OPEN ACCESS

entropy

Article

Multi-Scale Analysis Based Ball Bearing Defect Diagnostics Using Mahalanobis Distance and Support Vector Machine

Shuen-De Wu [1], Chiu-Wen Wu [1], Tian-Yau Wu [2,*] and Chun-Chieh Wang [3]

[1] Department of Mechatronic Technology, National Taiwan Normal University, Taipei, 10610, Taiwan; E-Mails: sdwu@ntnu.edu.tw (S.D.W); im113qr@gmail.com (C.W.W)

[2] Research Center for Adaptive Data Analysis, National Central University, Jhongli, 32001, Taiwan

[3] Advanced Mechanical Technology Department, Industrial Technology Research Institute, Chutung Township, 31040, Taiwan; E-Mail: chunchiehwang@itri.org.tw

* Author to whom correspondence should be addressed; E-Mail: tianyauw@ncu.edu.tw;
Tel.: +886-3-422-7151; Fax: +886-3-426-9736.

Abstract: The objective of this research is to investigate the feasibility of utilizing the multi-scale analysis and support vector machine (SVM) classification scheme to diagnose the bearing faults in rotating machinery. For complicated signals, the characteristics of dynamic systems may not be apparently observed in a scale, particularly for the fault-related features of rotating machinery. In this research, the multi-scale analysis is employed to extract the possible fault-related features in different scales, such as the multi-scale entropy (MSE), multi-scale permutation entropy (MPE), multi-scale root-mean-square (MSRMS) and multi-band spectrum entropy (MBSE). Some of the features are then selected as the inputs of the support vector machine (SVM) classifier through the Fisher score (FS) as well as the Mahalanobis distance (MD) evaluations. The vibration signals of bearing test data at Case Western Reserve University (CWRU) are utilized as the illustrated examples. The analysis results demonstrate that an accurate bearing defect diagnosis can be achieved by using the extracted machine features in different scales. It can be also noted that the diagnostic results of bearing faults can be further enhanced through the feature selection procedures of FS and MD evaluations.

Keywords: multi-scale entropy (MSE); multi-scale permutation entropy (MPE); multi-scale root-mean-square (MSRMS); multi-band spectrum entropy (MBSE); fisher score (FS); mahalanobis distance (MD); support vector machine (SVM)

1. Introduction

Machine health monitoring and fault diagnosis have attracted considerable attention in both the academic fields as well as the industrial applications. Among the enormous number of components and devices, bearings are one of the most crucial parts in rotating machinery. Due to improper initial assembly, manufacturing inaccuracy, applied repetitive stress or impurity entanglement, the occurrence of bearing damage is inevitable during long term operation. Therefore, bearing fault detection in the early stages is an important task in the fields of mechanical and aerospace engineering. An accurate detection of bearing faults can reduce the maintenance costs and assure the machine operation safety. The bearing components are generally inaccessible for inspection due to the disassembly problem and other possible hazards. Therefore, the vibration analysis is one of the most commonly used techniques for detecting the malfunctions of bearings in rotating systems since the mechanical vibration signals consist of plentiful information related to the system dynamical characteristics.

Bearing fault diagnosis has been studied in the past decades and offered promising results in a large volume of prior literature. Among the previous investigations, Randall [1,2] provided a complete and systematic review for summarizing the measurement techniques as well as the fault diagnosis of rotating mechanical systems. In addition to the conventional spectral analysis of faulty bearing vibration signals [3,4], the envelope analysis and the demodulation technique are also widely utilized for diagnosis of different bearing defects [5–8]. Because of the non-stationarity of vibration signals in the damaged machine, on the other hand, the time-frequency analysis methods, such as wavelet transform [9–13] and Hilbert-Huang Transform, are also employed to analyze the complicated signals with faulty features [10,14–17].

Even though the aforementioned traditional methods have demonstrated the capability of diagnosing bearing defects and shown promising results, the bearing fault-related features are generally clouded by massive uncorrelated signals, and thus it is not easy to observe the fault features of bearings when the previous methods are used. Therefore, some filtering process is usually used to remove the uncorrelated components and preserve the faulty feature-correlated signals. Nevertheless, due to the lack of the rigorous rules for selecting the central frequency and bandwidth of the filters, these filter parameters are generally decided based on empirical experience. Moreover, most of the filtering processes will inevitably cause some phase shifts or waveform distortion of the signals, which may interfere with the analysis and diagnosis of the vibration signals.

As presented in the previous literature, the entropy can be used to measure the regularity or orderliness of the signals, and hence different calculations of entropy, such as the approximate entropy [18], spectral entropy [19], pattern spectrum entropy [20], Hilbert-Huang entropy [21], and energy entropy [22], are defined to characterize the system dynamics and disorderliness that are related to machinery defects and malfunctions. In addition, Costa *et al.* [23,24] indicated that the heartbeat time series of different pathological patients cannot be classified while the sample entropy of the heartbeat time series in one scale is analyzed. Therefore, they proposed the idea of multi-scale entropy (MSE) for separating the human heartbeat signals of healthy and pathological groups. In light of the effectiveness of MSE in physiological and biological signal analysis, different MSE-based algorithms and improved methods were proposed to reduce the influence of noise and employed in vibration analysis of rotating machinery [25–28]. Furthermore, the concept of multi-scale permutation entropy (MPE) was applied

to diagnose bearing faults and demonstrated enhanced performance compared with the single scale permutation entropy method [29].

In general, the classification schemes, such as artificial neural-fuzzy network [22,25], support vector machine (SVM) [8,12,26] and decision tree [13,30], are utilized to effectively identify the different types of machine faults. In order to increase the computational efficiency of the classification procedure and enhance the accuracy of the diagnosis results, the principal component analysis (PCA) method is to select the certain features of high priority [30,31]. However, the selected features will lose the original physical representations through the feature space transformation process of PCA. Alternatively, Chen *et al.* [32] proposed the calculation of distinction index (DI) to extract the certain significant features among the numerous features. The DI is also known as the Fisher score (FS) and applied to select the signatures for face recognition [33]. The assessment of the Mahalanobis distance (MD) can be used to measure the distinction between the characteristic vectors and demonstrates the effectiveness of identifying the component looseness at different locations of rotating machinery [16,34].

