applied sciences

MDPI

Article

A Novel Rolling Bearing Fault Diagnosis and Severity Analysis Method

Yinsheng Chen [1], Tinghao Zhang [2], Zhongming Luo [1,*] and Kun Sun [1]

[1] The Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, Harbin 150001, China; chen_yinsheng@126.com (Y.C.); Sunkun1982@126.com (K.S.)

[2] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China; zhangtinghaohit@163.com

* Correspondence: luozhongming@hrbust.edu.cn; Tel.: +86-0451-86392308

Received: 26 May 2019; Accepted: 4 June 2019; Published: 8 June 2019

Abstract: To improve the fault identification accuracy of rolling bearing and effectively analyze the fault severity, a novel rolling bearing fault diagnosis and severity analysis method based on the fast sample entropy, the wavelet packet energy entropy, and a multiclass relevance vector machine is proposed in this paper. A fast sample entropy calculation method based on a kd tree is adopted to improve the real-time performance of fault detection in this paper. In view of the non-linearity and non-stationarity of the vibration signals, the vibration signal of the rolling bearing is decomposed into several sub-signals containing fault information by using a wavelet packet. Then, the energy entropy values of the sub-signals decomposed by the wavelet packet are calculated to generate the feature vectors for describing different fault types and severity levels of rolling bearings. The multiclass relevance vector machine modeled by the feature vectors of different fault types and severity levels is used to realize fault type identification and a fault severity analysis of the bearings. The proposed fault diagnosis and severity analysis method is fully evaluated by experiments. The experimental results demonstrate that the fault detection method based on the sample entropy can effectively detect rolling bearing failure. The fault feature extraction method based on the wavelet packet energy entropy can effectively extract the fault features of vibration signals and a multiclass relevance vector machine can identify the fault type and severity by means of the fault features contained in these signals. Compared with some existing bearing rolling fault diagnosis methods, the proposed method is excellent for fault diagnosis and severity analysis and improves the fault identification rate reaching as high as 99.47%.

Keywords: rolling bearing; fault diagnosis; fault severity; sample entropy; wavelet packet energy entropy; multiclass relevance vector machine

1. Introduction

Rolling bearings, one of the important parts of rotating machinery, reduce the friction loss between mechanical components. Moreover, rolling bearings are a highly standardized precision mechanical device with the advantages of low friction, convenient assembly and use, and high working efficiency. As the connection between rotating parts, rolling bearings also serve as supporting parts and bear certain loads. Therefore, it is inevitable that these components suffer from a series of physical effects, such as mechanical stress and mechanical wear, which over time cause deformation and corrosion of the bearings, resulting in the degradation of their working performance. Rolling bearings are one of the most vulnerable parts of mechanical equipment because of their poor impact capacity [1]. Many failure modes of mechanical equipment, especially rotating machinery, are closely related to rolling bearing

failure. It is of great practical significance to carry out research on the condition monitoring and fault diagnosis technology of rolling bearings to effectively improve the reliability of mechanical equipment operations [2]. Many rolling bearing fault-diagnosis-related methods have been proposed in succession. Ocak et al. developed a scheme based on wavelet packet decomposition and hidden Markov modelling (HMM) for tracking the severity of bearing faults [3]. In this scheme, the wavelet packet decomposed vibration signals into a series of sub-signals and the energies of the sub-signals were extracted as features. The probabilities of HMM trained by features were used to track the condition of rolling bearings. Vafaei et al. proposed a methodology termed indicated repeatable runout with wavelet decomposition that represents a novel approach for the determination of specific sources of vibration [4]. Guo et al. presented a signal compression method based on ensemble empirical mode decomposition (EEMD) [5]. On the basis of ensuring the accuracy of the fault diagnosis, the burden of wireless transmission was reduced.

Given their importance, the condition monitoring and fault diagnosis of rotating machinery are always a research focus and have attracted the attention of numerous scholars for years [6]. The model-based fault diagnosis technology is formed with the idea of analytic redundancy. The advantage of this method is that it can delve into the essence of a dynamic system and detect and diagnose faults in real time. The difficulty of its implementation lies in the need for an accurate system model. Model-based fault detection methods use residuals to indicate changes between the process and the model. One general assumption is that the residuals are changed significantly so that a detection is possible. The model-based method is often used for fault detection, but the analysis of the fault type and fault severity is insufficient [7,8]. The vibration-based signal processing technique is one of the principal tools for diagnosing the malfunctions of rolling bearings [9]. Generally, rolling bearing fault diagnoses based on vibration analysis methods consist of two main parts: The feature extraction of vibration signals and the pattern recognition of faults [10]. First, the performance of the feature extraction method is the key point of rolling bearing fault diagnosis. Current feature extraction methods of vibration signals can be classified into three categories: Time domain feature extraction, frequency domain feature extraction, and time–frequency feature extraction. The vibration signals of faulty rolling bearing exhibit non-stationarity and non-linearity, leading to complex feature extraction processes and low separability of feature extraction results. In recent years, time–frequencyanalysis has become the main feature extraction method of rolling bearing fault diagnosis, predominating over the other feature extraction methods [11–13]. Second, an appropriate classification algorithm should be adopted to identify the fault types on the basis of accurately extracting the fault features of rolling bearings. The adopted classification algorithm should have good generalization, multi-classification performance, and a simple model structure [14,15]. Obviously, the selection of an appropriate pattern recognition algorithm is another key point in determining the recognition rate of rolling bearing fault diagnoses.

The vibration signal of the rolling bearing is composed of the vibration of the rolling bearing itself, the vibration caused by other moving parts of the electrical machinery, and the external disturbances. These vibration signals are superimposed on each other, resulting in extremely complex vibration signals acquired by the vibration sensors. To effectively extract the fault features from the non-linear and non-stationary vibration signals, many time–frequency analysis methods have been applied to rolling bearing fault diagnoses for vibration signal decomposition. The decomposed fault signal is helpful for further highlighting the fault characteristics of the vibration signal. The fast Fourier transform (FFT) is a classical time–frequency analysis method, but it is inherently unsuitable for non-stationary and transient signal analysis. The window shape and size of the short-time Fourier transform (STFT) is fixed for all frequencies, which causes the STFT to have obvious limitations regarding time-varying signals. In contrast to the FFT and STFT, the wavelet transform (WT) is a typical time–frequency analysis technique that can change both the time and frequency windows, thus enabling a multi-scale analysis of the vibration signal. However, the frequency resolution decreases with increasing frequency. Therefore, the WT cannot effectively split the high-frequency

band containing rich fault modulation information [16]. Empirical mode decomposition (EMD) is a self-adaptive time–frequency decomposition technique that can self-adaptively decompose a non-stationary non-linear signal into a series of intrinsic mode functions (IMFs) that include different frequency characteristics [1]. Nevertheless, EMD has an end effect and mode mixing problem that directly influences the stability of the decomposition results and the subsequent feature extraction effects [17]. To improve the above disadvantages, many derivative methods of EMD, such as EEMD and complete ensemble empirical mode decomposition (CEEMD) are presented. Local mean decomposition (LMD) can automatically decompose multi-component amplitude-modulated (AM) and frequency-modulated (FM) signals into a series of product functions (PFs), each of which is a mono-component AM–FM signal that contains the signal characteristics. However, as with the EMD, the LMD suffers from the problem of edge effects caused by the local extreme points [18]. The wavelet packet transform (WPT) has been introduced to overcome WT's application problems. The WPT is an extension of the WT that can select an arbitrary frequency resolution, which makes it relatively flexible [19]. Although some studies suggest that WT-based decomposition is not an adaptive method due to the need to preselect the wavelet kernel function, the WPT is still an effective time–frequency analysis method for the vibration signal characteristics of rotating machinery because of its complete theoretical basis and excellent decomposition results compared with those of EMD and LMD.

To further extract the fault features of the decomposed vibration signal by the time–frequency analysis method, many entropy-based feature extraction methods have been investigated for rolling bearing fault diagnosis, such as approximate entropy (ApEn), fuzzy entropy (FE), permutation entropy (PE), and sample entropy (SampEn). Zhang et al. adopted EEMD to decompose the vibration signal into a set of IMF and calculated the PE values of the first several IMFs to reveal the intrinsic characteristics of a faulty vibration signal [11]. Tiwari et al. used the multi-scale permutation entropy to calculate the PE values at different time scales and generate the fault feature vector of a rolling bearing vibration signal [20]. Shi et al. proposed a feature extraction method of vibration signals based on improved LMD and PE to highlight the fault characteristics [18]. Li et al. presented a fault feature extraction method based on the hierarchical fuzzy entropy for extracting the characteristics of a non-linear and non-stationary vibration signal in complex operating conditions [15]. Although all of the above methods have had achievements in rolling bearing fault diagnosis, the accuracy of these fault diagnosis methods needs to be improved. In particular, the description of different fault severity levels is an urgent problem that needs to be solved. Therefore, a high-performance fault feature extraction method for rolling bearing should be effectively conducted before performing fault identification.

