

Article

Rolling Bearings Fault Diagnosis Based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, Nonlinear Entropy, and Ensemble SVM

Rui Li ¹, Chao Ran ¹, Bin Zhang ^{2,3,*}, Leng Han ² and Song Feng ² 

¹ School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; lirui@cqupt.edu.cn (R.L.); ranchaoxp@163.com (C.R.)

² School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; hanleng@cqupt.edu.cn (L.H.); fengsong@cqupt.edu.cn (S.F.)

³ The State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing 400044, China

* Correspondence: zhangbin@cqupt.edu.cn

Received: 21 July 2020; Accepted: 7 August 2020; Published: 11 August 2020



Abstract: Rolling bearings are fundamental elements that play a crucial role in the functioning of rotating machines; thus, fault diagnosis of rolling bearings is of great significance to reduce catastrophic failures and heavy economic loss. However, the vibration signals of rolling bearings are often nonlinear and nonstationary, resulting in difficulty for feature extraction and fault recognition. In this paper, a hybrid method for multiple fault diagnosis of rolling bearings is presented. The bearing vibration signals are decomposed with the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) to denoise and extract nonlinear entropy features. The nonlinear entropy features are further processed to select the more discriminative fault features and to reduce feature dimension. Then a multi-class intelligent recognition model based on ensemble support vector machine (ESVM) is constructed to diagnose different bearing fault modes as well as fault severities. The effectiveness of the proposed method is assessed via experimental case studies of rolling bearings under multiple operational conditions (i.e., speeds and loads). The results show that our method gives better diagnosis results as compared to some existing approaches.

Keywords: fault diagnosis; rolling bearing; nonlinear entropy; improved complete ensemble empirical mode decomposition with adaptive noise; ensemble SVM; rotating machines

1. Introduction

Rolling bearings are fundamental components that are widely used in rotating machinery, for example, in airplanes, machining centers, and wind turbines. As a vulnerable link in these machines, rolling bearing faults can result in breakdown of the whole machinery. According to reference statistics, about 40% (for large machines) to about 90% (for small machines) of failures can be attributed to bearing faults [1]. Therefore, rolling bearing fault diagnosis has been the research hot topic in machine health monitoring over the past few decades.

The process of rolling bearing fault diagnosis basically comprises fault feature extraction and fault classification. To extract the fault features from vibration signals is the first and most essential step. Traditionally, bearing fault features are extracted with time-domain or frequency-domain analysis methods [2,3]. However, due to variations such as friction, damping, and operational conditions, bearing vibration signals often are nonlinear and nonstationary. Thus, bearing fault feature extraction is now more frequently analyzed through time-frequency/scale methods, including short-time Fourier

transform (STFT) [4], wavelet transform (WT) [5], and empirical mode decomposition (EMD). Without requirement of basis functions, EMD [6] is a data-driven method that adaptively decomposes a nonstationary and nonlinear time series into several approximately stationary time series defined as the intrinsic mode function (IMF) plus a final monotonic trend. In this way, EMD is complete and has been widely applied for fault feature extraction [7]. Nevertheless, EMD suffers problems of mode mixing, false mode, and endpoint effect. In [8], ensemble empirical mode decomposition (EEMD) was proposed to alleviate the mode aliasing problem of EMD by taking advantage of the property of frequency uniform distribution of Gaussian white noise. When analyzed with the EEMD method, the reconstructed signal containing residual noise and different realizations of signal plus noise may produce different numbers of modes, bringing difficulties to the final average operation. Then, complementary ensemble empirical mode decomposition (CEEMD) was proposed in [9] to improve the reconstruction error with EEMD by adding complementary pairs of Gaussian white noise, but the final averaging problem remained unsolved. This promoted the emergence of the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [10], achieving a negligible reconstruction error and solving the problem of different numbers of modes at the final average. Although CEEMDAN has been proven to be effective in fault feature extraction [11,12], it still has some aspects to be improved: (i) the presence of residual noise in the modes and (ii) the existence of spurious modes. To address these two issues, an improved CEEMDAN (ICEEMDAN) was further proposed and modes with less noise as well as more physical meaning can be obtained [13]. Although ICEEMDAN was originally proposed for biomedical signal processing [13,14], very recently few work [15] has researched its potential for fault diagnosis.

With the development of nonlinear dynamics theory, many nonlinear parameter estimation methods provide good alternatives to extracting the fault-related features hidden in vibration signals stemming from nonlinear systems. Among these parameters, entropy-based ones quantify the complexity of a time series and have shown very good performances in analyzing nonlinear vibration signals of mechanical systems. Recently, entropy and its variants attracted increasing attention in fault diagnosis [16]. The commonly used entropies are sample entropy (SampEn) [17,18] and permutation entropy (PerEn) [19–21]. Although SampEn is one improved version of approximate entropy, it still leads to unreliable estimations for short signals and is not fast enough for long signals in real-time applications. PerEn is much simpler to calculate and considerably faster than SampEn; however, it considers only the order relations of the signal amplitude values, so some information regarding the amplitude differences may be lost. To tackle these problems, Rostaghi and Azami [22,23] recently introduced one new nonlinear parameter, termed dispersion entropy (DisEn), to quantify the complexity of a time series. DisEn is relatively insensitive to noise and leads to more reliable estimations. Also, DisEn is simple and computationally fast as PerEn, so it is suitable for applications of real-time monitoring. Moreover, the mean value of signal amplitudes and differences between amplitude values are given consideration in DisEn. Li et al. [24] combined DisEn and EMD to extract the rolling bearing fault feature. In [25], DisEn was calculated with the morphology envelope of the original signal for multi-fault recognition of rolling bearings. To enhance the stability of multiscale DisEn [26], Yan et al. [27] further proposed an improved multiscale DisEn for fault diagnosis of rotating machinery. In view of the facts mentioned above, nonlinear entropy can be good symptom features for fault diagnosis of rolling bearings.

After the extraction of fault features, fault classification is essentially a problem of pattern recognition and is often intelligently accomplished with some kind of machine learning models. In [28], the rolling element bearing fault was diagnosed using a hidden Markov model with the extracted singular features. Ali et al. [29] detected bearing malfunction with a probabilistic neural network and then categorized bearing faulty states with a simplified fuzzy adaptive resonance theory map. Li et al. [30] studied comparisons of four representative fuzzy clustering algorithms for bearing fault diagnosis. A sparse autoencoder was proposed in [31] both for feature extraction and fault classification of rolling bearing. Support vector machine (SVM) [32] based on structural risk minimization is

a promising classification method for small samples and has been successfully applied to many engineering fields, such as remote sensing, credit scoring, mechanical fault diagnosis, etc. [33–35]. However, many SVMs should be combined for a multi-class classification since SVM is designed for binary classification problem. Secondly, SVM learning for big data applications is of low efficiency. To overcome these issues, the ensemble SVM (ESVM) was proposed and has been verified with better classification performance than a single SVM for multi-pattern recognition [36–38]. ESVM is a collection of SVMs, and the classification results of these SVMs are combined in some way to output the final decision; thus, the classifying generalization performance is to be improved by utilizing the difference of each individual SVM.

Through the above literature analysis, a novel method for multi-fault diagnosis based on ICEEMDAN, nonlinear dynamics entropy, and ensemble SVM is presented for rolling bearings. In the proposed method, the nonstationary bearing vibration signal is decomposed by ICEEMDAN, and three kinds of entropy features are extracted from each of the obtained IMFs. Then, a distance-based method is formulated for selecting discriminative features, and a multi-class ESVM is constructed and trained for intelligent classification of multi-fault rolling bearings. The innovative contributions of this work can be summarized as follows: (1) a hybrid method by the integration of nonstationary signal adaptive decomposition, nonlinear entropy analysis, and ensemble learning for multi-fault diagnosis of the rolling bearing is systematically presented; (2) nonlinear entropy features including DisEn, PerEn, and SampEn of IMFs from ICEEMDAN are extracted to describe the characteristics of different rolling bearing fault categories; and (3) an ensemble of multi-class SVMs trained with the selected discriminative entropy features is used to accurately and intelligently classify different faults and different severities of the rolling bearing.

The remainder of this paper is organized as follows. Section 2 briefly introduces the background theory knowledge. Section 3 systematically outlines the proposed approach. Section 4 presents application studies of the presented method to rolling bearing experimental data to verify its effectiveness and validity. Finally, Section 5 concludes the paper with summaries and future directions.

2. Background Theories

2.1. Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN)

To enhance the performance of the EMD, some noise-assisted improvements have been proposed. The main idea of these improved versions is to add controllable noise to the signal so that the problems with the original EMD can be solved to some extent. Most recently, ICEEMDAN was proposed to obtain less noise and more physical meaning components with a lower computational cost [13]. The basic principle of ICEEMDAN is reformulated as follows.