To advance the state-of-art of bearing defect diagnostics, a new approach is proposed to investigate the effectiveness of utilizing the multi-scale analysis and feature selection methods for diagnosis of different bearing faults in this research. The MSE, MPE, multi-scale root-mean-square (MSRMS) and multi-band spectrum entropy (MBSE) of the machine vibration signals are determined as the features in different scales. Through the FS evaluation, the features of high distinguishabilities are selected for diagnosis of different bearing defects. On the other hand, the combinations of features with high distinguishabilities can be also extracted by the MD evaluation procedure. The selected features are casted into the one-against-one SVM structure to classify the different classes of bearing faults. With the feature selection process, either the FS or MD evaluation, both the accuracy of classification results and the computational efficiency can be enhanced. Moreover, the features selected by means of the above feature selection methods can preserve the original physical representations since either the FS or MD evaluation process consists of statistical calculation without space transformation as PCA.

The research in this paper utilizes the data of the Bearing Center at Case Western Reserve University (CWRU) [35] to validate the feasibility of the proposed method and evaluate the effectiveness of this approach. The analysis results show that the different classes of bearing defects can be diagnosed accurately through the multi-scale analysis and SVM classification. The results demonstrate that the diagnostic accuracy can be enhanced by imposing the feature selection process. It is also noted that the MD evaluation can achieve more accurate diagnostic results than the FS evaluation.

2. Entropy and Multi-Scale Analysis

The concept of entropy is conventionally applied in the fields of information theory as well as thermodynamics. In 1948, Shannon proposed the method of entropy quantification to characterize the information buried in the data, and thus the method is broadly employed to measure the regularity or orderliness of a time series [36]. In general, the entropy increases with the degree of disorder and is the maximum for a completely random system [23]. Since the entropy analysis has been utilized for feature extraction of pathological and biological systems and presented the effectiveness of disease diagnosis, the concept of entropy quantification can be also employed for machine fault detection [25–28].

Suppose a single discrete random series **S** has N outcomes in which there exist n classes. The entropy $En(\mathbf{S})$ of **S** is defined as:

$$En(\mathbf{S}) = -\sum_{i=1}^{n} p(s_i) \log(p(s_i)), \ s_i \in \mathbf{S}, \ 1 \le i \le n \tag{1}$$

where $p(s_i)$ is the probability density function of the random series **S**, and log represents the natural logarithmic function or logarithmic function to base of 2. The entropy is generally applied to indicate the complexity of a random variable series and can be utilized to predict the system behavior of the time series. According to the computations of the probability density functions, the entropy can be formulated in terms of the following different representations.

2.1. Sample Entropy

Consider a time series with data length of N, $\mathbf{S}=\{x_1, x_2,..., x_N\}$. Let m sequential points of the time series be a pattern. For instance, $\mathbf{X}_i=[x_i, x_{i+1},..., x_{i+m-1}]$ represents the i-th pattern. Therefore, this time series consists of N-m+1 patterns. Define the pattern space **X** as:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-m+1} & x_{N-m+2} & \cdots & x_N \end{bmatrix} \tag{2}$$

The sample entropy can be determined by the following steps:

(1) Measure the mean self-similarity value of the pattern of length m, $\varphi_m(r)$, where r is the tolerance.
(2) Expand the pattern length m to m+1, and measure the mean value of $\varphi_{m+1}(r)$.
(3) The sample entropy (SEn) is then determined as [37]:

$$SEn(\mathbf{S}, m, r) = -\log \frac{\varphi_{m+1}(r)}{\varphi_m(r)} \tag{3}$$

In Equation (3), the self-similarity value is formulated as:

$$\varphi_m(r) = \frac{1}{N-m} \frac{1}{N-m+1} \sum_{i=1}^{N-m} \sum_{j=1}^{N-m+1} G(d_{ij}, r) \tag{4}$$

where $G(\cdot)$ represents the Heaviside function and $d_{ij} = \left\| \mathbf{X}_i - \mathbf{X}_j \right\|_{\infty}$.

2.2. Spectral Entropy

The concept of spectral entropy was first proposed by Powell *et al.* [38] and has been applied in fault diagnosis of gearboxes and bearings [39]. Like Shannon entropy, the spectral entropy determines the power spectrum density of the signals in the probability density function as shown in Equation (1). In this research, therefore, the Discrete Fourier Transform (DFT) is utilized to calculate the spectrum of the signals. Let $Y(f_i)$ be the magnitude of the signal at the frequency f_i, and the power spectrum density $\hat{Y}(f_i)$ is then defined as:

$$\hat{Y}(f_i) = \frac{Y(f_i)}{\sum_{i=1}^{N} Y(f_i)}, \ f_i \in \left[0, \frac{f_s}{2}\right] \tag{5}$$

where f_s represents the sampling frequency. The spectral entropy *SpEn* is thus formulated as:

$$SpEn(\mathbf{S}) = -\sum_{i=1}^{N} \hat{Y}(f_i) \log(\hat{Y}(f_i)) \tag{6}$$

2.3. Permutation Entropy

Permutation entropy, which is an alternative method of defining the quantity of irregularity or complexity, was first proposed by Bandt *et al.* in 2002 and applied to speech signal processing and analysis [40]. The concept of permutation entropy was also employed to analyze electroencephalographic signals [41,42] as well as the working status characterization of rotary machines [43]. Based on the Shannon entropy, the permutation entropy is defined by calculating the probability density function of a time series with permutation of specific order. Consider a time series $\{x(t)\}_{t=1,2\ldots,N}$. We consider an order-*m* segment $\mathbf{X}_i^m$ of the time series:

$$\mathbf{X}_i^m = [x(i), x(i+1), \ldots, x(i+m-1)] \tag{7}$$

and hence there are *m*! permutation π of order *m*. For each permutation π, the relative frequency is determined by the following formula:

$$p(\pi) = \frac{\#\left\{\mathbf{X}_i^m \text{ has type } \pi, i \middle| 1, 2, \ldots, N-m+1\right\}}{N-m+1} \tag{8}$$

where # represents the number. The permutation entropy is thus expressed in terms of the probability density function which is represented by the relative frequency, *i.e*:

$$PEn(\mathbf{S}, m) = -\sum_{i=1}^{m!} p(\pi_i) \log(p(\pi_i)) \tag{9}$$

Let us take an example of a seven-value series [40]:

$$\mathbf{X} = [4, 7, 9, 10, 6, 11, 3] \tag{10}$$

For order *m*=3, the series **X** can be segmented as:

$$\begin{bmatrix} \mathbf{X}_1^3 \\ \mathbf{X}_2^3 \\ \mathbf{X}_3^3 \\ \mathbf{X}_4^3 \\ \mathbf{X}_5^3 \end{bmatrix} = \begin{bmatrix} 4 & 7 & 9 \\ 7 & 9 & 10 \\ 9 & 10 & 6 \\ 10 & 6 & 11 \\ 6 & 11 & 3 \end{bmatrix} \tag{11}$$