After feature extraction, an appropriate classification algorithm should be used to achieve fault type identification. As an effective predictor and classifier, neural network has been successfully applied in mechanical fault prediction and diagnosis [21]. The support vector machine (SVM) is a powerful classifier [22] that is currently one of the most widely used classifiers in rolling bearing fault diagnosis. However, the classification performance of SVMs is related to the setting of a penalty factor value. Improper setting of the penalty factor leads to an over-fitting problem when the number of training samples is large. A linear increase in the number of support vectors leads to a significant increase in the test time. The relevance vector machine (RVM) introduced by Tipping in 2001 has the same function as the SVM, but it is a new type of machine-learning method based on the sparse Bayesian framework [23]. The classification accuracy of the RVM is no less than that of the SVM. Compared with SVMs, RVMs have special advantages in solving the classification problem [24,25]; the penalty factor is assigned automatically in RVMs, not subjectively by users, the sparsity of RVMs is better than that of SVMs, making them more suitable for online monitoring, and the kernel functions of RVM algorithms are no longer constrained by the Mercer theorem. It is worth mentioning that the RVM provides a probabilistic interpretation of its outputs. However, similar to the SVM, the RVM is also a binary classification algorithm; when dealing with multiclass issues, it usually transforms them into a series of binary classification problems, which may cause an accumulative error and overlapping classification. A multiclass relevance vector machine (mRVM) is proposed by Psorakis, which extends

the original RVM to multiclass issues by introducing auxiliary variables, which act as intermediate regression targets, thus realizing the classification of multiclasses by outputting the estimation of class membership probabilities [26,27]. Lei et al. applied an mRVM to the on-line fault diagnosis of an analogue circuit and obtained promising results [28]. Xu et al. presented a novel fault diagnosis strategy based on principal component analysis (PCA) and mRVM, and the simulation results have validated that the mRVM is useful in multi-fault diagnosis, producing results superior to those of the SVM [29]. The above studies have shown that the mRVM has excellent multi-classification performance and can provide a very good tool for the fault diagnosis and severity analysis of rolling bearings.

Through the above analysis, a novel fault diagnosis and severity analysis method based on fast sample entropy (SampEn), wavelet packet energy entropy (WPEE), and multiclass relevance vector machine (mRVM) is presented for rolling bearings. The SampEn can effectively detect the dynamic changes of the vibration signal and can be applied to detect the rolling bearing faults. The basic SampEn algorithm for computing the complexity of a signal is $O(N^2)$ and the calculation time is closely related to the length of the time series. With the growth of the time series length, the computing time grows almost exponentially and, therefore, this technique is not realistic for applications with long data sets. For that reason, a fast SampEn algorithm based on a kd tree is adopted to detect whether the rolling bearing contains faults or not in a timely manner. The WPEE is a feature extraction method that can be used in conjunction with wavelet packets and the Shannon entropy. The vibration signal is decomposed into several sub-signals by using a wavelet packet. Then, the energy entropy values of the sub-signals are calculated to generate the fault feature vector. The WPEE reflects the evenness of the signal energy distribution in various frequency bands. The WPEE can effectively describe the characteristics of a vibration signal and is a powerful tool for vibration signal analysis. In this paper, the proposed method extracts the fault features of bearing vibration signals by the WPEE and forms a feature vector for further fault identification and severity analysis. The mRVM is used to identify the fault type of rolling bearings and indicate the level of fault severity. The proposed fault diagnosis and severity analysis method is fully evaluated by experiments and comparative studies in this paper. The innovative contributions of this paper can be summarized as follows:

(1) This paper systematically presents a rolling bearing fault diagnosis and severity analysis method. The fault severity analysis of different fault types can be realized on the premise of accurate fault type identification.
(2) A fast sample entropy algorithm based on a kd tree is adopted to improve the real-time performance of the fault detection process for rolling bearings.
(3) The wavelet packet energy entropy is used to describe the characteristics and severity of different fault types. This approach can solve the problem of a low fault diagnosis accuracy caused by the weak separability of feature vectors extracted from non-stationary and non-linear vibration signals.
(4) A multiclass relevance vector machine is used to accurately classify different fault types and this tool is effectively extended to analyze the same fault with different fault severities.

The rest of this paper is organized as follows: Section 2 briefly introduces the fundamental theories of the SampEn and the fast algorithm of the SampEn. The feature extraction method based on the WPEE is detailed in Section 3. In Section 4, the fundamental theories of the mRVM are briefly described. Section 5 provides a description of the proposed rolling bearing fault diagnosis and severity analysis method. In Section 6, the effectiveness of the proposed method is demonstrated by experiments. Finally, conclusions are drawn in Section 7.

2. Fast Sample Entropy

2.1. Principle of Sample Entropy

The sample entropy is a measure of time series complexity proposed on the basis of the approximate entropy (ApEn) by Richman and Moorman in 2000 [30] and has been widely used in signal analysis.

The physical significance of the SampEn is the rate at which the non-linear dynamic system generates new information, which is consistent with the mechanism of mechanical failure.

The basic calculation steps of the SampEn are as follows:

(1) For a given time series $x(n), n = 1, 2, \ldots, N$, an m-dimensional vector can be composed of $x(n)$,

$$\boldsymbol{X}(i) = [x(i), x(i+1), \ldots, x(i+m-1)], \tag{1}$$

in which $i = 1, 2, \ldots, N-m+1$.

(2) The distance between any two vectors $\boldsymbol{X}(i)$ and $\boldsymbol{X}(j)$ in Equation (1) is the maximum coordinate difference:

$$d[\boldsymbol{X}(i), \boldsymbol{X}(j)] = \max[|x(i+k) - x(j+k)|], \tag{2}$$

where $k = 1, 2, \ldots, m-1$ and $i, j = 1, 2, \ldots, N-m+1$.

(3) For a given similarity tolerance r, count the number of differences between $\boldsymbol{X}(i)$ and other $N-m$ vectors that are less than r, the ratio of which to $N-m$ is as follows:

$$C_i^m(r) = \begin{cases} \frac{\sum_{j=1, j\neq i}^{N-m} \Theta(r - d[\boldsymbol{X}(i), \boldsymbol{X}(j)])}{(N-m)} & , i \leq N-m \\ 0 & , i > N-m \end{cases}, \tag{3}$$

in which Θ is the Heaviside function defined as

$$\Theta(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}. \tag{4}$$

(4) Perform the step (3) operations on $N-m+1$ vectors and calculate the average of all $C_i^m(r)$, written as

$$C^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} C_i^m(r). \tag{5}$$

(5) Increase the vector dimension to $m+1$, and repeat above steps to obtain $C^{m+1}(r)$.

(6) The SampEn value of time series $x(n)$ is expressed as

$$SampEn(m, r) = \lim_{N \to \infty} [-\ln \frac{C^{m+1}(r)}{C^m(r)}]. \tag{6}$$

When the time series length N is finite, Equation (6) is converted to

$$SampEn(m, r, N) = -\ln \frac{C^{m+1}(r)}{C^m(r)}. \tag{7}$$

It can be seen from the above calculation steps of the SampEn that the SampEn value of the time series depends on the similarity tolerance r, embedding dimension m and time series length N. The larger the value of m is, the more data are calculated for the SampEn and the longer the calculation time is. Generally, the time series length N should be equal to $10^m \sim 30^m$. Referring to the parameter setting of the ApEn [31], the settings of $m = 1$ or $m = 2$ and $r = 0.1 \sim 0.25std$ enable the SampEn to have good statistical properties [17].

2.2. Fast Algorithm of the Sample Entropy

Although the SampEn is an effective tool for time series analysis, the time computing complexity is $O(N^2)(m = 2)$, which makes SampEn not suitable for the occasions requiring analysis of a long time series or requiring high real-time performance. The literature [32] indicates that the calculation of $n_i^m = \sum_{j=1, j\neq i}^{N-m} \Theta(r - d[\boldsymbol{X}(i), \boldsymbol{X}(j)])$ can be transformed into an orthogonal range search problem.

The time series $\boldsymbol{X}(i)$ is denoted by $(\boldsymbol{X}_i)_m$, which is convenient to illustrate the orthogonal range search. The implementation process is outlined below:

For each element $i(i = 1, 2, \ldots, N)$ in the time series, $(\boldsymbol{X}_i)_m$ can be converted to the m-dimensional point set $\boldsymbol{P}_i(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i, \ldots)$,

$$\boldsymbol{x}_i = \boldsymbol{X}_i, \boldsymbol{y}_i = \boldsymbol{X}_{i+1}, \boldsymbol{z}_i = \boldsymbol{X}_{i+2}, \ldots. \tag{8}$$

Then, $\Theta(r - d[\boldsymbol{X}(i), \boldsymbol{X}(j)])$ is equal to the number of points contained within the bounding box W_i, which can be expressed as

$$W_i = [(x_{LB})_i : (x_{UB})_i] \times \left[(y_{LB})_i : (y_{UB})_i\right] \times [(z_{LB})_i : (z_{UB})_i] \times \cdots, \tag{9}$$

in which LB and UB, respectively, represent the lower and upper bounds of the bounding box W_i.

$$\begin{array}{l} (x_{LB})_i = x_i - r, (x_{UB})_i = x_i + r \\ (y_{LB})_i = y_i - r, (y_{UB})_i = y_i + r \\ (z_{LB})_i = z_i - r, (z_{UB})_i = z_i + r \end{array} \tag{10}$$

The number of points inside each bounding box queried by an orthogonal range in a d-dimensional space is called an orthogonal range search problem in computational geometry. Therefore, the calculation process of n_i^m is equivalent to an m-dimensional orthogonal range search for each point P_i and the corresponding boundary box W_i. Once n_i^m and n_i^{m+1} are calculated, $SampEn(m, r, N)$ can be obtained by Equation (7) directly and the time computing complexity depends on n_i^{m+1}.

