth International Joint Conference on Artificial intelligence, New York, NY, USA, 9–15 July 2016.
25. Tran, Q.H.; Hasegawa, Y. Topological time-series analysis with delay-variant embedding. *Phys. Rev. E* **2019**, *99*, 032209.
26. Wei, Z.; Wang, Y.; He, S.; Bao, J. A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. *Knowl Based. Syst.* **2017**, *116*, 1–12. [[CrossRef](#)]
27. Rauber, T.W.; de Assis Boldt, F.; Varejão, F.M. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2014**, *62*, 637–646. [[CrossRef](#)]
28. Prieto, M.D.; Cirrincione, G.; Espinosa, A.G.; Ortega, J.A.; Henao, H. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *IEEE Trans. Ind. Electron.* **2012**, *60*, 3398–3407. [[CrossRef](#)]
29. Williams, T.; Ribadeneira, X.; Billington, S.; Kurfess, T. Rolling element bearing diagnostics in run-to-failure lifetime testing. *Mech. Syst. Signal. Pr.* **2001**, *15*, 979–993. [[CrossRef](#)]
30. Xia, P.; Zhang, L.; Li, F. Learning similarity with cosine similarity ensemble. *Inf. Sci.* **2015**, *307*, 39–52. [[CrossRef](#)]
31. Huang, H.; Baddour, N. Bearing vibration data collected under time-varying rotational speed conditions. *Data. Brief.* **2018**, *21*, 1745–1749. [[CrossRef](#)] [[PubMed](#)]
32. Staszewski, W.J.; Worden, K.; Tomlinson, G.R. Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition. *Mech. Syst. Signal. Pr.* **1997**, *11*, 673–692. [[CrossRef](#)]