Since the Eauations (4), (7) and (9) and Equations (7), (9) and (10) are in increasing order ($x_t < x_{t+1} < x_{t+2}$), they represent the permutation 012. Equations (9), (10) and (6) and Equations (6), (11) and (3) correspond to permutation 201 since $x_{t+2} < x_t < x_{t+1}$, while Equations (10), (6) and (11) is categorized to permutation type 102 with $x_{t+1} < x_t < x_{t+2}$. Therefore, the permutation entropy of order *m*=3 is determined as:

$$PEn(\mathbf{X},3) = -\frac{2}{5}\log(\frac{2}{5}) - \frac{1}{5}\log(\frac{1}{5}) - \frac{2}{5}\log(\frac{2}{5}) \approx 1.522 \tag{12}$$

2.4. Multi-Scale Analysis

2.4.1. Coarse-Grain Process

The concept of multi-scale analysis was proposed by Costa *et al.* and utilized to investigate the feasibility of identifying the pathological human's physiological signals [23,24]. As depicted in their study, it is generally difficult to distinguish the inter-beat interval time series of different diseased and healthy systems if only a single-scale sample entropy is analyzed. Therefore, they proposed the concept of MSE to transform the original signal into different scales of data through the coarse-grained process. Their analysis results demonstrated that the MSE analysis is capable of classifying the different diseased systems.

For a given time series signal, $\mathbf{S}=\{x_1, x_2,..., x_N\}$, it is segmented into several data sets of length τ. The new time series sets $\{y_j^{(\tau)}\}$ are then reconstructed by taking the mean values of segmented data, according to the equation:

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \ 1 \le j \le \frac{N}{\tau} \tag{13}$$

where τ is called the scale factor. Obviously, the coarse-grain is equivalent to the process of sliding window of length τ and taking the average of the original signal within the window in the way of non-overlap. In the point of view of signal processing, the coarse-grain process includes the following two steps:

(1) using the moving average filter to remove the high frequency components.
(2) down-sampling the signal.

Through the coarse-grain process, the multi-scale analysis can be utilized to characterize the dynamics of the signals in different scales.

2.4.2. Multi-Scale Entropy

According to coarse-grain process for multi-scale analysis, the MSE is defined by calculating the *SEn* of $y_j^{(\tau)}$ with different scale τ, *i.e*:

$$MSE(\mathbf{S},\tau,m,r) = SEn(\mathbf{y}^{(\tau)},m,r) \tag{14}$$

where $\mathbf{y}^{(\tau)} = \{y_j^{(\tau)}\}$ represents the series obtained through the coarse-grain process with the scale τ. It is noted that the tolerance r in calculating the *SEn* of different scales is fixed without variation to the scale τ. In general, the tolerance r is selected to be 0.15 times the standard deviation of the time series signal $\mathbf{S}=\{x_1, x_2,..., x_N\}$ [44]. Therefore, the MSE analysis is not affected by the original signal amplitude.

2.4.3. Multi-Scale Permutation Entropy

The concept of multi-scale permutation entropy (MPE) was proposed by Aziz and Arif in 2005 and applied in measuring the entropy of physiological signals [45]. In their study, it is demonstrated that the MPE has higher robustness in tolerating the signal noise than the MSE does. The MPE is defined as:

$$MPE(\mathbf{S}, \tau, m) = nPEn(\mathbf{y}^{(\tau)}, m) \tag{15}$$

where *nPEn* represents the normalized *PEn* that is defined by:

$$nPEn(\mathbf{S}, m) = \frac{PEn(\mathbf{S}, m)}{\log(m!)}, 0 \le nPEn(\mathbf{S}, m) \le 1 \tag{16}$$

Since the MSE and MPE correspond to the different features of dynamical systems, the two measurements of entropy can be utilized to characterize the signals of rotary system with faulted bearings.

2.4.4. Multi-Scale Root-Mean-Square

The root-mean-square (RMS) value is a traditional method to measure the signal's energy statistically and has been broadly utilized for machine condition monitoring. Based on the aforementioned concept of multi-scale analysis, the RMS measurement can be extended to quantify the energy distributions at different scales or frequency bands. The multi-scale RMS (MSRMS) is defined as:

$$MSRMS(\mathbf{S}, \tau) = RMS(\mathbf{y}^{(\tau)}) \tag{17}$$

where $\mathbf{y}^{(\tau)} = \left\{ y_j^{(\tau)} \right\}$ represents the series obtained through the coarse-grain process with the scale τ. While the machine dynamics changes, particularly for the occurrence of component faults, the vibration energy distributions of different scales change correspondingly. Therefore, the MSRMS values can be also used as the indices for bearing defect diagnosis of rotary systems.

2.4.5. Multi-Band Spectrum Entropy

Through the coarse-grain process as expressed in Equation (13), the calculation of multi-band spectrum entropy is employed to measure the disorderliness of the signal spectrum within the different scales. Let $\mathbf{y}^{(\tau)}$ represent the time series through the coarse-grain process with the scale τ. The multi-band spectrum entropy (MBSE) is defined as:

$$MBSE(\mathbf{S}, \tau) = -\sum_{i=1}^{n^{(\tau)}} \hat{X}^{(\tau)}(f_i) \log(\hat{X}^{(\tau)}(f_i)), \ f_i \in \left[0, \frac{f_s}{2\tau} \right] \tag{18}$$

where $\hat{X}^{(\tau)}(f_i)$ represents the power spectrum density of $\mathbf{y}^{(\tau)}$ at the frequency f_i, and $n^{(\tau)}$ is the length of the power spectrum density. The normalized MBSE (nMBSE) can be thus formulated as:

$$nMBSE(\mathbf{S}, \tau) = -\frac{1}{\log\left(\dfrac{n}{\tau} \right)} MBSE(\mathbf{S}, \tau) \tag{19}$$

3. Feature Selection

As introduced in Section 2, multi-scale analysis can be utilized to extract the features of different scales in the signals, such as MSE, MPE, MSRMS and MBSE. However, superfluous features with different scales may interfere with the diagnosis of machine faults, and inevitably result in the inaccuracy of identification as well as the increase of computational efforts. Therefore, it is beneficial to select the high-priority features among the large number of features for fault diagnosis purpose before the feature classification process. In this research, two indices, FS and MD, are employed to evaluate the significance of the extracted features.