Given a univariate signal $\mathbf{s} = [s_1, s_2, \dots, s_L]^T$ of length L , let $\mathbf{w}^{(i)}$ be a realization of zero mean unit variance white noise vector of the same length, $E_k(\cdot)$ be the operator which produces the k th mode obtained by EMD, $M(\cdot)$ be the operator which produces the local mean of the signal that is applied to, and $\langle \cdot \rangle$ be the operator of an ensemble averaging throughout the realizations. Then, the procedures of ICEEMDAN can be briefly described as follows:

Step 1: Calculate the first residue $\mathbf{r}_1 = M(\langle \mathbf{s}^{(i)} \rangle)$ by EMD with the local means of realizations $\mathbf{s}^{(i)} = \mathbf{s} + \beta_0 E_1(\mathbf{w}^{(i)})$ ($i = 1, 2, \dots, I$);

Step 2: Obtain the first mode as $\mathbf{d}_1 = \mathbf{s} - \mathbf{r}_1$;

Step 3: Estimate the k th residue $\mathbf{r}_k = M(\langle \mathbf{r}_{k-1} + \beta_{k-1} E_k(\mathbf{w}^{(i)}) \rangle)$ for $k = 2, \dots, K$.

Step 4: Compute the k th mode as $\mathbf{d}_k = \mathbf{r}_{k-1} - \mathbf{r}_k$.

Step 5: Go to step 4 for next k until K .

Through the ICEEMDAN algorithm, the original signal \mathbf{s} is adaptively decomposed into a set of meaningful intrinsic mode functions (IMFs) and a residual so that useful information can be obtained by further processing these components.

2.2. Dispersion Entropy

Similar to PerEn, DisEn is also a symbolic dynamic (i.e., dispersion pattern) entropy to make a fast and robust measure of time-series complexity. That is, a time series is transformed into a new signal with only a few different elements, and thus, study of the nonlinear dynamics of a time series is simplified to the corresponding distribution of symbol sequences. Although some of detailed information may be lost, some of the invariant and robust dynamic properties can be kept. For the given time series $\mathbf{d} = [d_1, d_2, \dots, d_L]^T$, calculation of its DisEn follows 4 steps.

Step 1: Map each element d_j of the signal \mathbf{d} to one of the c classes. To do so, first, employ the normal cumulative distribution function to map d_j into y_j . Next, y_j is assigned to an integer from 1 to c by a linear algorithm $z_j^c = \text{round}(c \cdot y_j + 0.5)$.

Step 2: Map embedding vectors to dispersion patterns. For each vector $\mathbf{z}_i^{m,c} = [z_i^c, z_{i+\tau}^c, \dots, z_{i+(m-1)\tau}^c]^T$ ($i = 1, 2, \dots, L - (m-1)\tau$) with embedding dimension m and time delay τ , it is mapped to a dispersion pattern $\pi_{v_0 v_1 \dots v_{m-1}}$, where $v_0 = z_i^c, v_1 = z_{i+\tau}^c, \dots, v_{m-1} = z_{i+(m-1)\tau}^c$.

Step 3: Calculate the relative frequency of potential dispersion patterns

$$p(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{\#\{i | i \leq L - (m-1)\tau, \mathbf{z}_i^{m,c} \text{ has pattern } \pi_{v_0 v_1 \dots v_{m-1}}\}}{L - (m-1)\tau} \tag{1}$$

Step 4: Obtain the DisEn in the normalized form as

$$\text{DisEn}(\mathbf{d}, c, m, \tau) = - \sum p(\pi_{v_0 v_1 \dots v_{m-1}}) \ln p(\pi_{v_0 v_1 \dots v_{m-1}}) / \ln c^m \tag{2}$$

When all the elements of a signal are assigned to one class (i.e., $c = 1$), the time series \mathbf{d} is perfectly predictable and then its DisEn value is equal to zero. In contrast, if all the possible dispersion patterns have equal probability, all values of the \mathbf{d} are independent and randomly distributed, thus leading to the maximum value of DisEn.

2.3. Ensemble Support Vector Machine

Support vector machine (SVM), a kind of kernel-based learning algorithm, is a state-of-the-art approach that learns a discriminating hyperplane to maximize the margin and to produce a good generalization ability for classification. It works by implicitly mapping original examples into a high-dimensional space with a kind of nonlinear function and then by optimizing the decision function with the maximal margin to classify different classes. Actually, SVM is originally designed for two-class problems. Denoting the n original examples and the corresponding class labels as set $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, n\}$, where $\mathbf{x}_i \in R^d$ is the d -dimensional input features for the i -sample and $y_i \in \{+1, -1\}$ is the label for the positive and negative class and with the decision function $f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b$ (T is the transpose operator), the primal problem of soft-margin SVM is formulated as

$$\begin{aligned} & \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ & \text{s.t. } y_i (\mathbf{w}^T \varphi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \text{ with } \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \tag{3}$$

where \mathbf{w} and b are the weight vector and bias, $\varphi(\mathbf{x})$ is the nonlinear function that maps \mathbf{x} to the high-dimensional space, ξ_i are slack variables, and C is the penalty factor.

Since $\|\mathbf{w}\|^2$ is convex, the above minimizing problem can be achieved with Lagrange multipliers method and the dual problem follows:

$$\begin{aligned} & \min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_j) - \sum_{i=1}^n \alpha_i \\ & \text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, \text{ with } 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \end{aligned} \tag{4}$$

where α_i is the Lagrange multiplier coefficients.

The solution of Equation (4) gives $\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \varphi(\mathbf{x}_i)$ and the estimated bias b , inserting back to the linear decision function $f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b$ and defining $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$ as a kernel function to calculate the inner product; finally, the decision function of soft-margin SVM is represented with the signature function $\text{sgn}(\cdot)$ as

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (5)$$

In practice, there are abundant multi-class classification applications, so it is required that the original SVM is extended to multi-category problems. For this purpose, the task has been solved by decomposing into multiple two-class problems. Initially, the two strategies of one-against-all and one-against-one have been adopted to construct a number of binary SVMs for discrimination of more than two classes. However, a single SVM may not learn exact parameters for the global optimum and SVM learning is time-consuming for a large scale of data. To address these problems, the ensemble SVM (ESVM) has been proposed and has been verified to improve the classification performance greatly compared to a single SVM [36].

ESVM is a collection of several SVMs in which individual decisions are combined in some way to output the classification result. Each individual SVM is independently trained with the randomly chosen training samples to form one base learner. In such a way, each base learner learns a certain area of the data samples space and is unlikely to be wrong at the same time, so ensemble of these base learners provides complementary information and thus more accurate classification results can be obtained. To ensure the differences among base learners (i.e., SVMs) for enhanced performance, methods including bootstrap and AdaBoost have been considered for selecting data samples for training individual SVM [37]. After training, the outputs from the trained base learners are aggregated in an appropriate manner to decide the final classification result. Majority voting, winner takes all, and weighted averaging are commonly used aggregating rules in case of the multi-class classification.

3. The Proposed Method

To secure properly operation of the rotary machinery, a hybrid method for multi-fault diagnosis of rolling bearings is presented in this paper. Figure 1 shows the framework of the method, which includes signal preprocessing and feature extraction, discriminative feature selection, and intelligent recognition.

3.1. Preprocessing with ICEEMDAN and Feature Extraction

The vibration-based monitoring technique is one of the most effective approaches for fault diagnosis of rolling bearings. However, vibration signals of rolling bearings are highly nonstationary and nonlinear, which complicates the process of fault symptom extraction. To address the signal nonstationarity, the recently developed ICEEMDAN is properly introduced to preprocess the acquired vibrational accelerator signals, so that each monitoring signal \mathbf{s} is decomposed as described in Section 2.1 and a set of IMFs \mathbf{d}_k ($k = 1, 2, \dots, K$) is thus obtained. Before the decomposition, some parameters need to be determined. Firstly, to control the signal-to-noise ratio (SNR) between the added noise and the residue to which the noise is added, the constants β_k ($k = 0, 1, \dots, K$) should be chosen. To simulate the same increasing SNR as that in the EEMD, it was suggested in [13] that $\beta_0 = 0.2 \times \text{std}(\mathbf{s})/\text{std}(E_1(\mathbf{w}^{(i)}))$ and $\beta_k = 0.2 \times \text{std}(\mathbf{r}_k)$ ($k \geq 1$), where $\text{std}(\cdot)$ is the standard deviation operator. This strategy is also considered in this work. For the ensemble realizations I of the final averaging, a number of 100 usually leads to satisfying analysis results for real applications.

To further address the nonlinearity of the vibration signals, the IMFs are processed to extract nonlinear dynamic features of entropy as rolling bearing fault features in this work. Specifically, dispersion entropy (DisEn), permutation entropy (PerEn), as well as sample entropy (SampEn) are considered for their effectiveness and stability. The DisEn features for all the IMFs are obtained as detailed in Section 2.2, while PerEn and SampEn for each IMF are calculated following the procedures in Table 1.