The k-dimensional tree (kd tree) is a high-dimensional index tree data structure that is always used for the nearest neighbor domain search or approximate nearest neighbor domain search in large-scale high-dimensional data spaces. The kd tree can be used to resolve orthogonal range search problems to improve the computational efficiency of the SampEn [33]. The basic principle of the SampEn calculation based on the kd tree is to store the point set $\boldsymbol{P}$ in the kd tree structure, which has a faster query speed for the points inside the specified boundary box.

The main calculation process of the fast sample entropy based on the kd tree is summarized as follows:

(1) Coarse granulation of the time series $\boldsymbol{X}$ is carried out. When the maximum time scale is 1, it means that time series $\boldsymbol{X}$ has not been coarsely granulated;
(2) The time series is transformed into the discrete space point set $\boldsymbol{P}_i$, as shown in Equation (8);
(3) Set $k = m - 1$ and build the kd tree structure by using the discrete space point set $\boldsymbol{P}_i$;
(4) For each discrete space point, obtain the bounding box W_i through Equation (9);
(5) Obtain the number of discrete space points within bounding box W_i through the kd tree algorithm;
(6) Set $k = m$ and repeat steps (2) to (5), calculating $n_d(i)$;
(7) Utilize $n_n(i)$ and $n_d(i)$ to calculate $SampEn(m, r, N)$.

The time computing complexity of the above kd tree-based sample entropy calculation method is mainly composed of the spatial point transformation process, the kd tree establishment process, and the query process. According to the literature [34], the time computing complexity of the whole SampEn calculating process based on the kd tree is $O(N \bullet N^{1-(1/d)})$. When $m = 2$, the time computing complexity of the SampEn calculation method based on the kd tree is reduced to $O(N^{5/3})$.

3. Wavelet Packet Energy Entropy

The WT is increasingly used in engineering applications, especially in signal feature extraction. Unlike the FFT, which is frequency-localized, the WT is a suitable tool for detecting and characterizing specific phenomena in both the time and frequency spaces, but it does not split the high-frequency bands. The wavelet packet transform (WPT) is a further generalization of the WT and is an effective method for vibration signal processing because it can process the entire set of frequency bands, including both

the high- and low-frequency bands [13]. Notably, this coverage includes the high-frequency bands where the information of a rolling bearing fault always exists. A signal can be decomposed into a set of wavelet packet nodes with the form of a full binary tree by the WPT.

The structure diagram of the wavelet packet three-layer is shown in Figure 1, which can also be called the wavelet packet decomposition tree. In the binary tree structure, each wavelet packet node is indexed by (j, n) and the corresponding wavelet packet coefficient is defined as d_j^n, where j represents the decomposition levels and n indicates the node index in level j. $(j, n) = (0, 0)$ is the root of the decomposition tree corresponding to the original signal $S_{0,0}$. It is clear that the original signal $S_{0,0}$ splits into two parts in the first level during the WPT. The left branch undergoes low-pass filtering and a vector of approximation coefficients d_1^0 can be obtained. The right branch undergoes high-pass filtering and a vector of detail coefficients d_1^1 can be acquired. Using the same scheme, d_1^0 and d_1^1 are split into two parts, obtaining d_2^0, d_2^1, d_2^2, and d_2^3. If the decomposition level is j, the number of nodes after decomposition is 2^j. That is, the frequency band of the sampling signals is subdivided into 2^j small segments.

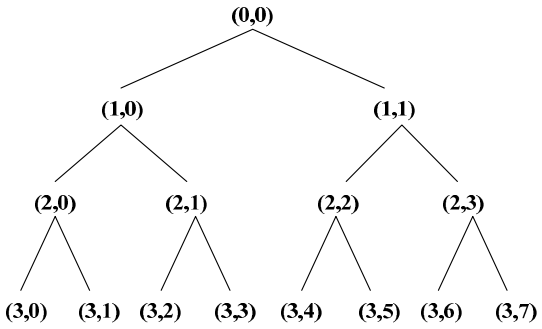

Figure 1. Structure diagram of wavelet packet three-layer decomposition.

The wavelet packet algorithm consists of two parts: The decomposition algorithm and the reconstruction algorithm. When the orthogonal scaling function $\phi(t)$ and the wavelet function $\psi(t)$ are set, the relationships between them are as follows:

$$\begin{cases} \phi(t) = \sqrt{2}\sum\limits_k h_k \phi(2t-k) \\ \psi(t) = \sqrt{2}\sum\limits_k g_k \phi(2t-k) \end{cases}, \tag{11}$$

where k is a transformation parameter, $h_k = 1/\sqrt{2}\left\langle \phi(t), \phi(2t-k) \right\rangle$ is the low-pass filter coefficient, and $g_k = 1/\sqrt{2}\langle \psi(t), \psi(2t-k) \rangle$ is the high-pass filter coefficient. Here, $\langle \cdot, \cdot \rangle$ represents the inner product operator.

The decomposition algorithm for the coefficients is defined by the following recursive relationships:

$$\begin{cases} d_{j+1}^{2n}[k] = \sqrt{2}\sum\limits_l h_{l-2k} d_j^n[k] \\ d_{j+1}^{2n+1}[k] = \sqrt{2}\sum\limits_l g_{l-2k} d_j^n[k] \end{cases}, \tag{12}$$

where $d_j^n[k]$ are the wavelet packet coefficients, $d_{j+1}^{2n}[k]$ are the approximation coefficients, $d_{j+1}^{2n+1}[k]$ are the detailed coefficients, and h_{l-2k} and g_{l-2k} are the low-pass and high-pass filter coefficients for decomposition, respectively.

After decomposition by the wavelet packet, the reconstruction algorithm for the wavelet packet coefficients is deduced as:

$$d_j^n[k] = \sum_l h_{k-2l} d_{j+1}^{2n}[k] + \sum_l g_{k-2l} d_{j+1}^{2n+1}[k]. \tag{13}$$

The corresponding reconstructed signal of wavelet packet coefficients d_j^n is defined as $S_{j,n}$. Assuming that the decomposition level $j = 2$, the original signal can be represented as:

$$S_{0,0} = d_0^0 = S_{2,0} + S_{2,1} + S_{2,2} + S_{2,3} \tag{14}$$

The reconfiguration signal of each node after wavelet packet decomposition reflects the corresponding frequency component in the distribution of the original signal. This information can be used to show the change of the frequency characteristic when applied in the fault diagnosis.

After the frequency band of the vibration signal with the fault has been subdivided into several small segments, the energy of the fault vibration signal changes with the frequency distribution. The wavelet packet energy entropy (WPEE) is proposed to illustrate this change.

The signal is decomposed by the *J*-level wavelet packet and its total energy is decomposed into *J* non-overlapping frequency intervals. Therefore, the energy of each decomposed segment can be regarded as the characteristic information of the original signal by measuring the distribution of the signal energy in each frequency band. Assume that (j, i) is the ith node in the j level and that the corresponding reconstructed signal is expressed as $S_{j,i}$, the energy of which is denoted as $E_{j,i}$. The energy calculation formula corresponding to node (j, i) is as follows:

$$E_{j,n} = \int |S_{j,n}(t)|^2 dt = \sum_{k=1}^{n} |S_{j,n}[k]|^2, \tag{15}$$

where j is the decomposition level, $n = [0, 1, \cdots, 2^j - 1]$ is the nth node on the jth level, and $S_{j,n}$ is the reconstructed signal.

The total energy of the original signal can be defined as:

$$E_j = \sum_{n=0}^{2^j-1} E_{j,n}. \tag{16}$$

The proportion of the reconstructed signal energy corresponding to each node in the total energy is

$$p_{j,n} = \frac{E_{j,n}}{E_j}, \tag{17}$$

in which $p_{j,n}$ represents the wavelet packet energy distribution in the relevant wavelet packet node.

Finally, according to the definition of the Shannon entropy, the WPEE can be calculated as follows:

$$Entropy_{j,n} = -p_{j,n} ln\left(p_{j,n}\right) \tag{18}$$

$$WPEE_j = -\sum_n Entropy_{j,n}. \tag{19}$$

The relative WPEE of each sub-signal can be calculated as follows:

$$WPEE_{j,n} = Entropy_{j,n} / WPEE_j. \tag{20}$$

4. Multiclass Relevance Vector Machine

4.1. Relevance Vector Machine

The relevance vector machine (RVM) is a machine-learning technique that uses Bayesian inference to provide probabilistic predictions [22]. RVMs achieve sparsity by producing models that have both a structure and a parameterization process.

There is a set of training data $\boldsymbol{T} = \{\boldsymbol{x}_i, t_i\}_{i=1}^{N}$, where $\boldsymbol{x}_i \in R^D$ is the input vector and t_i is the corresponding class label. The classification function of the RVM is defined as

$$y(\boldsymbol{x}, \boldsymbol{w}) = \sum_{i=1}^{N} w_i K(\boldsymbol{x}, \boldsymbol{x}_i) + w_0, \tag{21}$$

where $K(\boldsymbol{x}, \boldsymbol{x}_i)$ is a kernel function, $\boldsymbol{w} = [w_1, w_2, \cdots, w_N]$ is the weight vector, and w_0 is the bias.