3.1. Fisher Score

FS is used to evaluate the distinguishability of the k-th feature for classifying the data sets between the i-th and the j-th classes. Suppose there are N_i data sets in the i-th class and $\{v_{ik}(1), v_{ik}(2),...,v_{ik}(N_i)\}$ represent the k-th feature of all the N_i data sets in the i-th class. The mean value and standard deviation of the k-th feature are then determined as:

$$\mu_{ik} = \frac{1}{N_i}\sum_{p=1}^{N_i} v_{ik}(p)$$

$$\sigma_{ik} = \sqrt{\frac{1}{N_i}\sum_{p=1}^{N_i}(v_{ik}(p)-\mu_{ik})^2} \tag{20}$$

The FS of the k-th feature between the i-th and the j-th classes is defined as [32]:

$$I_{ij}^{(k)} = \frac{\left(\mu_{ik}-\mu_{jk}\right)^2}{\sigma_{ik}^2+\sigma_{jk}^2} \tag{21}$$

By observing Equation (21), obviously, a large FS quantity represents that the k-th feature has high distinction between i-th and the j-th classes. Namely, the k-th feature has high priority for classifying the data sets between the i-th and the j-th classes.

3.2. Mahalanobis Distance

Mahalanobis [46] first proposed the determination of MD in which a statistical method is introduced to measure the distance between different data sets of two different classes. For the data sets of two different classes, let μ_i and μ_j represent the mean values of the characteristic vectors in the i-th and the j-th classes, and the MD between the characteristic vectors of the i-th and the j-th classes is defined as:

$$MD_{ij} = \sqrt{\left(\mu_i-\mu_j\right)^T \Sigma^{-1} \left(\mu_i-\mu_j\right)}, \ \Sigma = \frac{N_i}{N_i+N_j}\mathbf{C}_i + \frac{N_j}{N_i+N_j}\mathbf{C}_j \tag{22}$$

where N_i and N_j are the number of data sets in the i-th and the j-th classes, and $\mathbf{C}_i$ and $\mathbf{C}_j$ represent the covariance matrices of the characteristic vectors in the i-th and the j-th classes, respectively. Apparently, the characteristic vectors consisting of multi-features of the data sets and the

corresponding covariance matrices are included, therefore, the MD measurement can statistically depict the distance between the two classes of characteristic vectors with variations.

It is noted that the FS formulates the distinguishability of individual feature for data sets of two classes, while the MD reflects the distinguishability of different combinations of features for data sets of two classes. Therefore, even though the features of highest FS are selected, the combination of the selected features may not have the most distinguishability for the data sets of two classes if the combination of the selected features does not have the highest MD quantity.

4. Support Vector Machine

The SVM aims to find the optimal plane, termed as the hyper plane, such that the data sets of two different classes can be separated by the hyper plane. The concept of SVM is illustrated as shown in Figure 1 [47]. This figure shows that the hyper plane is placed between the data sets of two different classes, circles and triangles, and is oriented in such a way that the margin between the date sets of two classes is maximized. Namely, the distance between the hyper plane and the nearest data point in each class is maximal. The nearest data points that are utilized for defining the margin form the support vectors.

Figure 1. Conceptual illustration of SVM.
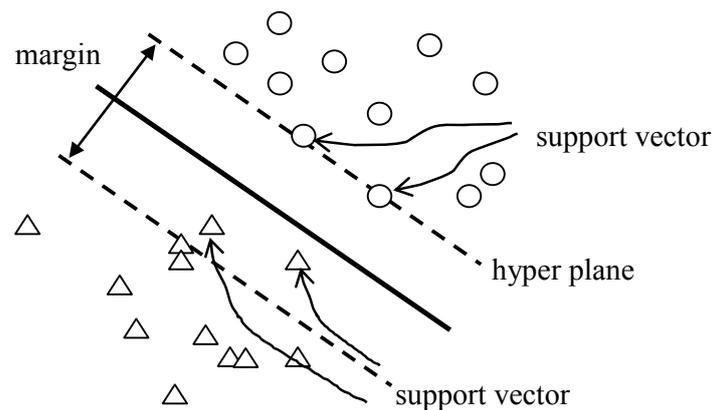
Let $\{(x_k, y_k), k=1,2,...,N\}$ be the given data sets as the training sample. Each sample data $x_k \in \mathbf{R}^d$ is distinguished as the *i*-class or *j*-class according to the value of y_k ($y_k \in \{+1, -1\}$). The hyper plane can be formulated as:

$$w \cdot x + b = 0 \tag{23}$$

where w represents the weighting vector and b is the bias. The data point x_k is thus distinguished as the class *i* or *j* according to the value of decision function, that is:

$$D(x_k) = sign(w \cdot x_k + b) \tag{24}$$

Therefore, the SVM is equivalent to finding the solution of the constrained optimization problem:

$$J = \frac{1}{2} \min_{w,b} w^T w \tag{25}$$

which is subjected to the constraint:

$$y_k \left(w \cdot x_k + b \right) - 1 \geq 0, \, k=1, 2,..., N \tag{26}$$

This standard optimization problem can be solved through the quadratic programming process.

While the linear hyper plane is incapable of separating the data sets effectively in the original space, the nonlinear SVM can be applied to classify the data sets of classes in a higher dimension through the space transformation that is achieved by the kernel function mapping procedure. The commonly used kernel functions include homogenous polynomial function, inhomogeneous polynomial function, Gaussian radial basis function, and sigmoid function [8].

The fundamental theory of SVM as stated above is developed and based on the classification of data sets in two classes. In general, it is necessary to classify the data sets of more than two classes in most practical applications, therefore, the three structures of SVMs, including one-against-one SVM, one-against-all SVM and hierarchical SVM, are the well-known classification tools that are frequently utilized to separate the data sets of N different classes. Among the three different structures of SVMs, the hierarchical SVM requires the least number of classifiers: at most N-1 classifiers are needed to separate the data sets of N different classes. Although the one-against-all SVM requires N classifiers to distinguish the data sets of N classes, the parallel computation technique can be employed to concurrently execute the classification of N classifiers, and hence the computing time of classification is not affected by the class number of data sets. However, the drawback of the above two SVM structures is that the SVMs require relatively more time and data sets for training procedures inevitably. The one-against-one SVM, on the other hand, requires more classifiers ($N!/[2!(N$-2$)!]$) than the above two structures. Nevertheless, the classification accuracy of one-against-one SVM is more reliable than the above two SVMs based on test experience [48]. Therefore, the diagnosis results in this research is based on the classification of the one-against-one SVM.