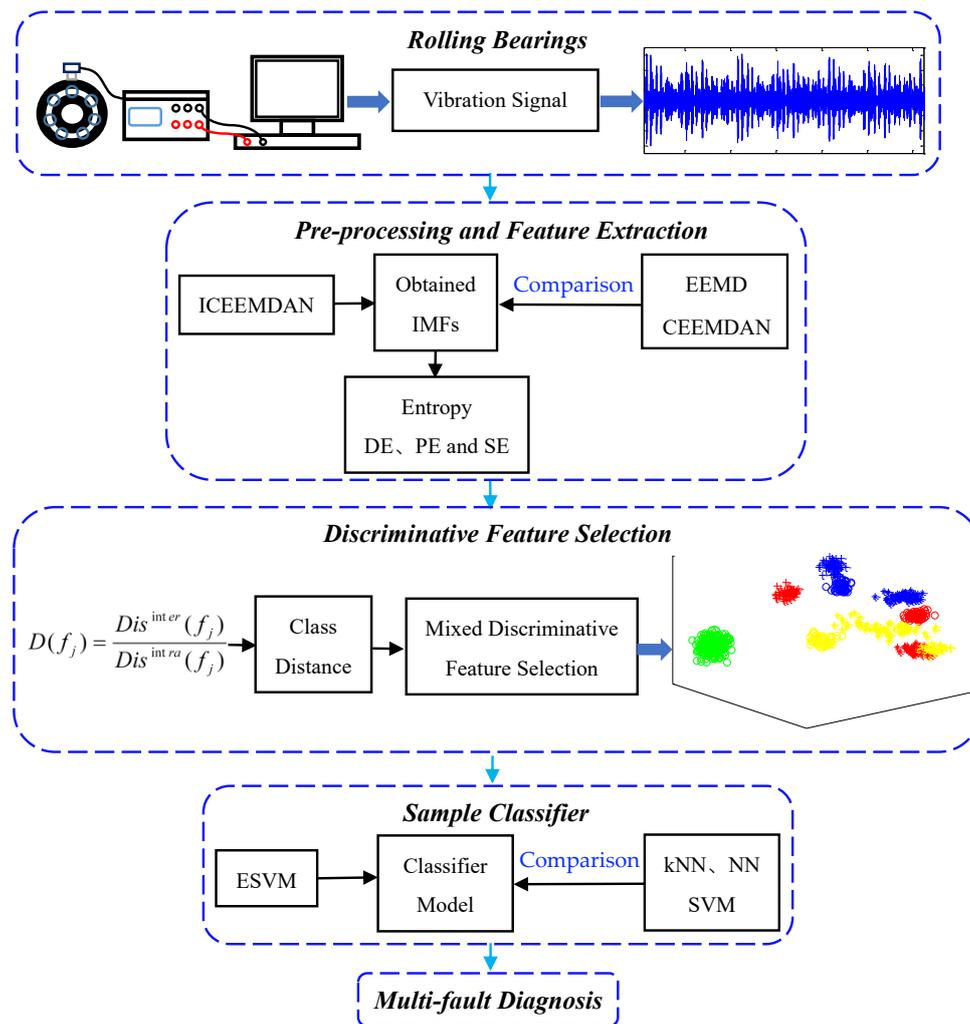


Figure 1. Framework of the proposed multi-fault diagnosis method for rolling bearings.

Table 1. Procedures for calculation of PerEn and SampEn.

| Steps | PerEn | SampEn |
|-------|---|---|
| 1 | Reconstruct each IMF d into state vector $d_i = (d_i, d_{i+\tau}, \dots, d_{i+(m-1)\tau})$ ($i = 1, 2, \dots, L-(m-1)\tau$) with embedding dimension m , time delay τ , and $\tau = 1$ for SampEn. | |
| 2 | Sort d_i in ascending order $d_i = (d_{i_1}, d_{i_2}, \dots, d_{i_m})$ and the ordinal $i_1 i_2 \dots i_m$ is mapped as a permutation pattern $\pi_{u_0 u_1 \dots u_{m-1}}$. | Calculate distance $D_{ij} = D(d_i, d_j) = \max_{l=0,1,m-1} (d_{i+l} - d_{j+l})$ and $C_i^m(r) = \frac{\sum_j H(r - D_{ij})}{L - m - 1}$ ($j \neq i$), where r is the tolerance for Heaviside function $H(\cdot)$. |
| 3 | Calculate frequency of patterns: $p(\pi_{u_0 u_1 \dots u_{m-1}}) = \frac{\#(i_1 i_2 \dots i_m \text{ has the pattern } \pi_{u_0 u_1 \dots u_{m-1}})}{L - (m-1)\tau}$ | Summarize across i : $C^m(r) = \frac{\sum_i C_i^m(r)}{L - m}$ and $m \leftarrow m + 1$ to repeat Steps 1–2 to further obtain $C^{m+1}(r)$. |
| 4 | $\text{PerEn}(\mathbf{d}, m, \tau) = -\sum p(\pi_{u_0 u_1 \dots u_{m-1}}) \ln p(\pi_{u_0 u_1 \dots u_{m-1}})$ | $\text{SampEn}(\mathbf{d}, m, r) = \ln C^m(r) - \ln C^{m+1}(r)$ |

In entropy-based analyzing methods of time series, as in other types of nonlinear dynamics analysis, appropriate choice of the related parameters is an important issue. For DisEn, there are three parameters to be determined, i.e., the number of classes c , embedding dimension m , and time delay τ . Theoretically, two amplitude values that are far from each other are more likely to be assigned to the same class if c is too small while a very small disturbance (such as noise) may change the original classes with a too large c . Additionally, the computational burden increases with the number of class c . To make a tradeoff between useful information detection and computational efficiency, it is recommended to choose c from 4–8 for practical uses of DisEn. For the embedding dimension m , if it is too small, the dynamic changes of the time series might not be detected. In contrast, if m is chosen too

large, small variations may be unable to be observed. Based on the abovementioned facts, embedding dimension m is usually set as 2. According to previous studies [23,27], τ has a very small effect on the estimation of DisEn and aliasing may occur when $\tau > 1$. Hence, τ is usually selected as 1 to guarantee computational efficiency and to provide a reliable analysis.

From Table 1, two parameters of embedding dimension m and time delay τ should be chosen in PerEn analysis. For the embedding dimension, it is appropriate to choose m from 3–7 for bearing vibration analysis based on previous literature [20]. To make full use of the original vibrational signal, τ is often set to be 1. Also, there are two parameters in the calculation of SampEn: reconstruction embedding dimension m and similarity comparison tolerance r . Practically, the tolerance r is a value from 0.1–0.25 $\text{std}(\mathbf{d})$ and the parameter m is set to be 2 in SampEn analysis.

3.2. Discriminative Feature Selection

In general, the extracted entropy features can be directly combined into a longer vector to comprehensively quantify operational status of rolling bearings. However, the high dimensionality and redundancy of the original features will increase the computation time and decrease the pattern recognition accuracy. Thus, it is necessary to select the most representative features to form a low-dimensional discriminative feature vector for fault diagnosis.

The proposed discriminative feature selection is based on the distance principle: the larger distance between different classes and the smaller distance within the same category of one feature. This feature is regarded as discriminative for fault diagnosis. To be specific, denote the feature set as $\mathbf{f}_n \in R^J$ ($n = 1, 2, \dots, N$), where N is the total number of samples and J is the feature dimensionality. The discriminativity of the j th feature is

$$D(f_j) = \frac{Dis^{inter}(f_j)}{Dis^{intra}(f_j)} \tag{6}$$

where Dis^{inter} and Dis^{intra} are the interclass distance and the intraclass distance for the j th feature, respectively, calculated as Equation (7).

$$\begin{aligned} Dis^{inter}(f_j) &= \frac{1}{Q(Q-1)} \sum_{q,p=1; q \neq p}^Q \left| \frac{1}{N_q} \sum_{i=1}^{N_q} f_{i,j}^q - \frac{1}{N_p} \sum_{i=1}^{N_p} f_{i,j}^p \right| \\ Dis^{intra}(f_j) &= \frac{1}{Q} \sum_{q=1}^Q \left(\frac{1}{N_q(N_q-1)} \sum_{i,k=1; i \neq k}^{N_q} |f_{i,j}^q - f_{k,j}^q| \right) \end{aligned} \tag{7}$$

where $f_{i,j}^q$ are samples for the q th ($q = 1, 2, \dots, Q$) class and N_q is its number.

The discriminativity $D(f_j)$ measures ease of classifying the Q classes using the j th feature. A larger discriminativity is, the easier to classify the Q classes that a feature is. Thus, $D(f_j)$ ($j = 1, 2, \dots, J$) is sorted in descending order, and the number of features for fault diagnosis is increased one by one. When the desired accuracy is achieved, the first few features are selected as the most discriminative features.

3.3. ESVM for Multi-Fault Diagnosis

After the discriminative feature selection, one ESVM is constructed and trained for multi-fault intelligent diagnosis of rolling bearings. To be consistent, denote the training samples and the corresponding labels set as $TR = \{(\mathbf{x}_i, y_i), i = 1, 2, \dots, N\}$, where $\mathbf{x}_i = (f_{i,1}, f_{i,2}, \dots, f_{i,d})$ is the d -dimensional discriminative features for the i -sample and $y_i \in \{1, 2, \dots, Q\}$ is the label from the Q fault conditions of rolling bearings. From the statistical view, the training sample sets of each individual SVM (i.e., base learner) should be as different as possible to obtain higher improvement of the ensemble aggregation. Thus, the bagging technique is considered to construct an individual SVM classifier in this paper. Suppose the ensemble number of SVM is P ; bootstrapping then builds replicated data sets TR_p ($p = 1, 2, \dots, P$) by randomly resampling with replacement from the original training data set TR . In this way, each pair \mathbf{x}_i and y_i in the N given training set may appear repeated or not at all in any

particular replication. With the replicated data sets TR_p , an individual SVM is independently trained to act as one classifier as in Equation (5) by the one-against-one strategy.