The output target includes the additive noise, so the functional relationship between $\boldsymbol{x}_i$ and t_i is given as follows:

$$t_i = \sum_{i=1}^{N} W_i K(x, x_i) + w_0 + \varepsilon_i, \tag{22}$$

where ε_i is the ith independent sample from a noise process with zero mean and variance σ^2. According to the statistical convention, the linear model is generalized by applying the logistic sigmoid function $\sigma(y) = 1/(1 + e^{-y})$, so $y(\boldsymbol{x}, \boldsymbol{w})$ becomes a linear method. Given that $P(\boldsymbol{t}|\boldsymbol{x})$ is the Bernoulli distribution, its probability density function is $f(k) = p^k(1-p)^{1-k}(k = 0 \text{ or } 1)$. The likelihood estimation probability $P(\boldsymbol{t}|\boldsymbol{x})$ is written as

$$P(\boldsymbol{t}|\boldsymbol{w}) = \prod_{i=1}^{N} \sigma\{y(\boldsymbol{x}_i; \boldsymbol{w})\}^{t_i} [1 - \sigma\{y(\boldsymbol{x}_i; \boldsymbol{w})\}]^{1-t_i}. \tag{23}$$

However, the weights $\boldsymbol{w}$ cannot be analytically obtained in the process of classification because of the non-normal distribution of $P(\boldsymbol{t}|\boldsymbol{w})$. Thus, the closed-form expression $P(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha})$ or the marginal likelihood $P(\boldsymbol{t}|\boldsymbol{\alpha})$ cannot be deduced, where $\boldsymbol{\alpha}$ is a vector of $N + 1$ hyper-parameters. Therefore, a Laplace-based approximation procedure proposed by Mackay [31] is used as follows:

(1) For a fixed value of $\boldsymbol{\alpha}$, because $P(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}) \propto P(\boldsymbol{t}|\boldsymbol{w})P(\boldsymbol{w}|\boldsymbol{\alpha})$, find the maximum posteriori weight through the following equation:

$$\log\{P(\boldsymbol{t}|\boldsymbol{w})P(\boldsymbol{w}|\boldsymbol{\alpha})\} = \sum_{i=1}^{n} [t_i \log y_i + (1 - t_i) \log(1 - y_i)] - \frac{1}{2}\boldsymbol{w}^T \boldsymbol{A} \boldsymbol{w}, \tag{24}$$

where $y_i = \sigma\{y(\boldsymbol{x}_i; \boldsymbol{w})\}$ and $\boldsymbol{A} = diag(\alpha_0, \alpha_1, \cdots, \alpha_n)$ are composed of the current values of $\boldsymbol{\alpha}$.

(2) Differentiate the quantity above twice by applying Laplace's method to yield the following equation:

$$\nabla_w \nabla_w \log p(w|t, \boldsymbol{\alpha})|_{w_{MP}} = -(\boldsymbol{\Phi}^T B \boldsymbol{\Phi} + \boldsymbol{A}), \tag{25}$$

where $\boldsymbol{B} = diag(\beta_1, \beta_2, \cdots, \beta_n)$ is a diagonal matrix with $\beta_i = \sigma\{y(x_i; \boldsymbol{w}_{MP})\}[1 - \sigma\{y(x_i; \boldsymbol{w}_{MP})\}]$, $\boldsymbol{\Phi}$ is the design matrix with $\boldsymbol{\Phi}_{ij} = K(x_i, x_j)$, and ∇ denotes the gradient operator.

(3) The hyper-parameters α_i can be updated by the following equation:

$$\alpha_i^{new} = \frac{1 - \alpha_i^{old} \sum_{ii}}{w_{MP_i}^2}, \tag{26}$$

where the covariance $\sum_{ii}$ is the ith diagonal element of the posterior weight covariance $\sum = (\boldsymbol{\Phi}^T \boldsymbol{B} \boldsymbol{\Phi} + \boldsymbol{A})^{-1}$, w_{MP_i} is the ith element of $\boldsymbol{w}_{MP}$, and $\boldsymbol{w}_{MP} = \sum \boldsymbol{\Phi}^T \boldsymbol{B} \boldsymbol{t}$ is the most likely maximum posterior weight vector.

4.2. Multiclass Relevance Vector Machine

The multiclass relevance vector machine (mRVM) expands the original RVM to the multi-class setting by introducing the auxiliary variables $\boldsymbol{Y} \in \Re^{C \times N}$, which act as intermediate regression targets. This approach uses the hierarchical Bayesian framework to solve multiclass issues and introduces the multinomial probit likelihood to output the probability of the members. The structure of the hierarchical Bayesian model is shown in Figure 2.

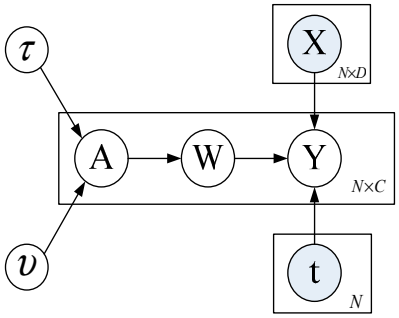

Figure 2. Schematic diagram of the hierarchical Bayesian model.

There are two classification algorithms, mRVM_1 and mRVM_2, which expand the original RVM to the multi-class multi-kernel setting. The theoretical bases of the two versions of mRVM are consistent.

First, a training set $\boldsymbol{T} = \{\boldsymbol{x}_i, t_i\}_{i=1}^{N}$ should be given, where $\boldsymbol{x}_i \in \Re^D$ is the D-dimensional feature vector and $\boldsymbol{t} \in \{1, 2, \cdots, C\}$ is the respective class label. The kernel function of the training set is $\boldsymbol{K} \in \Re^{N \times N}$. By introducing the auxiliary variables $\boldsymbol{Y} \in \Re^{C \times N}$ to act as the regression targets of the weighting parameter $\boldsymbol{W}^T\boldsymbol{K}$, the standardized noise model is deduced as follows:

$$\boldsymbol{y}_{cn} | w_c, \boldsymbol{k}_n \sim N_{y_{cn}}(w_c^T \boldsymbol{k}_n, 1). \tag{27}$$

Convert the regression targets above into the category labels by introducing the multinomial probit link:

$$\boldsymbol{t}_n = i, \boldsymbol{y}_{ni} > \boldsymbol{y}_{nj}, \forall i \neq j. \tag{28}$$

The probabilistic output for class membership is obtained through the resultant multinomial probit likelihood function:

$$P(\boldsymbol{t}_n = i | \boldsymbol{w}, \boldsymbol{k}_n) = \varepsilon_{p(u)} \left\{ \prod_{j \neq i} \boldsymbol{\Phi}(u + (w_i - w_j)^T \boldsymbol{k}_n) \right\}, \tag{29}$$

where $u \sim N(0, 1)$ and $\boldsymbol{\Phi}$ denotes the Gaussian cumulative distribution function.

To ensure the sparse model, the weight w_{nc} follows a standard normal distribution with zero-mean and variance α_{nc}^{-1}, where α_{nc} belongs to the scales matrix $\boldsymbol{A} \in \Re^{N \times C}$ and follows a gamma distribution. With sufficiently small τ and v ($<10^{-5}$), most $\boldsymbol{w}$ values are restricted around zero, which naturally leads to a sparse solution.

The closed-form posterior of the weight $\boldsymbol{w}$ can be derived as follows:

$$P(\boldsymbol{w}|\boldsymbol{Y}) \propto P(\boldsymbol{Y}|\boldsymbol{w}) P(\boldsymbol{w}|\boldsymbol{A}) \propto \prod_{c=1}^{C} N((\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{A}_c)^{-1} \boldsymbol{K} y_c^T, (\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{A}_c))^{-1}, \tag{30}$$

where $\boldsymbol{A}_c$ is a diagonal matrix derived from the c column of $\boldsymbol{A}$. The maximum a posteriori estimator is as follows:

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{arg\,max} P(\boldsymbol{w}|\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{K}). \tag{31}$$

Then, the parameters are updated based on the maximum a posterior estimator value when a class i is given:

$$\hat{w}_c = (\boldsymbol{KK} + \boldsymbol{A}_c)^{-1}\boldsymbol{K}y_c. \tag{32}$$

Finally, the posterior probability distribution of the priori parameters of the weight vector is as follows:

$$P(\boldsymbol{A}|\boldsymbol{w}) \propto P(\boldsymbol{w}|\boldsymbol{A})P(\boldsymbol{A}|\tau, \upsilon) \propto \prod_{c=1}^{C}\prod_{n=1}^{N} G(\tau + \frac{1}{2}, \frac{w_{nc}^2 + 2\upsilon}{2}).$$

Psorakis et al. introduced and provided the theoretical background of the two multi-class multi-kernel mRVMs [25]. The difference between the two versions of the mRVM is how the mRVM manipulates the kernel during the training phase. $mRVM_1$ follows a constructive approach starting with an empty model and then adding or removing samples from the training kernel based on their contribution. However, $mRVM_2$ follows a top-down approach, loading the entire training set into memory and iteratively pruning uninformative samples. The study focused on their multi-class discrimination aspect and provided an extensive evaluation of $mRVM_1$ and $mRVM_2$ following a thorough experimentation on real-world data sets. $mRVM_1$ has a better ability of prototypical sample identification properties and naturally leads to more confident predictions, while $mRVM_2$ is more accurate in terms of predictive power and is better able to detect outliers. $mRVM_1$ is very suitable for large-scale problems by using the fast type-II ML procedure, but it yields lower class recognition rates than $mRVM_2$ because it is the less expressive model. In terms of sparsity, both models use only a small fraction of the overall training set to achieve good class recognition accuracy.