5. Experimental Validation

In order to verify the proposed approach for identifying the bearing defects, the experimental vibration data of the Bearing Center at Case Western Reserve University (CWRU) [35] are utilized in this research to evaluate the effectiveness of the proposed method. The test rig of this experiment is shown in Figure 2. The fault conditions of the bearing consist of four classes: (1) normal condition; (2) inner race defect; (3) rolling element defect; (4) outer race defect. The defects are artificially produced on the bearing of the drive end at the test rig by drilling holes of 7, 14 and 21 mils in the corresponding defective locations respectively to simulate the different levels of bearing defect. The associated rotating speed of driving motor is set to be 1,730, 1,750, 1,772 and 1,797 RPM, respectively. The accelerometer is placed at the driving motor to measure the vibration signals in different test conditions. The sampling rate of the data acquisition device is set to 48 KHz and the recorded data length is set to 2048 data points. The numbers of data sets corresponding to the different faulted classes, defective levels and rotation speeds are shown in Table 1.

Figure 2. Picture of the bearing fault test rig.
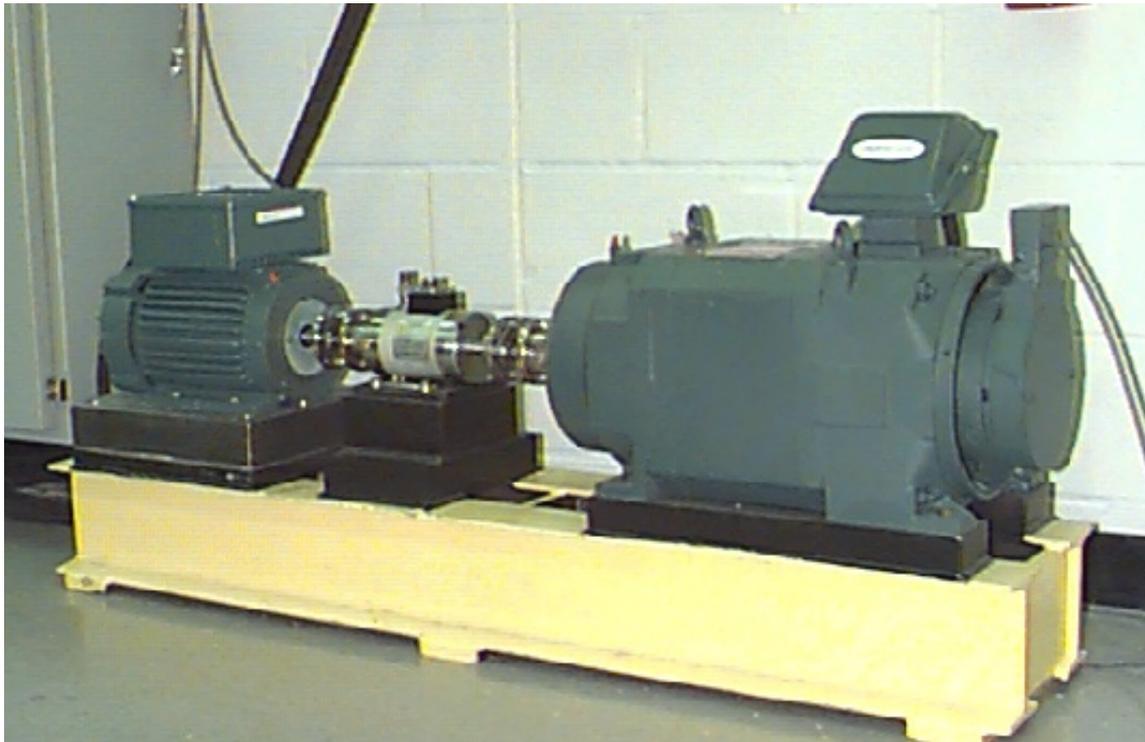
Table 1. Numbers of data sets corresponding to different faulted classes, defective levels and rotation speeds.

|	Rotation Speed (rpm)											
Shaft Speed / Defect Level	1730			1750			1772			1797		
	Diameter of defective hole (mm)											
Fault Class	7	14	21	7	14	21	7	14	21	7	14	21
Normal		237			236			236			119	
Ball	238	237	237	237	237	237	237	237	237		121	119
Inner race	237	236	238	237	238	239	237	186	236		31	119
Outer race (3)	237		236	237		237	236		239			62
Outer race (6)	238	238	238	237	237	238	237	236	238		119	120
Outer race (12)	236		237	235		237	235		237			63

Fault class: Outer race (3): Outer race defect at location of 3 o'clock; Outer race (6): Outer race defect at location of 6 o'clock; Outer race (12): Outer race defect at location of 12 o'clock;

The collected vibration data sets are processed through the coarse-grain procedure for multi-scale analysis. The MSE, MPE, MSRMS and MBSE of the vibration signals are then calculated respectively for different scales. In this research, the aforementioned quantities in twenty scales are selected as the features for bearing fault diagnosis. The FS and MD of the corresponding features are determined to evaluate the distinguishabilities of the features between data sets of different classes. According to the values of FS and MD, the features of high distinguishabilities are selected and fed into the SVM for fault classification, so that the computational efforts can be decreased and the classification accuracy can be enhanced simultaneously.

In this research, the one-against-one SVM model with a radial basis function as the hyper plane is utilized to classify the faulted features. For comparison, ten and twenty percent of the collected vibration data sets are randomly selected for training the SVM model, respectively. The remaining ninety (eighty) percent of data sets are then utilized for classification test. The statistical results are obtained by repeating the above procedures 100 times.

Figure 3 shows the diagnostic accuracy results of selecting different features through the FS and MD evaluations respectively, in which ten percent of the collected vibration data sets are used for training and the remaining ninety percent of data sets are used for fault diagnosis. It is observed that the highest classification accuracy can be achieved while the fifteen of the most distinguishable features are selected through the FS evaluation. On the other hand, the most accurate diagnosis result can be obtained by selecting the combination of 10 features through the MD evaluation. Table 2 and 3 show the contingency tables of classification results that the FS and MD evaluations for feature selection are employed respectively. The classification results in which the total 80 features (20 scales for MSE, MPE, MSRMS and MBSE, respectively) are utilized for fault diagnosis are also displayed in the tables for comparison. As observed in these two tables, it can be inferred that the accuracy of bearing fault diagnosis can be enhanced by using the evaluation of feature distinguishability. Concurrently, the computational efforts can be certainly reduced with the features that are selected by the FS evaluation as well as the MD evaluation. Additionally, the classification results in the tables also indicate that the diagnostic accuracy of utilizing the MD evaluation outperform the results of utilizing the FS evaluation. However, it is a dilemmatic problem since the computational effort of MD evaluation is much higher than that of FS evaluation.