For one test sample $\mathbf{x} = (f_1, f_2, \dots, f_d)$, it is input to the trained SVMs to output a total of P fault labels. To obtain the fault class label of the new sample, some kind of aggregating rule should be designed. Among the frequently used ensemble rules, majority voting is the simplest one for combining several SVMs and is also utilized here. Let $f_p(\mathbf{x})$ be the decision function for the p th SVM in the ESVM and $\#(p|f_p(\mathbf{x}) = q)$ be the counted number of SVMs in which the decisions are made to the q th class label. Then, final fault class label of the ESVM for the test sample \mathbf{x} is

$$f_{ensemble}(\mathbf{x}) = \max_q(\#(p|f_p(\mathbf{x}) = q)) \quad (8)$$

4. Application to Rolling Bearings Fault Diagnosis

In this section, the proposed multi-fault diagnosis method is applied to the rolling bearing experiments under different operational conditions and is also compared with some existing methods.

4.1. Description of the Experimental Setup

To validate our proposed multi-fault diagnosis method, the publicly available rolling bearing vibration data shared by the bearing data center of Case Western Reserve University (CWRU) [39] is researched in this paper. Since its availability, this data set has been intensively studied for bearing fault diagnostics [40,41]. The test stand of the rolling bearing experiment is illustrated in Figure 2, and the motor shaft (on the left side) support bearings were designed as the testing bearings, which are SKF deep groove ball bearings of 6205-2RS JEM. During the experiments, single point faults ranging from 0.007 to 0.028 inches were introduced to the bearing elements (i.e., ball, and inner and outer raceway) using electro-discharge machining (EDM). Faulted bearings were then reinstalled into the test motor, and vibration data were recorded for motor loads of 0 to 3 horsepower (approximate motor speeds of 1797–1720 rpm).

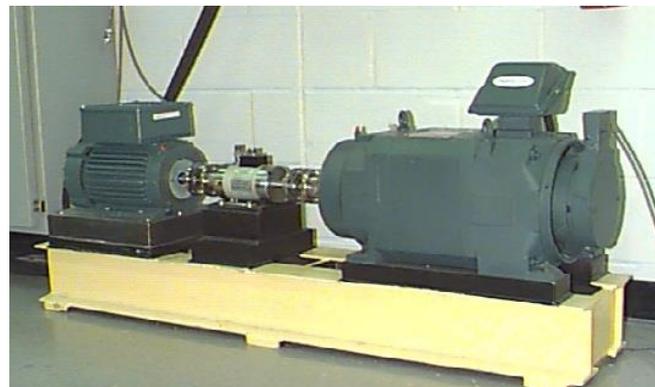


Figure 2. The rolling bearing experiment test stand of Case Western Reserve University (CWRU) bearing data center.

Specifically, the accelerator signals of the drive-end bearing were recorded with the sampling frequency 12 kHz and a duration of one second. In this scenario, there are a total of ten bearing fault classes. Besides the normal state, three defect severities are respectively simulated for the ball, inner raceway, as well as outer raceway. For each class, the original vibration signal is divided into a number of training and testing samples with a length of 2048 data points (about 0.171 s). Details of the analyzed multi-fault bearing data set are listed in Table 2.

Table 2. Description of the bearing multi-fault conditions and their class labels.

| Fault Mode | Fault Size in Inches | Class Labels | Number of Training Samples | Number of Testing Samples |
|---------------|----------------------|--------------|----------------------------|---------------------------|
| Ball | 0.007 | 1 | 47 | 12 |
| | 0.014 | 2 | 47 | 12 |
| | 0.021 | 3 | 47 | 12 |
| Inner Raceway | 0.007 | 4 | 47 | 12 |
| | 0.014 | 5 | 47 | 12 |
| | 0.021 | 6 | 47 | 12 |
| Outer Raceway | 0.007 | 7 | 47 | 12 |
| | 0.014 | 8 | 47 | 12 |
| | 0.021 | 9 | 47 | 12 |
| Normal state | 0 | 10 | 95 | 24 |

4.2. Results and Analysis

After the above experimental configuration, training and testing vibrational samples for different bearing conditions under four motor loads are prepared. One vibration signal sample for each of the ten fault classes under 0 horsepower motor load is illustrated in Figure 3, which demonstrates that it is problematic to discern each bearing fault class directly from the time waveforms, especially for the ball and inner raceway fault modes.

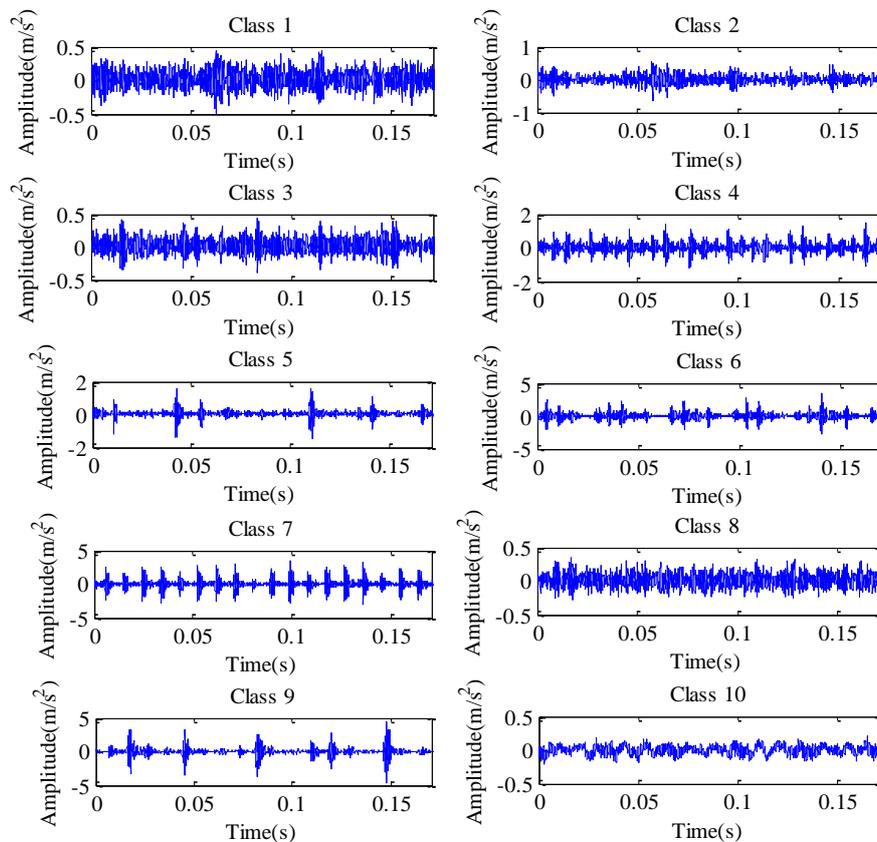


Figure 3. One vibration signal sample for each of the ten bearing fault classes.

Then, the ICEEMDAN algorithm is utilized to preprocess time waveform signals with the parameters chosen as detailed in Section 3.1, and the results for one sample of bearing fault class 7 (i.e., outer raceway fault mode with fault size of 0.007 inches) under 0 horsepower motor load are displayed

in Figure 4. For comparison purposes, the decompositions by CEEMDAN and EEMD are performed with the same parameters. It can be observed from Figure 4 that the periodic impulses generated by an outer raceway point fault are successful detected by all three algorithms. However, there are ten IMFs using CEEMDAN and EEMD algorithms while nine IMFs using ICEEMDAN. Similar results are also found for other vibration samples. Thus, more compact decomposition results can be obtained with the ICEEMDAN algorithm. In the following, the obtained IMFs are further processed to extract fault features of the rolling bearing.