5. Proposed Fault Diagnosis and Severity Analysis Method

5.1. Fault Detection Process

For rolling bearings, when bearing failure occurs, the fault defect triggers an instantaneous impact with non-stationarity. When a rolling bearing undergoes local failure, every contact in the process of operation triggers an instantaneous impact, resulting in a large number of impact waveforms in the measured vibration signal of the rolling bearing. Therefore, the vibration signal of the rolling bearing shows impact characteristics and the impact waveform changes with the severity of the fault. This impact waveform causes the complexity of the vibration signal to increase, leading to abrupt changes in the SampEn value. The fault detection of the rolling bearing vibration signals can be realized by using the SampEn to measure the complexity of the time series. The more serious the fault is, the more complex the vibration signal is. The threshold of fault detection is determined by the SampEn of the vibration signal under normal conditions and different fault conditions of the rolling bearing. If the SampEn value of the vibration signal of the rolling bearing is greater than the set threshold value, it indicates that the rolling bearing has failed. To improve the real-time performance of the fault detection process, the calculation method based on the kd tree is applied to the calculation of the SampEn. Although rolling bearing faults can be detected through SampEn, the identification of the fault types needs to be realized through the fault diagnosis process.

5.2. Fault Diagnosis Process

Once the rolling bearing has failed, the vibration signal of the rolling bearing is decomposed and restructured by the WPT. The WPEE of each node is calculated to generate the fault feature vector of the vibration signal. The mRVM classifier is established by the feature vectors extracted by different fault samples. The fault feature vector is fed into the mRVM to identify the fault type. The flow chart of the proposed fault diagnosis method is shown in Figure 3. The process of the fault diagnosis method is divided into two parts: The training process and the testing process.

In the training process, the task is to determine the threshold of the SampEn for fault detection and train the mRVM multiple classifier. The main steps of the training process are as follows:

(1) Collect sufficient vibration samples of the rolling bearing under different fault conditions and construct the vibration sample set;
(2) For each vibration sample, calculate the SampEn value and determine the fault detection threshold;
(3) In the vibration sample set, calculate the WPEEs of vibration samples under different fault conditions and generate the feature vector set under different fault conditions;
(4) Establish the mRVM classifier by using the feature vector set.

The main steps of the testing process are as follows:

(1) Acquire the vibration signal of the rolling bearing and partition it into non-overlapping segments with the setting length N;
(2) Calculate the SampEn value of the vibration signal by the fast algorithm of the sample entropy according to Section 2.2;
(3) Compare the SampEn value of vibration signal with the threshold value to determine whether the rolling bearing has faults or not. If the SampEn value is smaller than the threshold value, this result suggests that the rolling bearing is running at normal conditions; jump to step (1) to continue collecting the vibration signals of the rolling bearing. In contrast, if the SampEn value is larger than the threshold value, then proceed to step (4);
(4) Decompose the vibration signal by the WPT, calculate the WPEE of the reconstructed signal of each node in the last level, and generate the fault feature vector;
(5) Input the fault feature vector into the trained mRVM to identify the fault type;
(6) Output the fault diagnosis result.

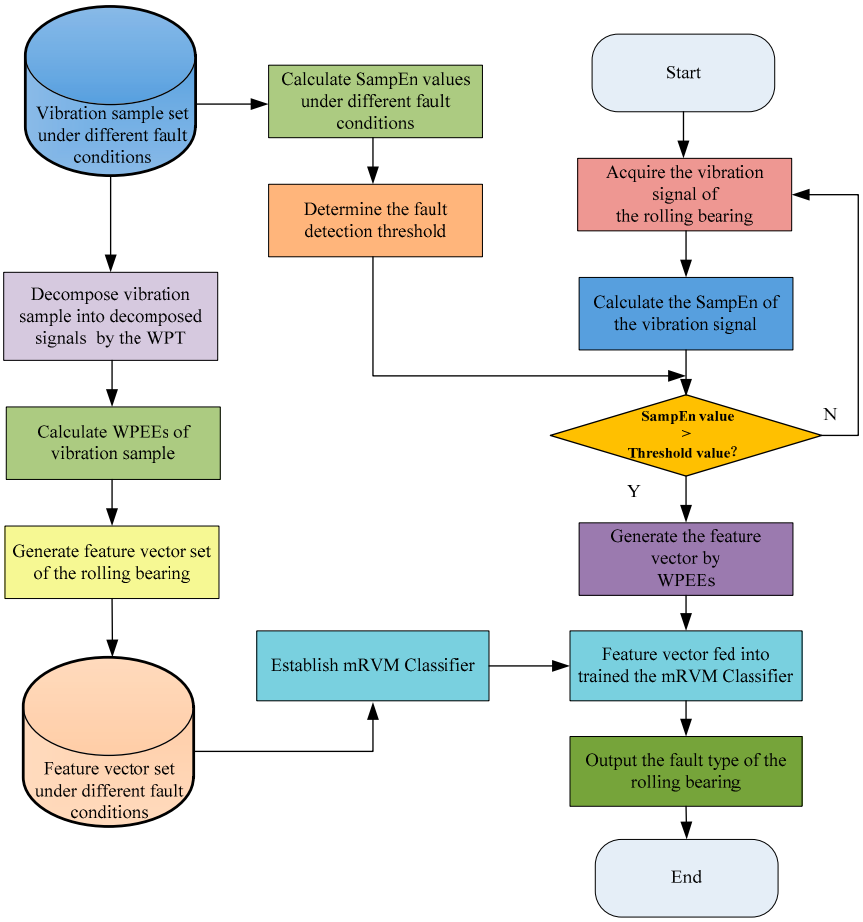

Figure 3. Flow chart of the proposed fault diagnosis method.

5.3. Fault Severity Analysis Process

The fault diameter of the rolling bearing directly affects the running state of the rolling bearing, which is a manifestation of the fault severity. Therefore, the energy contained in the vibration signal of the rolling bearing under different fault severities is different. In this paper, the wavelet packet energy entropy is used to describe the fault severity of the rolling bearing. To further analyze the fault severity on the basis of accurate fault type identification, this paper adopts the mRVM to further identify the fault severity under the same fault type. The fault severity analysis process is described below:

(1) Vibration signals of the rolling bearing with different fault diameters under the same fault are sampled and segmented according to a certain length to form a sample set of vibration signals under different fault severities;
(2) WPEE feature extraction is performed on the sample set composed of vibration signals with different fault severities under different fault types and the set of feature vectors representing the fault severity is composed;
(3) The mRVM model is modeled by using the sample set obtained in step (2), which can be used to analyze the fault severity of the rolling bearing.

As long as there are enough samples of vibration signals of the rolling bearing with different fault diameters, the WPEE can be used to describe the fault severity and the mRVM can be used to identify the fault severity.

6. Experiments and Results

6.1. Experimental Data and Setup

To verify the effectiveness of the proposed rolling bearing fault diagnosis and severity analysis method, the experimental data selected in this paper are all obtained from the Bearing Data Center of Case Western Reserve University and the rolling bearing experiment system and its sketch are shown in Figure 4. The accelerometers were placed at two positions, one at the drive end of the motor and the other at the fan end. Faults were introduced separately at the inner raceway, outer raceway, and rolling element by using electro-discharge machining. The vibration signals of the rolling bearing included four types: Normal, ball fault (BF), inner race fault (IRF), and outer race fault (ORF). The rolling bearing was tested under four different loads of 0, 1, 2, and 3 horse power (hp), each type of fault ranging from 0.007 inches to 0.040 inches in diameter. Motor bearings were seeded with faults using electro-discharge machining (EDM). Faults ranging from 0.007 inches in diameter to 0.040 inches in diameter were introduced separately at the inner raceway, rolling element (i.e., ball), and outer raceway. Faulted bearings were reinstalled into the test motor and vibration data was recorded for motor loads of 0 to 3 horsepower (motor speeds of 1797 to 1720 rpm). The sampling rates of the data were 12 kHz and 24 kHz. The drive-end and fan-end bearing specifications, including the bearing geometry and defect frequencies are listed in Tables 1 and 2, respectively. Figure 5 shows the vibration signals collected under different rolling bearing conditions for a load of 0 hp. Further details can be found on the Case Western Reserve University website [35].

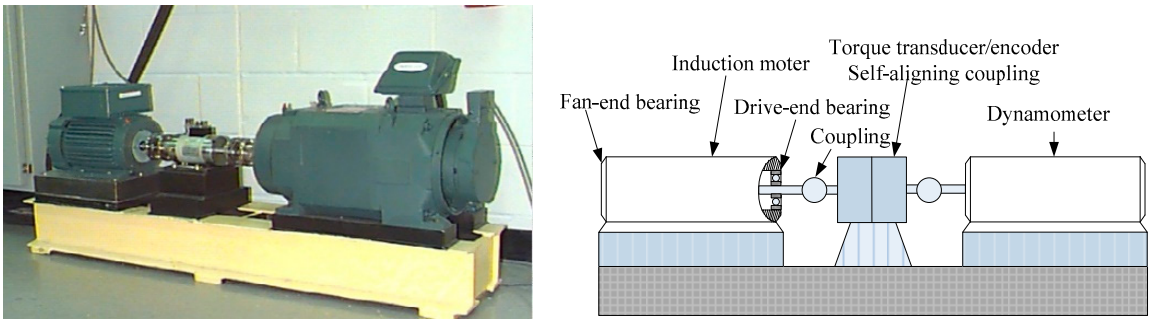

Figure 4. The rolling bearing experiment system and its sketch of Bearing Data Center of Case Western Reserve University.