Figure 3. Diagnostic accuracy with FS and MD evaluations (10% data sets for SVM training).
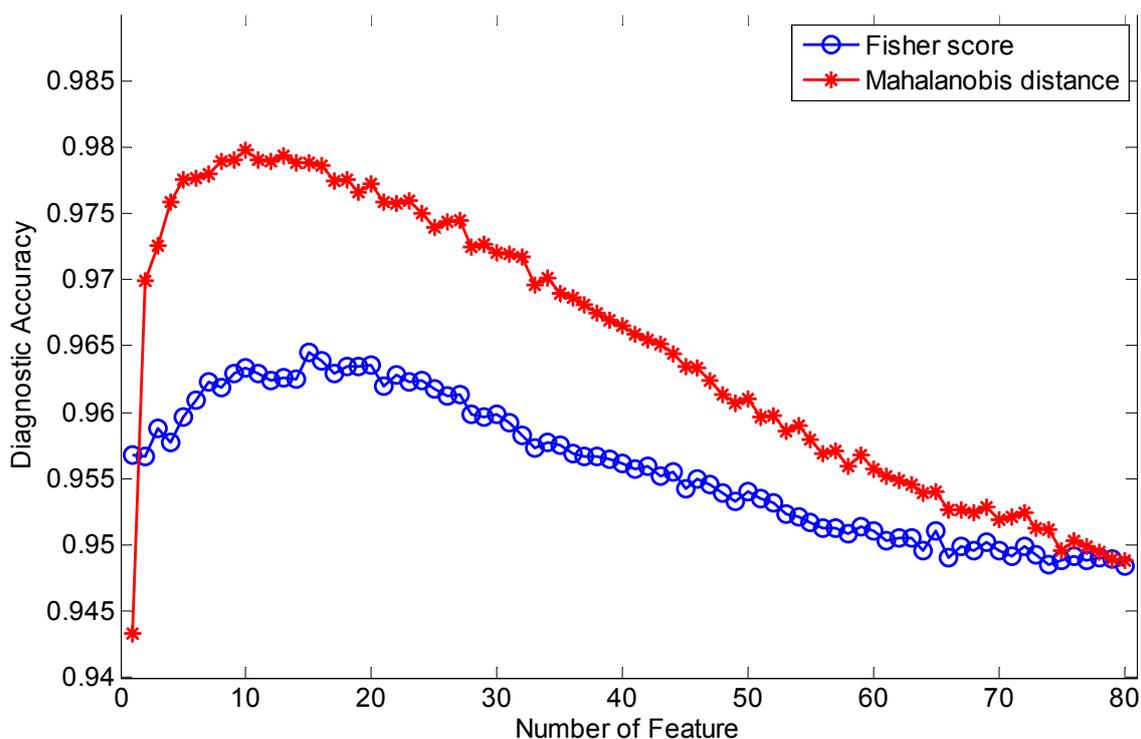
Table 2. Contingency table of classification results with FS evaluations (10% data sets for SVM training).

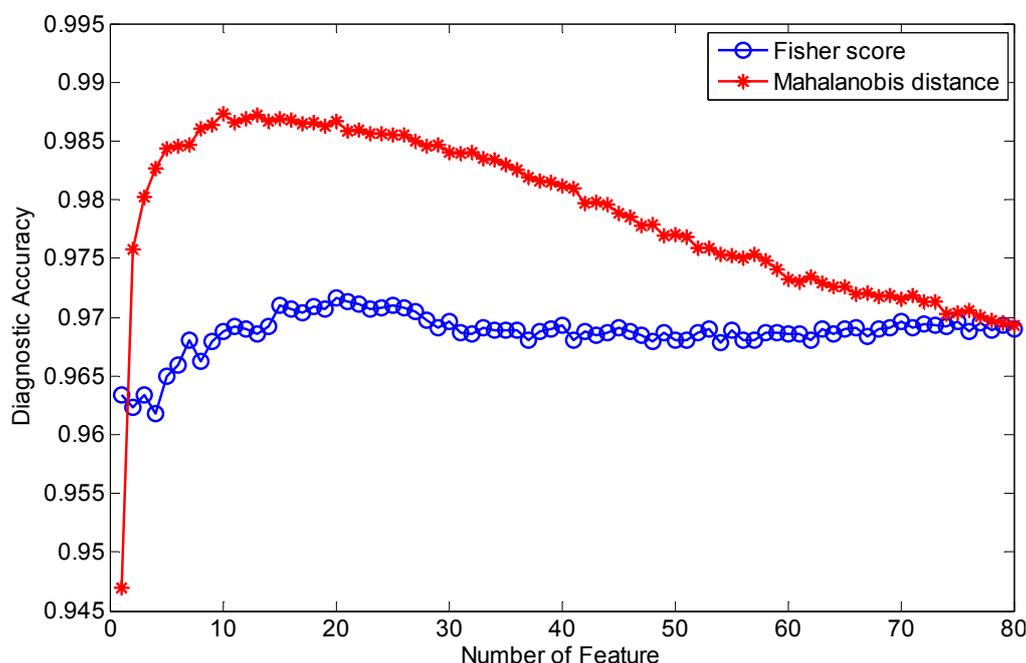| Classification Result	Fifteen Selected Features Through FS Evaluation.			
	Actual Class			
	Normal	Ball	Inner Race	Outer Race
Normal	98.31%	1.69%	0.00%	0.00%
Ball	0.06%	95.04%	2.70%	2.21%
Inner race	0.00%	1.25%	97.73%	1.02%
Outer race	0.00%	1.58%	0.41%	98.02%
Classification Result	Total of 80 Features Without FS Evaluation			
	Actual Class			
	Normal	Ball	Inner Race	Outer Race
Normal	95.59%	4.41%	0.00%	0.00%
Ball	0.05%	94.42%	2.30%	3.23%
Inner race	0.00%	2.71%	93.73%	3.55%
Outer race	0.00%	1.78%	1.00%	97.22%

Table 3. Contingency table of classification results with MD evaluations (10% data sets for SVM training).