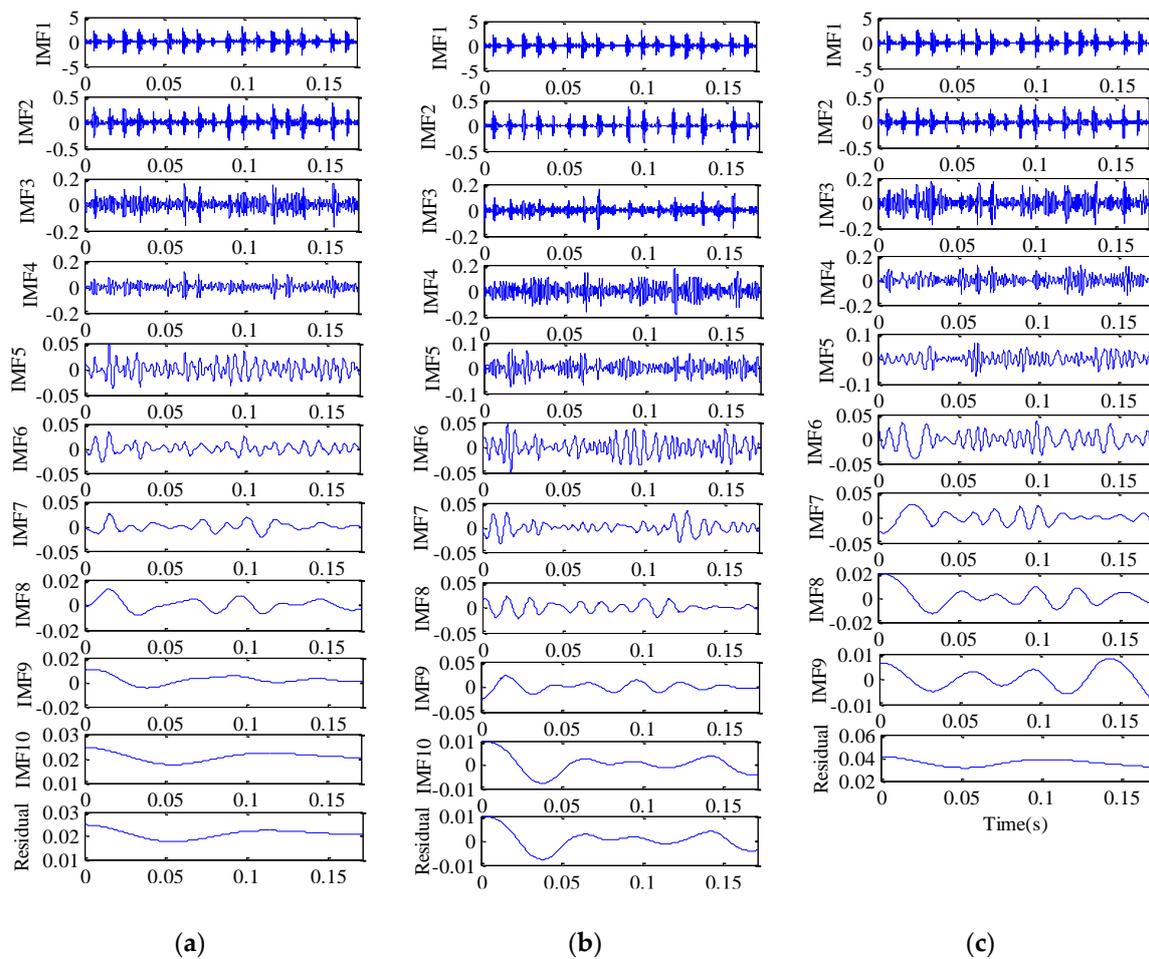


Figure 4. Decomposition results for one vibration signal sample of bearing fault class 7 by (a) ensemble empirical mode decomposition (EEMD), (b) complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and (c) improved CEEMDAN (ICEEMDAN).

Specifically, nonlinear entropy features of DisEn, PerEn, and SampEn are extracted for fault diagnosis of rolling bearings. Before calculation of the three entropy features, the related parameters are to be set. Based on the analysis in Section 3.2 and with some initial trials, the parameters are finally chosen for all cases of bearing conditions under four motor loads as $c = 8$, $m = 2$, and $\tau = 1$ for DisEn; $m = 6$ and $\tau = 1$ for PerEn; and $m = 2$ and $0.2\text{std}(d)$ for SampEn. Actually, the parameters of DisEn and SampEn show limited effects on the extracted features; the embedding dimension has a relatively greater influence on PerEn. To compromise between calculating complexity and effectiveness, these parameters are determined as mentioned above.

The scatter plots of entropy features under 0 horsepower motor load are shown in Figure 5. In previous studies, the IMF1 from such decomposition is usually selected as the components for further analysis; thus, the three kinds of entropy features of IMF1 are also plotted in Figure 5. Some observations can be made. Firstly, different bearing conditions are easier to be discerned in the extracted entropy

feature space than the original time waveforms (refer to Figure 3). This is especially evident for the fault mode of the inner raceway, since the three classes with increasing fault sizes are well clustered and demarcated from the remaining seven classes. Nevertheless, such improvements are limited for the cases of the three ball fault sizes because of their large dispersion. Secondly, the different fault sizes of the rolling bearing outer raceway are both well clustered with and without discriminative feature selection of the extracted nonlinear entropies. Additionally, the proposed discriminative feature selection method can better cluster and demarcate the normal as well as the three ball-fault-bearing conditions compared to that of the results of IMF1.

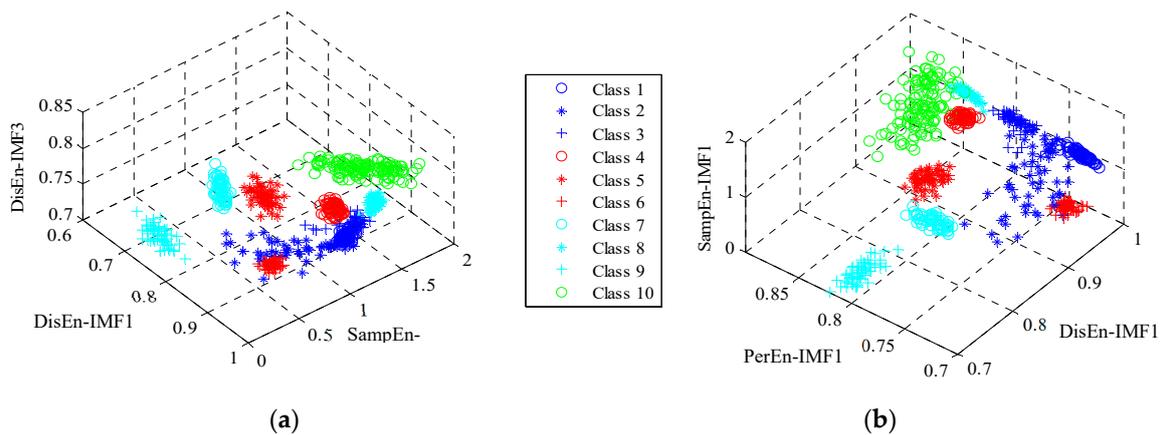


Figure 5. Scatter plots for nonlinear entropy features of the ten bearing fault classes: (a) selected discriminative features and (b) DisEn, PerEn, and SampEn of IMF1.

To have a sense of the effect of the number of selected features on the classifying accuracy, a multi-class ESVM of ensemble number ten is constructed and trained with the extracted nonlinear entropy of the ten bearing conditions. Specifically, the kernel function is a radial basis function with unit bandwidth and the penalty factor $C = 10,000$. The results for class 3 (i.e., ball fault mode with fault size of 0.021 inches) are depicted in Figure 6.

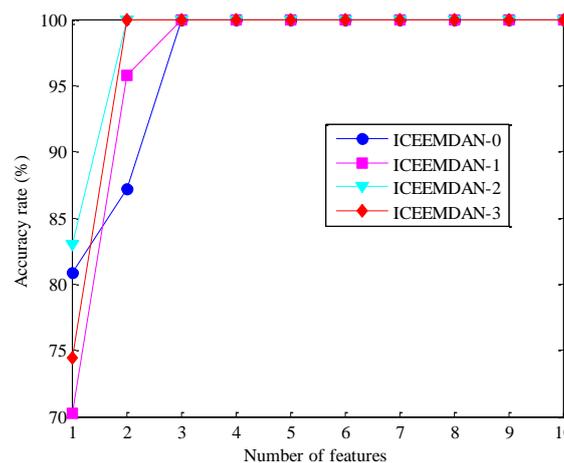


Figure 6. Training classification accuracy for the bearing fault class 3.

Although the accuracy under the four motor loads (i.e., 0 to 3 horsepower) varies differently with the increased number of features, a perfect recognition can be obtained based on the first three discriminative entropy features for this bearing fault class. Furthermore, a detailed analysis of the other nine bearing status shows that more than 85% training accuracy can be obtained if the number of discriminative features is set to be three. Therefore, the first three extracted entropy features are selected as the discriminative features for each class to diagnose rolling bearing faults.

With the above analysis, a multi-fault diagnosis of rolling bearings can then be performed online. To validate the performance of the proposed approach, training and testing diagnostic results of the method as well as the non-discriminative features scenario (i.e., DisEn, PerEn, and SampEn of IMF1) under 0 horsepower motor load are summarized in Table 3. Interestingly, both the compromised training and testing accuracy for the two scenarios are in cases of ball fault mode (i.e., bearing fault classes 1, 2, and 3), which was highlighted for further research in the benchmark literature [41]. For the proposed method with the selected discriminative features, the training and testing accuracies are respectively compromised in fault classes 2 and 3 while the training accuracy is compromised in all three fault classes of ball fault mode and the testing accuracy is compromised in fault classes 2 and 3 for the non-discriminative features scenario.