Table 1. Drive-end bearing defect frequencies (multiples of the running speed in Hz).

Inner Race	Outer Race	Cage Train	Rolling Element
5.4152	3.5848	0.39828	4.7135

Table 2. Fan-end bearing defect frequencies (multiples of the running speed in Hz).

Inner Race	Outer Race	Cage Train	Rolling Element
4.9469	3.0530	0.3817	3.9874

In this study, the sampling rate was set to 12 kHz and the drive-end rolling bearing data under four different loads of 0, 1, 2, and 3 hp with fault diameters of 0.007 in, 0.014 in, 0.021 in, and 0.028 in were collected. The working conditions of the collected rolling bearing data are illustrated in Table 3, where "✓" indicates the data sets that are selected and "*" indicates the data sets that are not available.

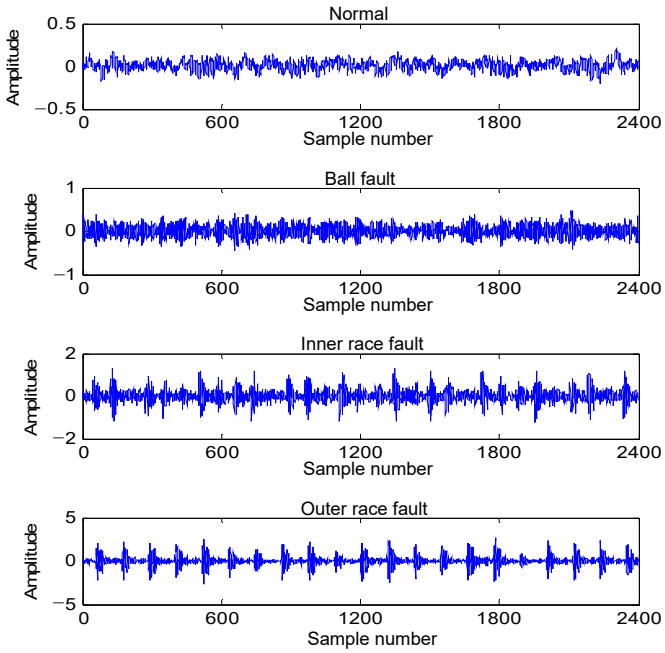

Figure 5. Vibration signals collected under different rolling bearing conditions for a load of 0 hp.

Table 3. Working conditions of the collected rolling bearing data.

Fault Type	Fault Diameter (in)	Load (hp)			
		0	1	2	3
Normal	—	✓	✓	✓	✓
BF	0.007	✓	✓	✓	✓
	0.014	✓	✓	✓	✓
	0.021	✓	✓	✓	✓
	0.028	✓	✓	✓	✓
IRF	0.007	✓	✓	✓	✓
	0.014	✓	✓	✓	✓
	0.021	✓	✓	✓	✓
	0.028	✓	✓	✓	✓
ORF	0.007	✓	✓	✓	✓
	0.014	✓	✓	✓	✓
	0.021	✓	✓	✓	✓
	0.028	*	*	*	*

6.2. Fault Detection Experiment

To demonstrate the effectiveness of the fast SampEn algorithm for rolling bearing fault detection, fault detection experiments are conducted in this section. When the SampEn is used to detect dynamical changes in vibration time series, the vibration signal should be partitioned into several non-overlapping segments with a suitable length. As discussed in Section 2, the length of the time series $\boldsymbol{N}$ has enormous implications for the calculation time. Thus, setting an appropriate length $\boldsymbol{N}$ is important for the SampEn calculational efficiency. Obviously, the length $\boldsymbol{N}$ should not be too large because the calculation time would be too long. However, $\boldsymbol{N}$ should not be too small because the SampEn value would not be effective for the fault detection of rolling bearings. Zhang et al. set $N = 2400$ and successfully detected and diagnosed a rolling bearing fault [36]. Zheng et al. [37] and Wang et al. [38] respectively set $N = 4096$ and $N = 2000$ with a sampling frequency of 20 kHz. Li et al. set $N = 2400$ with a sampling frequency of 12 kHz [15]. In comparison, this study sets $N = 2400$, which not only conserves the calculation time but also produces reasonable statistical validity.

The embedding dimension can be set as $m = 1$ or $m = 2$. To detect the dynamics of the change in the vibration signal and realize fault detection, set $m = 2$. The value of the similarity tolerance r affects the sensitivity degree of the SampEn to the time series complexity. The smaller the value of r is, the greater the impact of noise in the calculation results. However, the larger the value of r is, the more detailed the information loss in the time series. Therefore, the similarity tolerance is set to 0.1, which can detect the complexity changes more effectively than other settings.

According to the discussion above, this study set $N = 2400$; the first 120,000 points of each original collected signal were divided into 50 segments and all of the vibration signals under different conditions are partitioned into 2400 non-overlapping segments. The SampEn values for all 2400 samples are calculated and are shown in Figure 6.

In Figure 6, the SampEn values of samples with a fault are larger than those of the normal samples. In other words, samples with a fault have more complexity than normal samples. When the rolling bearing is working in normal conditions, the vibration mainly stems from the coupling between mechanical parts and the environmental noise, which have a certain regularity, so the SampEn values are relatively small. In contrast, when faults occur, the vibration signal of the bearing becomes more complex due to the introduction of some uncertain components. As a result, the SampEn values significantly change.

The appropriate threshold value is set to 0.3, which can distinctly divide the samples with faults from normal samples irrespective of the fault diameter and the load, as shown in Figure 6. As a result, the SampEn value of the vibration signal can be an effective method to detect bearing faults. However, the SampEn value alone cannot be used to distinguish between three types of faults from Figure 6. It also cannot be used to illustrate the fault diameter. Therefore, it is necessary to further analyze the vibration signals, extract more features, and then use an effective classifier to identify the fault type and the fault diameter.

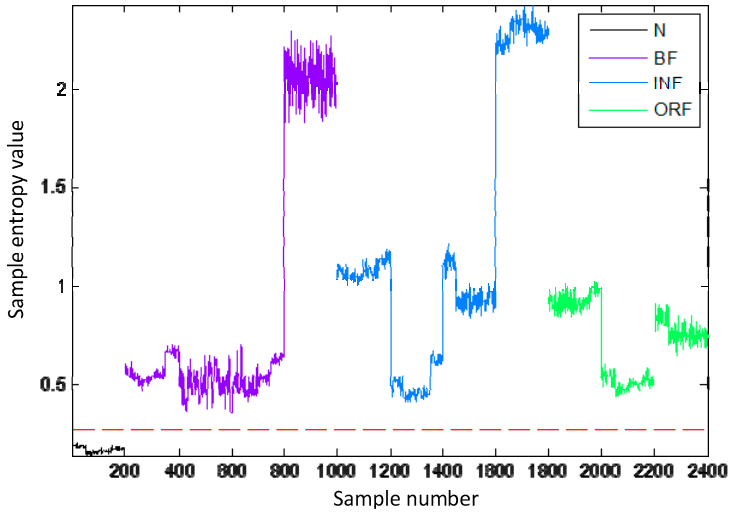

Figure 6. The SampEn values for all 2400 samples under different working conditions.

Several complexity analysis algorithms of time-domain data have also been selected to handle the same sample sets in this paper. In this experiment, different complexity analysis algorithms are calculated for all 2400 segments of the vibration signal sample set and their average running times are shown in Table 4. As shown in Table 4, the permutation entropy, fuzzy entropy, and SampEn are able to detect the fault in the rolling bearing signal and the fast SampEn has an enormous advantage in terms of algorithm speed. Therefore, the fast SampEn has the potential to be a real-time method for the fault detection of rolling bearings.

Table 4. The average running time of several complexity analysis algorithms applied in rolling bearing fault detection.

Complexity Analysis Algorithm	Can Detect Faults?	Average Running Time (s)
Approximate entropy	N	NA
Permutation entropy	Y	2.573
Fuzzy entropy	Y	41.851
Sample entropy	Y	0.744
Fast sample entropy	Y	0.273

6.3. Fault Feature Extraction Experiment

If the rolling bearing has faults, the fault feature should be extracted for further fault identification. In this paper, the WPEE is used to describe fault characteristics. To illustrate the effectiveness of the proposed feature extraction method, the relevant experiments are conducted in this section.