Classification Result	Ten Selected Features Through MD Evaluation			
	Actual Class			
	Normal	Ball	Inner Race	Outer Race
Normal	99.57%	0.42%	0.01%	0.00%
Ball	0.00%	96.86%	1.49%	1.65%
Inner race	0.00%	0.85%	98.33%	0.82%
Outer race	0.00%	0.93%	0.26%	98.81%
Classification Result	Total of 80 Features Without MD Evaluation			
	Actual Class			
	Normal	Ball	Inner Race	Outer Race
Normal	95.56%	4.44%	0.00%	0.00%
Ball	0.05%	94.47%	2.21%	3.27%
Inner race	0.00%	2.63%	93.83%	3.54%
Outer race	0.00%	1.78%	0.99%	97.22%

In order to further investigate the effect of utilizing different numbers of data sets for training the SVM model, twenty percent of the collected vibration data sets are randomly selected for training the SVM model and the remaining eighty percent of data sets are utilized for fault classification test. Figure 4 shows the diagnostic accuracy results when the FS and MD evaluations are used to select the different features, respectively. Similarly, the highest classification accuracy can be achieved when the twenty most distinguishable features are selected through the FS evaluation and the combination of ten features is utilized through the MD evaluation. The classification results of utilizing all the eighty features for bearing fault diagnosis are also displayed for comparison. Tables 4 and 5 show the contingency tables of the classification results. A similar conclusion can be inferred from the results by observing the tables: the diagnostic accuracy can be enhanced when either the FS or MD evaluation is

executed before the SVM classification process. It is also noted from Tables 2 to 5 that the diagnostic accuracy of the bearing faults can be improved if more data sets are utilized for SVM model training. In other words, a more accurate SVM model can achieve more accurate bearing fault classification results. In conclusion, the feature selection procedure is beneficial for bearing fault diagnosis: both in increasing the classification accuracy and in decreasing the computational efforts. The MD evaluation can render more accurate results of bearing fault diagnosis than the FS evaluation although the MD evaluation procedure needs higher computational efforts than the FS evaluation procedure. Additionally, an accurate diagnostic result of bearing faults can be obtained if the trained SVM model is precise.

Figure 4. Diagnostic accuracy with FS and MD evaluations (20% data sets for SVM training).

Table 4. Contingency table of classification results with FS evaluations (20% data sets for SVM training).

Twenty Selected Features Through FS Evaluation			
Classification Result	Actual Class		
	Normal	Ball	Inner Race	Outer Race
Normal	98.62%	1.38%	0.00%	0.00%
Ball	0.03%	96.45%	1.74%	1.78%
Inner race	0.00%	1.12%	97.92%	0.96%
Outer race	0.00%	1.38%	0.46%	98.16%
Total of 80 features without FS evaluation			
Classification Result	Actual Class		
	Normal	Ball	Inner Race	Outer Race
Normal	98.41%	1.59%	0.00%	0.00%
Ball	0.05%	96.65%	1.40%	1.90%
Inner race	0.00%	2.00%	96.01%	2.00%
Outer race	0.00%	1.15%	0.59%	98.26%

Table 5. Contingency table of classification results with MD evaluations (20% data sets for SVM training).

Ten Selected Features Through MD Evaluation				
Classification Result	Actual class			
	Normal	Ball	Inner race	Outer race
Normal	99.79%	0.21%	0.00%	0.00%
Ball	0.00%	97.91%	1.22%	0.87%
Inner race	0.00%	0.54%	99.00%	0.45%
Outer race	0.00%	0.60%	0.13%	99.27%
Total of 80 Features Without MD Evaluation				
Classification Result	Actual class			
	Normal	Ball	Inner race	Outer race
Normal	98.54%	1.46%	0.00%	0.00%
Ball	0.03%	96.59%	1.47%	1.91%
Inner race	0.00%	1.98%	96.05%	1.96%
Outer race	0.00%	1.15%	0.61%	98.24%

6. Conclusions

In this paper, multi-scale analysis is utilized to extract the features for detecting the different defects of bearings. Since the vibration signals of the machine normally consist of plentiful information related to the dynamic characteristics of the machine, the MSE, MPE, MSRMS and MBSE can be used to represent the fault features of the machine in different scales. The FS and MD evaluations of the feature selection procedure are proposed to enhance both the accuracy of bearing fault classification and the computational efficiency. The analysis results demonstrate that the proposed approach can achieve the accurate bearing fault diagnosis effectively.

Acknowledgement

References

1. Randall, R.B. State of the art in monitoring rotating machinery-Part 1. *Sound Vib.* **2004**, *38*, 14–21.
2. Randall, R.B. State of the art in monitoring rotating machinery-Part 2. *Sound Vib.* **2004,** *38*, 10–17.
3. McFadden, P.D.; Smith, J.D. Vibration monitoring of rolling element bearings by the high-frequency resonance technique—A Review. *Tribol. Int.* **1984**, *17*, 3–10.
4. McInerny, S.A.; Dai, Y.; Basic Vibration signal processing for bearing fault detection, *IEEE Trans. Educ.* **2003**, *46*, 149–156.
5. Ho, D.; Randall, R.B. Optimization of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mech. Syst. Signal. Process.* **2000**, *14*, 763–788.

6. Randall, R.B.; Amtoni, J.; Chobsaard, S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mech. Syst. Signal. Process.* **2001**, *15*, 945–962.

7. Antoni, J.; Randall, R.B. A stochastic model for simulation and diagnostics of rolling element bearings with localized faults. *Trans. ASME J. Vib. Acoust.* **2003**, *125*, 282–289.

8. Yang, Y.; Yu, D.; Cheng, J. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement* **2007**, *40*, 943–950.

9. Cheng, J.; Yu, D.; Yang, Y. Application of an impulse response wavelet to fault diagnosis of rolling bearings. *Mech. Syst. Signal. Process.* **2007**, *21*, 920–929.

10. Peng, Z.K.; Tse, P.W.; Chu, F.L. A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mech. Syst. Signal. Process.* **2005**, *19*, 974–988.

11. Li, H.; Zhang, Y.; Zheng, H. Bearing fault detection and diagnosis based on order tracking and Teager-Huang transform. *J. Mech. Sci.Technol.* **2010**, *24*, 811–822.

12. Saravanan, N.; Kumar Siddabattuni, V.N.S.; Ramachandran, K.I.; A comparative study on classification of features by SVM and PSVM extracted using morlet wavelet for fault diagnosis of spur bevel gear box. *Expert Syst. Appl.* **2008**, *35*, 1351–1366.