Table 3. The bearing fault diagnosis results using non-discriminative and selected discriminative features.

| Fault Class | Non-Discriminative Features | | Discriminative Features | |
|-------------|-----------------------------|----------------------|-------------------------|----------------------|
| | Training Accuracy (%) | Testing Accuracy (%) | Training Accuracy (%) | Testing Accuracy (%) |
| 1 | 95.74 | 100 | 100 | 100 |
| 2 | 87.23 | 91.67 | 85.11 | 100 |
| 3 | 93.62 | 91.67 | 100 | 91.67 |
| 4 | 100 | 100 | 100 | 100 |
| 5 | 100 | 100 | 100 | 100 |
| 6 | 100 | 100 | 100 | 100 |
| 7 | 100 | 100 | 100 | 100 |
| 8 | 100 | 100 | 100 | 100 |
| 9 | 100 | 100 | 100 | 100 |
| 10 | 100 | 100 | 100 | 100 |

The corresponding misclassified testing samples for the proposed method and the non-discriminative features scenario under 0 horsepower motor load are plotted in Figure 7. It is observed that one testing sample in fault classes 2 and 3 is respectively classified into fault classes 6 and 8 for the non-discriminative feature scenario. However, only one testing sample from fault class 3 is misclassified into fault class 8 with the selected discriminative features. The above results can be partially explained by the entropy feature scatters as in Figure 5, where the related cluster boundaries are somehow overlapped. To draw comprehensively comparisons between the non-discriminative features (NDF) and the proposed discriminative features (DF) scenarios, testing results of all the bearing fault classes under four motor loads are averaged.

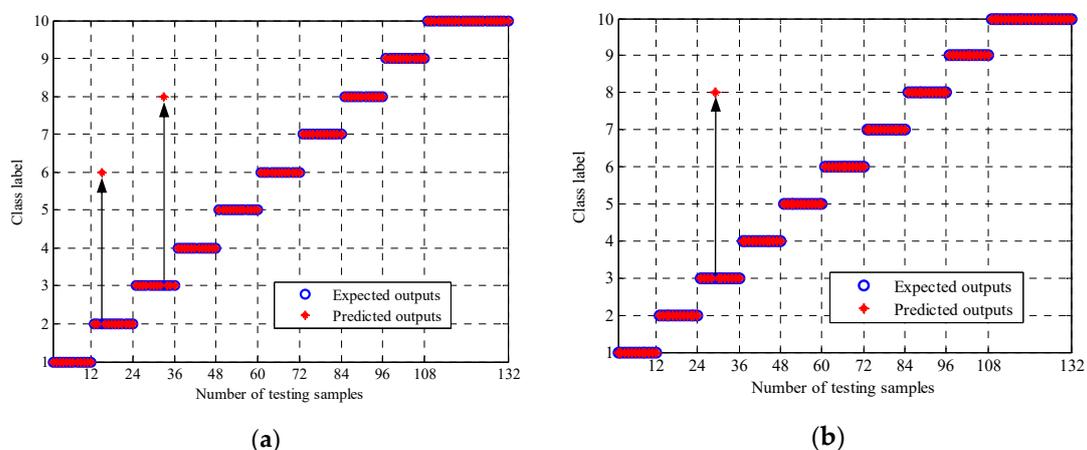


Figure 7. Misclassified testing samples with (a) the non-discriminative features and (b) discriminative features.

From the average recognition rates in Figure 8, only classes 2 and 3 do not have 100% accuracy. The recognition rates for class 2 (i.e., ball fault mode with fault size of 0.014 inches) are especially low. For both classes, a higher average recognition rate is obtained with the discriminative features. Thus, the proposed method with discriminative feature selection is validated to demonstrate much better performance.

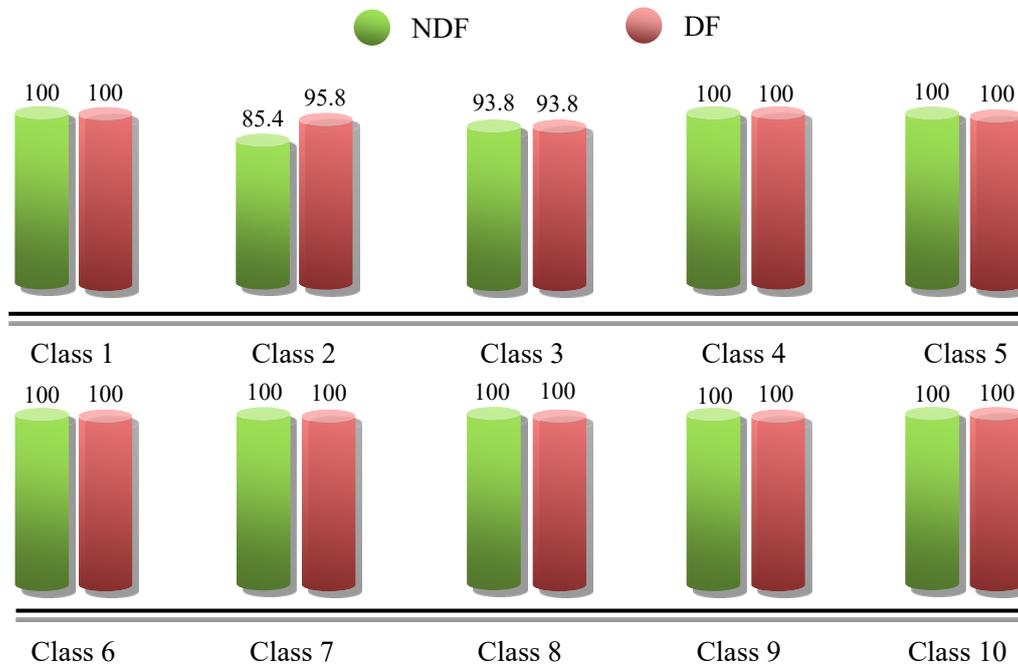


Figure 8. The average recognition rates with non-discriminative and selected discriminative features for the ten bearing fault classes.

4.3. Comparisons and Discussions

In this subsection, our proposed method is further compared with some other related methods.

For vibration signal preprocessing, the ICEEMDAN algorithm is encapsulated in our method. The two algorithms of CEEMDAN and EEMD have also been frequently utilized in literature for bearing signal decomposition. So, to study the preprocessing algorithm for multi-fault diagnosis performance, fault feature extraction and selection are also performed with the IMFs from CEEMDAN and EEMD to classify the ten bearing fault conditions.

The results under 0 horsepower motor load are listed in Table 4. The three algorithms all successfully classify the nine classes except class 3 from the ball fault mode. For this bearing state class under 0 horsepower motor load, two testing samples are misclassified with EEMD and only one is misclassified with CEEMDAN and ICEEMDAN. From the misclassified testing samples in Figure 9 for EEMD and CEEMDAN and in Figure 7b for ICEEMDAN, all the off-target samples of class 3 are labeled class 8 (i.e., outer raceway fault mode with fault size of 0.014 inches). Actually, the time waveforms of classed 3 and 8 in Figure 3 are very similar to each other, which complicates the correct classification of class 3. In addition, the average recognition rates by the three preprocessing algorithms for the ten bearing fault classes under four motor loads are summarized in Figure 10. It is also observed that only classes 2 and 3 do not have perfect recognition and that the best performance is obtained with the case of the ICEEMDAN algorithm, which further verifies the effectiveness of the proposed hybrid method.

Table 4. The bearing fault diagnosis results by the three preprocessing algorithms.

| Fault Class | Classification Accuracy (%) | | |
|-------------|-----------------------------|---------|----------|
| | EEMD | CEEMDAN | ICEEMDAN |
| 1 | 100 | 100 | 100 |
| 2 | 100 | 100 | 100 |
| 3 | 83.33 | 91.67 | 91.67 |
| 4 | 100 | 100 | 100 |
| 5 | 100 | 100 | 100 |
| 6 | 100 | 100 | 100 |
| 7 | 100 | 100 | 100 |
| 8 | 100 | 100 | 100 |
| 9 | 100 | 100 | 100 |
| 10 | 100 | 100 | 100 |

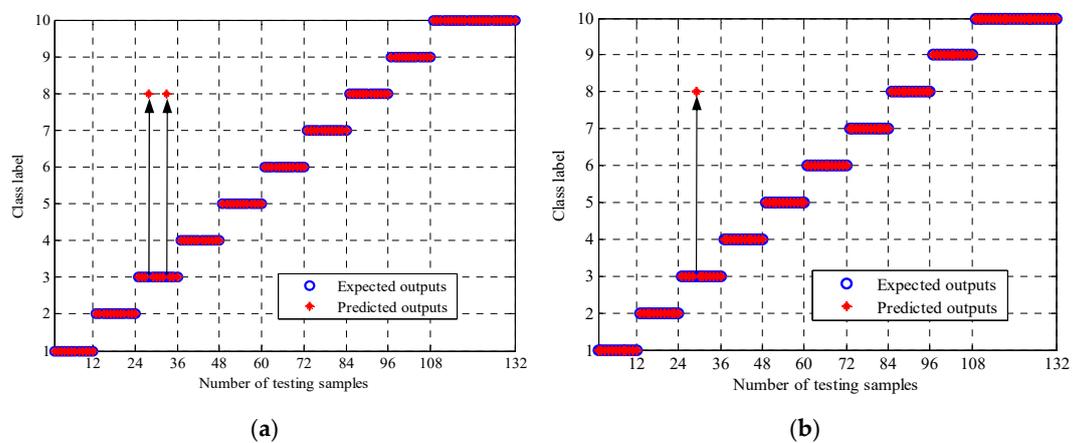


Figure 9. Misclassified testing samples with preprocessing by (a) EEMD and (b) CEEMDAN.