Figure 7 shows the vibration signal decomposed by the WPT with a load of 0 hp and a fault diameter of 0.007 in under a ball fault. The experimental vibration signal is from Bearing Data Center of Case Western Reserve University and the sampling frequency is 12 kHz. The selection of a wave basis not only has a great influence on the effectiveness of signal decomposition and fault feature extraction, but also further affects the accuracy of the rolling bearing fault diagnosis results. Generally, an orthogonal basis should be selected in the wavelet packet transformation and the common orthogonal basis functions are those of the Harr, Daubechies, Coiflets, Symlets, and Meyer families. Daubechies (abbreviated to *dbN*, where *N* denotes the number of wavelet order) wavelets can be applied to different signals by changing the order of the wavelet. In this study, *dbN* is selected as the wavelet basis for wavelet packet analysis and the vibration signal under rolling bearing failure state is decomposed and analyzed by setting the appropriate order *N*. As shown in Figure 7, the WPT can effectively decompose the vibration signal of a rolling bearing under a ball fault into a series

of decomposed signals over different frequency ranges and the spectrum of each node is distinctly different. The fault characteristics of the vibration signal are contained in the decomposed signals and the WPEEs of the vibration signal are used as the feature vector.

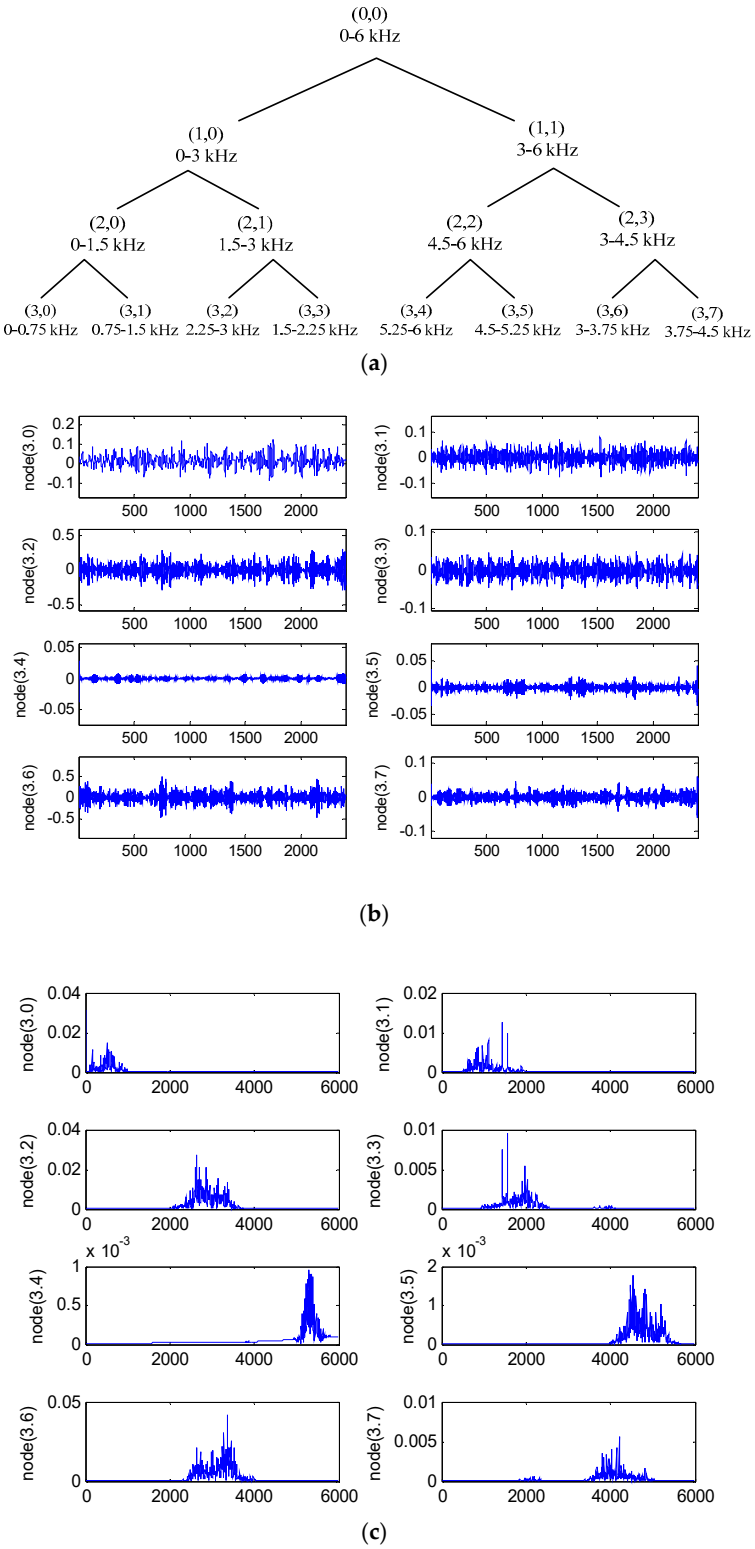

Figure 7. Vibration signal decomposed by WPT with a load of 0 hp and a fault diameter of 0.007 in under a ball fault. (**a**) Three-level wavelet decomposition tree; (**b**) decomposed signals corresponding to each node; (**c**) spectrum diagram of the decomposed vibration signal corresponding to each node.

The dimension of the feature vector is important for fault identification. If the number of wavelet packet levels is 5, the dimension of the feature vector is 32, which is too large to use machine-learning algorithms for fault identification. Moreover, if the number of wavelet packet levels is 3, the dimension of the feature vector is 8, which cannot highlight some features of the rolling bearing. To guarantee the analysis precision and speed, a 4-level WPT is chosen; the number of nodes after decomposition is 16 and the original vibration signals are decomposed into 16 sub-signals. After the signal is decomposed by the wavelet packet, the energy entropy of the reconstructed signal of each node is calculated and the feature vector is formed according to the node order. Then, each sample is represented by a 16-dimensional vector. The type of wavelet basis also has a certain influence on the analysis of the vibration signal; "db10" is chosen as the wavelet basis in this paper. The WPEE feature extraction results under different fault types of rolling bearings for a load of 2 hp are shown in Figure 7. Figure 7a presents the WPEE of a ball fault (BF) with different fault diameters, Figure 7b shows the WPEE of an inner race fault (IRF) with different fault diameters and Figure 7c shows the WPEE of an outer race fault (ORF) with different fault diameters. As shown in Figure 7a–c, choosing "db10" as the wavelet basis can clearly distinguish three types of faults and fault diameters. According to the above, "db10" is chosen as the wavelet basis and the WPT level is set to 4. Then, the energy entropy of the reconstructed signal of each node is calculated in the last level and forms the 16-dimensional feature vector for further fault diagnosis. It can be seen from this figure that the fault characteristics extracted from different fault types are also greatly different, which indicates that under different fault conditions, the energy concentration frequency band is different and different fault diameters also affect the energy entropy value of each frequency band. Therefore, the application of the WPEE in bearing fault feature extraction can reveal the energy distribution in various frequency bands in the form of entropy and effectively characterize the rolling bearing fault type.

As shown in Figure 8, there is a large difference in the energy entropy between node 2, node 6, and node 14 under the different fault types. The wavelet packet energy entropy clustering diagram under different fault conditions for a load of 2 hp is shown in Figure 9. Figure 9 shows the energy entropy clustering of the second node, the sixth node, and the fourteenth node in the fourth level, in which "BF1", "BF2", "BF3", and "BF4", respectively, represent ball fault samples with fault diameters of 0.007, 0.014, 0.021, and 0.027 in and the inner race fault and outer race fault are in turn. From this figure, the features of different fault conditions show good clustering, but since only 3 of the 16 dimensions in the feature vector are taken as the display, there is some overlap between some fault conditions.

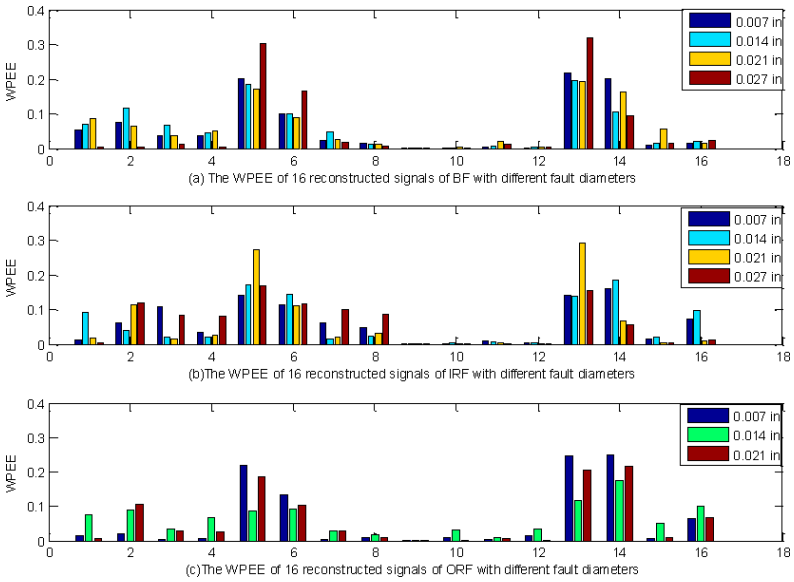

Figure 8. The WPEE feature extraction results under different fault types of rolling bearing with a load of 2 hp.