13. Saravanan, N.; Ramachandran, K.I. Fault diagnosis of spur bevel gear box using discrete wavelet features and decision tree classification. *Expert Syst. Appl.* **2009**, *36*, 9564–9573.

14. Yu, D.; Cheng, J.; Yang, Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mech. Syst. Signal. Process.* **2005**, *19*, 259–270.

15. Wu, T.Y.; Chung, Y.L. Misalignment diagnosis of rotating machinery through vibration analysis via hybrid EEMD and EMD approach. *Smart Mater. Struct.* **2009**, *18*, 095004.

16. Wu, T.Y.; Chung, Y.L.; Liu, C.H. Looseness diagnosis of rotating machinery via vibration analysis through Hilbert-Huang transform approach. *Trans. ASME J. Vib. Acoust.* **2010**, *132*, 031005.

17. Wu, T.Y.; Chen, J.C.; Wang, C.C. Characterization of gear faults in variable rotating speed using Hilbert-Huang transform and instantaneous dimensionless frequency normalization. *Mech. Syst. Signal. Process.* **2012**, *30*, 103–122.

18. Yan, R.; Gao, R.X. Approximate entropy as a diagnosis tool for machine health monitoring. *Mech. Syst. Signal. Process.* **2007**, *21*, 824–839.

19. Pen, Y.N.; Chen, J.; Li, X.L. Spectral Entropy: A complementary index for rolling element bearing performance degradation assessment. *Proc. Inst. Mech. Eng. Part C—J. Mech. Eng. Sci.* **2009**, *223*, 1223–1231.

20. Hao, R.; Peng, Z.; Feng, Z.; Chu, F. Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings. *Meas. Sci. Technol.* **2011**, *22*, 045708.

21. Yu, D.; Yang, Y.; Cheng, J. Application of time frequency entropy method based on Hilbert-Huang transform to gear fault diagnosis. *Measurement* **2007**, *40*, 823–830.

22. Lei, Y.G.; Zuo, M.J.; He, Z.J.; Zi, Y.Y. A multidimensional hybrid intelligent method for gear fault diagnosis. *Expert Syst. Appl.* **2010**, *37*, 1419–1430.

23. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102–1.

24. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E.* **2005**, *71*, 021906.

25. Zhang, L.; Xiong, G.; Liu, H.; Zou, H.; Guo, W. Bearing fault diagnosis using multi-scale entropy and adaptive Neuro-Fuzzy inference. *Expert Syst. Appl.* **2010**, *37*, 6077–6085.

26. Zhang, L.; Xiong, G.; Liu, H.; Zou, H.; Guo, W. Applying improved multiscale entropy and support vector machines for bearing health condition identification. *Proc. Inst. Mech. Eng. Part C—J. Mech. Eng. Sci.* **2010**, *224*, 1315–1325.

27. Lin, J.L.; Liu, J.Y.C.; Li, C.W.; Tsai, L.F.; Chung, H.Y. Motor shaft misalignment detection using multiscale entropy with wavelet denoising. *Expert Syst. Appl.* **2010**, *37*, 7200–7204.

28. Litak, G.; Syta, A.; Rusinek, R. Dynamical changes during composite milling: Recurrence and multiscale entropy analysis. *Int. J. Adv. Manuf. Technol.* **2011**, doi 10.1007/s00170–011–3195–8.

29. 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.

30. Sun, W.X.; Chen, J.; Li, J.Q. Decision tree and PCA-based fault diagnosis of rotating machinery. *Mech. Syst. Signal. Process.* **2007**, *21*, 1300–1317.

31. Li, Z.X.; Yan, X.P.; Yuan, C.Q.; Peng, Z.X.; Li, L. Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. *Mech. Syst. Signal. Process.* **2011**, *25*, 2589–2607.

32. Chen, P.; Toyota, T.; He, Z.J. Automated function generation of symptom parameters and application to fault diagnosis of machinery under variable operating conditions. *IEEE Trans. Syst. Man Cybern. Part A—Syst. Hum.* **2001**, *31*, 775–781.

33. Kao, W.C.; Hsu, M.C.; Yang, Y.Y. Local contrast enhancement and adaptive feature extraction for illumination-invariant face recognition. *Pattern Recogn.* **2010**, *43*, 1736–1747.

34. Wu, T.Y.; Hong, H.C.; Chung, Y.L. A looseness identification approach for rotating machinery based on post-processing of ensemble empirical mode decomposition and autoregressive modeling. *J. Vib. Control.* **2012**, *18*, 796–807.

35. Case western reserve university bearing data center website, Available online: http://www.eecs.case.edu/laboratory/bearing/welcome_overview.htm (accessed on 5 May 2011).

36. Shannon, E.C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.

37. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.—Heart Circul. Physiol.* **2000**, *278*, H2039-H2049.

38. Powell, G.E.; Percival, I.C. A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems. *J. Phys. A—Math. Gen.* **1979**, *12*, 2053–2071.

39. Pen, Y.N.; Chen, J.; Li, X.L. Spectral Entropy: A complementary index for rolling element bearing performance degradation assessment. *Proc. Ins. Mech. Eng. Part C—J. Mech. Eng. Sci.* **2009**, *223*, 1223–1231.

40. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102:1–174102:4.

41. Olofsen, E.; Sleigh, J.W.; Dahan, A. permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect. *Brit. J. Anaesth.* **2008**, *101*, 810–821.

42. Li, X.; Cui, S.; Voss, L.J. Using permutation entropy to measure the electroencephalographic effects of sevoflurane. *Anesthesiology* **2008**, *109*, 448–456.

43. Yan, R.; Liu, Y.; Gao, R.X. Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. *Mech. Syst. Signal. Process*. **2012**, *29*, 474–484.

44. Costa, M.; Peng, C.K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Physica. A* **2003**, *330*, 53–60.

45. Aziz, W.; Arif, M. Multiscale permutation entropy of physiological time series. In the 9th International Multitopic Conference, IEEE Inmic, Karachi, Pakistan, 24–25 December, 2005.

46. Mahalanobis, P.C. On the generalized distance in statistics. *Proc. Natl. Ins. Sci. India*. **1936**, *2*, 49–55.

47. Vapnik, V.N. Statistical Learning Theory; John Wiley & Sons: New York, NY, USA, 1998.

48. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Networ*. **2002**, *13*, 415–425.