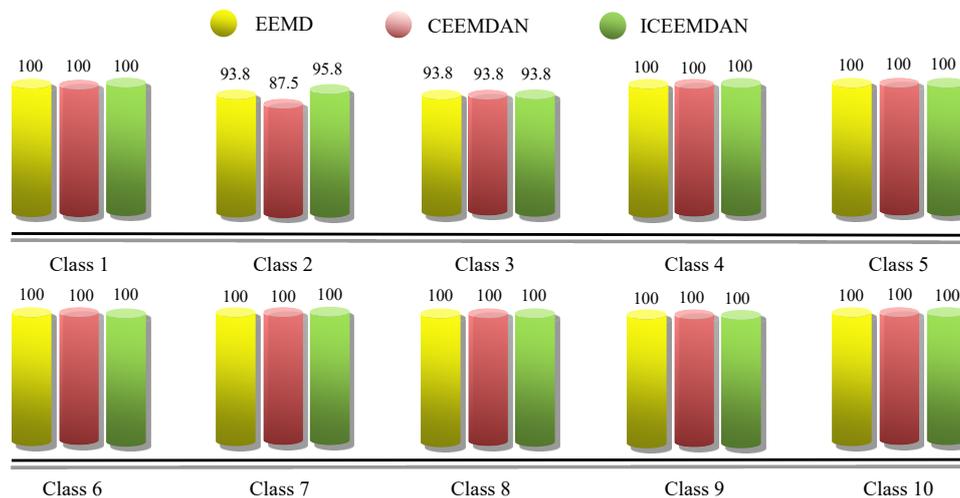


Figure 10. The average recognition rates by the three preprocessing algorithms for the ten bearing fault classes.

In the proposed method, the classifier is a multi-class ESVM. To evaluate the performance of the ESVM, three other pattern recognition models of neural network (NN), k-Nearest Neighbor (kNN), and plain multi-class SVM are also tested. For related settings, the SVM uses the same parameters as ESVM, $k = 1$ for the kNN classifier, and NN is of multilayer perceptron type with the hidden node of 25. The classification results under 0 horsepower motor load are listed in Table 5. It can be concluded that NN has the worst performance with misclassifications in four bearing conditions while the other

three classifiers miss the target in the hardest ball fault mode. To be specific, the kNN misclassifies two testing samples both in classes 2 and 3, SVM misses two targets in class 3, and ESVM only misclassified one testing sample in class 3.

Table 5. The bearing fault diagnosis results by different classifiers.

| Fault Class | Classification Accuracy (%) | | | |
|-------------|-----------------------------|-------|-------|-------|
| | SVM | NN | kNN | ESVM |
| 1 | 100 | 83.33 | 100 | 100 |
| 2 | 100 | 75.00 | 83.33 | 100 |
| 3 | 83.33 | 83.33 | 83.33 | 91.67 |
| 4 | 100 | 100 | 100 | 100 |
| 5 | 100 | 100 | 100 | 100 |
| 6 | 100 | 91.67 | 100 | 100 |
| 7 | 100 | 100 | 100 | 100 |
| 8 | 100 | 100 | 100 | 100 |
| 9 | 100 | 100 | 100 | 100 |
| 10 | 100 | 100 | 100 | 100 |

Detailed analysis can be seen from Figure 11, where the misclassifications are marked. For the multi-class SVM, the testing samples are erroneously classified into not only class 8 but also class 2, which is of poor performance compared to ESVM in Figure 7b. As regards the NN model, eight testing samples are misclassified, resulting in the poorest recognition rates. Finally, kNN classifies two testing samples from class 2 to class 6 and from class 3 respectively to class 8 and class 1. To further evaluate the average performance of the classifiers, the recognition rates under four motor loads are averaged and shown in Figure 12.

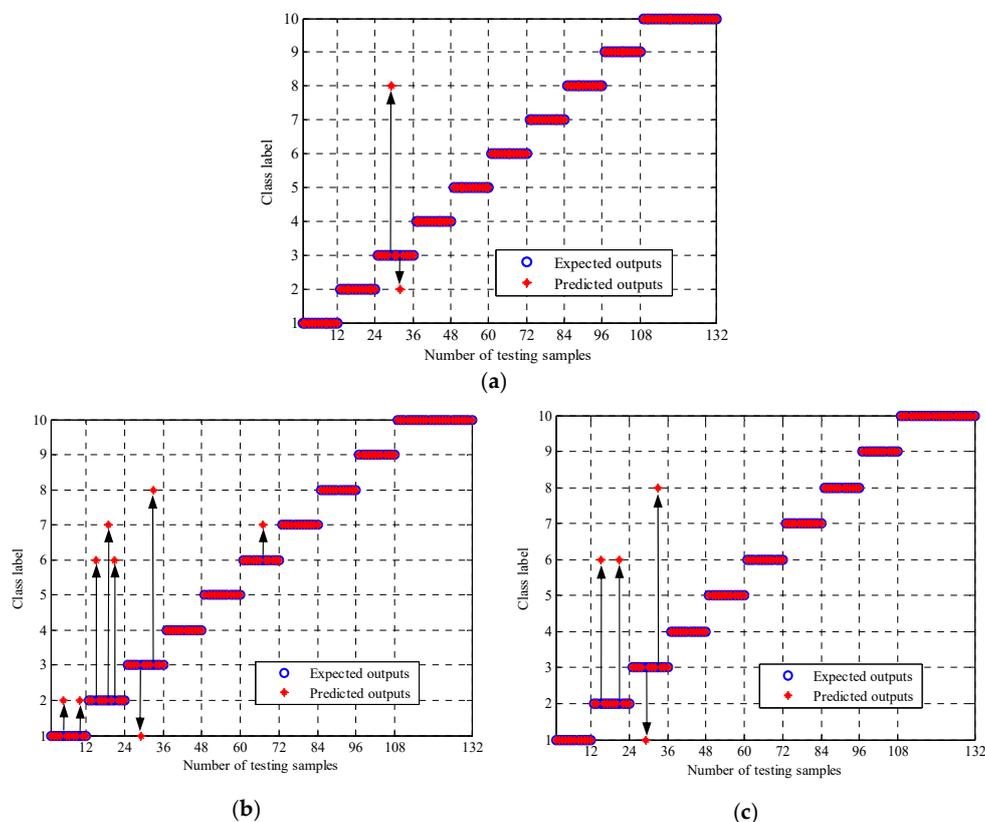


Figure 11. Misclassified testing samples by the classifier: (a) support vector machine (SVM), (b) neural network (NN), and (c) k-Nearest Neighbor (kNN).

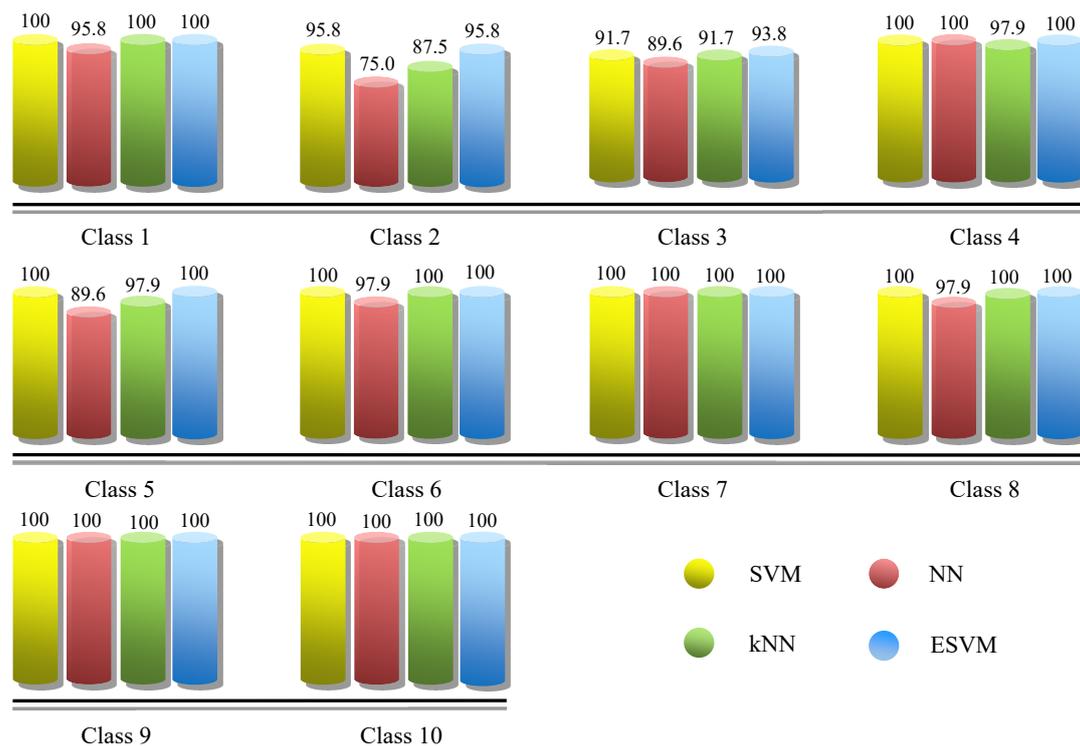


Figure 12. The average recognition rates by the four classifiers for the ten bearing fault classes.

From Figure 12, similar conclusions can be drawn: the NN classifier obtains the lowest classifications. Generally, the reduced recognition rates are also with fault classes 1–3 (i.e., ball fault mode). The classification accuracy is the highest with ESVM, which further highlights the performance of the proposed method.

5. Conclusions

For improved fault diagnosis of rolling bearings, a hybrid method based on ICEEMDAN, nonlinear dynamics entropy, and ESVM for multi-fault diagnosis is presented in this paper. The ICEEMDAN algorithm is utilized to decompose the nonlinear and nonstationary bearing vibration signals, and three kinds of entropy features are extracted from the obtained IMFs. Then, the discriminative features are selected based on the interclass and intraclass distances, and the multi-fault conditions of rolling bearings are intelligently recognized using the multi-class EVSM. The effectiveness of the proposed method is validated with extensive experimental case studies of rolling bearings with different fault modes as well as fault severities under multiple operational conditions (i.e., speeds and loads). Comparisons with some other decomposition algorithms and classifiers further demonstrate the performance of the proposed approach. In the future, the above method for multi-fault rolling bearing diagnosis across different operational conditions will be addressed to bridge the gap between research and real applications.

Author Contributions: Conceptualization, R.L. and B.Z.; methodology, R.L. and B.Z.; software, R.L. and B.Z.; validation, B.Z. and C.R.; formal analysis, L.H., R.L., and S.F.; investigation, B.Z.; resources, R.L.; data curation, B.Z.; writing—original draft preparation, B.Z.; writing—review and editing, R.L. and C.R.; visualization, B.Z.; supervision, B.Z.; project administration, B.Z.; funding acquisition, R.L. and B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Research Program of Chongqing Municipal Education Commission under grant KJ1704109 and KJZD-K201900602, and in part by the Open Found of The State Key Laboratory of Mechanical Transmissions under grant SKLMT-KFKT-201809.

Acknowledgments: The authors would like to appreciate CWRU for their kind permission to use their bearing data.

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

References

1. Shahriar, M.R.; Borghesani, P.; Tan, A.C.C. Electrical signature analysis-based detection of external bearing faults in electromechanical drivetrains. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5941–5950. [[CrossRef](#)]
2. Randall, R.B.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [[CrossRef](#)]
3. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [[CrossRef](#)]
4. Kaewkongka, T.; Joe, Y.H.; Rakowski, R.T.; Jones, B.E. A comparative study of short time Fourier transform and continuous wavelet transform for bearing condition monitoring. *Int. J. COMADEM* **2003**, *6*, 41–48.
5. Chen, J.L.; Li, Z.P.; Pan, J.; Chen, G.G.; Zi, Y.Y.; Yuan, J.; Chen, B.Q.; He, Z.J. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2016**, *70–71*, 1–35. [[CrossRef](#)]
6. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shi, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
7. Lei, Y.G.; Lin, J.; He, Z.J.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [[CrossRef](#)]
8. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
9. Yeh, J.R.; Shieh, J.S.; Huang, N.E. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2010**, *2*, 135–156. [[CrossRef](#)]
10. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 22–27 May 2011; pp. 4144–4147.
11. Lv, Y.; Yuan, R.; Wang, T.; Li, H.; Song, G.B. Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE. *Materials* **2018**, *11*, 1009. [[CrossRef](#)]
12. Kuai, M.K.; Cheng, G.; Pang, Y.S.; Li, Y. Research of planetary gear fault diagnosis based on permutation entropy of CEEMDAN and ANFIS. *Sensors* **2018**, *18*, 782. [[CrossRef](#)] [[PubMed](#)]
13. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control* **2014**, *14*, 19–29. [[CrossRef](#)]
14. Rajesh, K.N.; Dhuli, R. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomed. Signal Process. Control* **2018**, *41*, 242–254. [[CrossRef](#)]
15. Han, H.; Cho, S.; Kwon, S.; Cho, S.B. Fault diagnosis using improved complete ensemble empirical mode decomposition with adaptive noise and power-based intrinsic mode function selection algorithm. *Electronics* **2018**, *7*, 16. [[CrossRef](#)]
16. Li, Y.B.; Wang, X.Z.; Liu, Z.B.; Liang, X.H.; Si, S.B. The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review. *IEEE Access* **2018**, *6*, 66723–66741. [[CrossRef](#)]
17. Han, M.H.; Pan, J.L. A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings. *Measurement* **2015**, *76*, 7–19. [[CrossRef](#)]
18. Liu, H.H.; Han, M.H. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. *Mech. Mach. Theory* **2014**, *75*, 67–78. [[CrossRef](#)]
19. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 1–5. [[CrossRef](#)]
20. Yan, R.Q.; Liu, Y.B.; Gao, R.X. Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. *Mech. Syst. Signal Process.* **2012**, *29*, 474–484. [[CrossRef](#)]
21. Xue, X.M.; Li, C.S.; Cao, S.Q.; Sun, J.C.; Liu, L.Y. Fault diagnosis of rolling element bearings with a two-step scheme based on permutation entropy and random forests. *Entropy* **2019**, *21*, 96. [[CrossRef](#)]
22. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [[CrossRef](#)]

23. Rostaghi, M.; Ashory, M.R.; Azami, H. Application of dispersion entropy to status characterization of rotary machines. *J. Sound Vib.* **2019**, *438*, 291–308. [[CrossRef](#)]
24. Li, C.Z.; Zheng, J.D.; Pan, H.Y.; Liu, Q.Y. Fault diagnosis method of rolling bearings based on adaptive multi-scale dispersion entropy. *Noise Vib. Control* **2018**, *38*, 173–180.
25. Wan, S.T.; Peng, B. An integrated approach based on swarm decomposition, morphology envelope dispersion entropy, and random forest for multi-fault recognition of rolling bearing. *Entropy* **2019**, *21*, 354. [[CrossRef](#)]
26. Azami, H.; Kinney, L.E.; Ebied, A.; Fernandez, A.; Escudero, J. Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in Alzheimer’s disease. In Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Juje Island, Korea, 11–15 July 2017; pp. 3182–3185.
27. Yan, X.A.; Jia, M.P. Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection. *Knowl. Based Syst.* **2019**, *163*, 450–471. [[CrossRef](#)]
28. Liu, T.; Chen, J.; Dong, G.M. Singular spectrum analysis and continuous hidden Markov model for rolling element bearing fault diagnosis. *J. Vib. Control* **2015**, *21*, 1506–1521. [[CrossRef](#)]
29. Ali, J.B.; Saidi, L.; Mouelhi, A.; Chebel-Morello, B.; Fnaiech, F. Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations. *Eng. Appl. Artif. Intel.* **2015**, *42*, 67–81.
30. Li, C.; Cerrada, M.; Cabrera, D.; Sanchez, R.V.; Pacheco, F.; Ulutagay, G.; Valentede, O.J. A comparison of fuzzy clustering algorithms for bearing fault diagnosis. *J. Intell. Fuzzy Syst.* **2018**, *34*, 3565–3580. [[CrossRef](#)]
31. Ahmed, H.O.A.; Wong, M.L.D.; Nandi, A.K. Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features. *Mech. Syst. Signal Process.* **2018**, *99*, 459–477. [[CrossRef](#)]
32. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [[CrossRef](#)]
33. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
34. Harris, T. Credit scoring using the clustered support vector machine. *Expert Syst. Appl.* **2015**, *42*, 741–750. [[CrossRef](#)]
35. Islam, M.M.; Kim, J.M. Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines. *Rel. Eng. Syst. Saf.* **2019**, *184*, 55–66. [[CrossRef](#)]
36. Kim, H.C.; Pang, S.; Je, H.M.; Kim, D.; Bang, S.Y. Constructing support vector machine ensemble. *Pattern Recogn.* **2003**, *36*, 2757–2767. [[CrossRef](#)]
37. Zheng, J.D.; Pan, H.Y.; Cheng, J.S. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **2017**, *85*, 746–759. [[CrossRef](#)]
38. Zhou, S.H.; Qian, S.L.; Chang, W.B.; Xiao, Y.Y.; Cheng, Y. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. *Sensors* **2018**, *18*, 1934. [[CrossRef](#)]
39. Case Western Reserve University Bearing Data Center. Available online: <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 20 May 2020).
40. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [[CrossRef](#)]
41. Li, Y.B.; Wang, X.Z.; Si, S.B.; Huang, S.Q. Entropy based fault classification using the Case Western Reserve University data: A benchmark study. *IEEE Trans. Reliab.* **2019**, *69*, 754–767. [[CrossRef](#)]