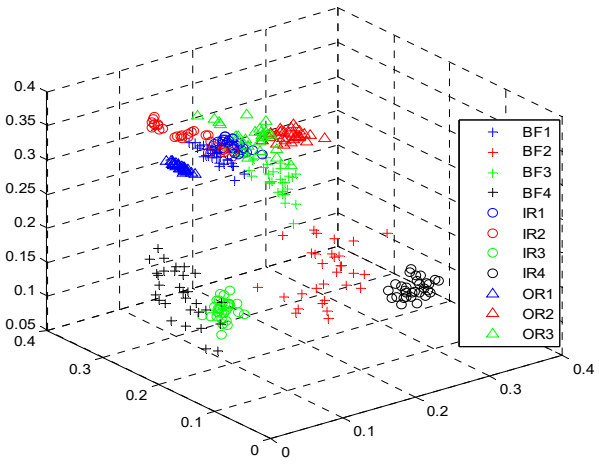

Figure 9. Wavelet packet energy entropy clustering diagram under different fault conditions for a load of 2 hp.

6.4. Fault Identification and Severity Analysis Experiment

To illustrate the fault identification and severity analysis effect of the proposed fault diagnosis method for rolling bearings, various types and fault diameters introduced in Section 6.1 are taken into account in the experiments. For each vibration signal in the vibration sample set, the length of signal segments is set to 2400 and the first 120,000 points in each type of vibration signal are divided into 50 non-overlapping sub-signals. Twenty sub-signals of each vibration signal in each fault type are randomly selected as training samples and the remaining 30 samples are all taken as test samples. Therefore, a total of 880 training samples and 1320 testing samples are used for experiments. The feature vectors of the above samples are extracted by the WPEE, thus composing the feature vector sets under different fault conditions. The mRVM classifier is established for performing the fault identification of the rolling bearings.

The Gaussian kernel function is selected; σ is the kernel parameter for $mRVM_1$ and $mRVM_2$. Each experiment is repeated 10 times and the fault identification accuracy irrespective of the load and fault diameter is shown in Table 5. It can be seen that both versions of the mRVM are able to classify the fault types of the rolling bearings with high recognition accuracy. $mRVM_2$ obtains satisfactory results in this experiment, with the classification accuracy reaching as high as 99.47%; however, the classification accuracy of $mRVM_1$ is slightly worse than that of $mRVM_2$. These results show that $mRVM_2$ offers better performance than $mRVM_1$ in an unknown sample classification.

Table 5. Fault identification accuracy irrespective of the load and the fault diameter.

Number of Categories	Number of Training Samples	Number of Testing Samples	$mRVM_1$		$mRVM_2$	
			σ	Average Accuracy (%)	σ	Average Accuracy (%)
3	880	2200	0.23	98.96	0.4	99.47

Table 6 shows the results of fault diameter identification with the same fault type irrespective of the load. The fault diameter identification of each fault type achieves excellent results and the fault severity identification accuracy exceeds 99%. Table 7 shows the results of the fault type and diameter identification irrespective of the load. The number of categories is 11 and the average classification accuracy is 98.75%. Table 8 shows the results of the fault type and diameter identification under the same load and the proposed fault diagnosis method still has a high identification accuracy. Tables 6–8 illustrates that the proposed fault severity analysis method based on the WPEE and mRVM exhibits good performance in the fault diagnosis of rolling bearings.

Table 6. Fault diameter identification accuracy with the same fault type irrespective of the load.

Fault Type	Number of Categories	Number of Training Samples	Number of Testing Samples	$mRVM_2$	
				σ	Average Accuracy (%)
BF	4	320	480	0.4	99.06
IRF	4	320	480	0.4	99.59
ORF	3	240	360	0.4	100

Table 7. Fault type and diameter identification irrespective of the load.

Number of Categories	Number of Training Samples	Number of Testing Samples	$mRVM_2$	
			σ	Average Accuracy (%)
11	880	2200	0.4	98.75

Table 8. Fault type and diameter identification under the same load.

Load (hp)	Number of Categories	Number of Training Samples	Number of Testing Samples	$mRVM_2$	
				σ	Average Accuracy (%)
0	3	550	2200	0.4	99.29
	11	550	2200	0.4	99.09
1	3	550	2200	0.4	99.80
	11	550	2200	0.4	99.39
2	3	550	2200	0.4	100
	11	550	2200	0.4	99.89
3	3	550	2200	0.4	99.70
	11	550	2200	0.4	99.55

To further experimentally demonstrate the effectiveness and superiority of the proposed method, some comparisons are conducted between the present work and recently published works regarding bearing fault diagnosis using different methods. Table 9 reports the comparative result between the current study and the recently published literature. In Table 9, fault types and fault severities are combined to be diagnosed and a comparison is made on the basis of the feature extraction methods and fault classifier. The proposed fault diagnosis method has a better diagnostic effect than that of the recently published works.

Table 9. Comparative study between the current work and previous works published in the literature.

Literature	Feature Extraction Method	Classifier	Average Accuracy (%)
[11]	EEMD + permutation entropy	Optimized SVM	97.91
[36]	Multi-scale entropy (MSE)	Adaptive neuro-fuzzy inference	99.38
[37]	EMD + statistical features selection	Maximum margin classification	97.10
Present work	WPEE	mRVM	99.47

The results of the experiments and the comparative studies prove the superiority and capacity of the proposed method for rolling bearing fault diagnosis and fault severity analysis; the proposed method has good development and application prospects in the field of rotating machinery fault diagnosis.

7. Conclusions

In this paper, a novel fault diagnosis and severity analysis for rolling bearings based on the fast sample entropy, the wavelet packet energy entropy, and the multiclass relevance vector machine is presented.

(1) The time complexity of the vibration signal increases when a fault occurs in the rolling bearing. To improve the real-time performance of the fault detection process, a fast sample entropy algorithm based on a kd tree is adopted. The rolling bearing failure can be effectively identified by calculating the fast sample entropy of the vibration signal and this fault detection method can effectively improve the real-time performance of fault detection process.

(2) Considering that the vibration signals are complex and exhibit non-stationarity and non-linearity, the wavelet packet energy entropy is adopted as the fault feature extraction method. First, the wavelet packet transform decomposes the fault signal into different frequency bands. Then, the wavelet packet energy entropy values of the reconstructed signals of each node in the last level of the wavelet packet are calculated and the fault feature vector is generated. The wavelet packet energy entropy can effectively extract the fault information of a rolling bearing vibration signal with good feature description performance and the extracted feature vector has strong separability.

(3) The fault feature vectors are fed into a multiclass relevance vector machine classifier to identify the fault type as well as the fault severity. The results of the experiments in this paper, which take into account multiple combinations of fault types and fault diameters, show that the proposed method can diagnose the fault type as well as the fault severity with high accuracy.

This paper provides a novel fault diagnosis and fault severity analysis method and an effective exploration method for fault severity analysis on the basis of accurate identification of fault types. In this paper, the classifier method is used to analyze the fault severity of rolling bearings; the future research will focus on the analysis of the fault severity by means of a prediction method and the influence of fault location on bearing fault diagnosis [39].

Author Contributions: Y.C. designed the method and wrote this paper. T.Z. and K.S. performed the experiments. Z.L. analyzed the data and proofread the manuscript.

Funding: This work received financial support from the National Natural Science Foundation of China (No. 61803128).

Acknowledgments: Thanks for the Bearing Data Center of Case Western Reserve University to supply the rolling bearing data set. Thanks to Mengyue Liu of Harbin Institute of Technology for the preliminary work of this paper.

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

Nomenclature

AM	Amplitude-modulated
ApEn	Approximate entropy
CEEMD	Complete ensemble empirical mode decomposition
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
FFT	Fast Fourier transform
FM	Frequency-modulated
FE	Fuzzy entropy
IMFs	Intrinsic mode functions
LMD	Local mean decomposition
HMM	Markov modeling
mRVM	Multiclass relevance vector machine
PE	Permutation entropy
PFs	Product functions
PCA	Principle component analysis

RVM	Relevance vector machine
SampEn	Sample entropy
STFT	Short-time Fourier transform
SVM	Support vector machine
WT	Wavelet transform
WPT	Wavelet packet transform
WPEE	Wavelet packet energy entropy

References

1. Li, Y.; Xu, M.; Wei, Y.; Huang, W. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* **2016**, *77*, 80–94. [CrossRef]
2. Li, K.; Su, L.; Wu, J.; Wang, H.; Chen, P. A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine. *Appl. Sci.* **2017**, *7*, 1004. [CrossRef]
3. Ocak, H.; Loparo, K.A.; Discenzo, F.M. Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics. *J. Sound Vib.* **2007**, *302*, 951–961. [CrossRef]
4. Vafaei, S.; Rahnejat, H. Indicated repeatable runout with wavelet decomposition (IRR-WD) for effective determination of bearing-induced vibration. *J. Sound Vib.* **2003**, *260*, 67–82. [CrossRef]
5. Guo, W.; Tse, P.W. A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. *J. Sound Vib.* **2013**, *332*, 423–441. [CrossRef]
6. Li, C.; Liang, M.; Zhang, Y.; Hou, S. Multi-scale autocorrelation via morphological wavelet slices for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **2012**, *31*, 428–446. [CrossRef]
7. Ding, S.X. *Model-Based Fault Diagnosis Techniques*; Springer: Berlin/Heidelberg, Germany, 2008.
8. Chen, J.; Patton, R.J. *Robust Model-Based Fault Diagnosis for Dynamic Systems*; Springer: New York, NY, USA, 1999.
9. Feng, Z.; Liang, M.; Chu, F. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech. Syst. Signal Process.* **2013**, *38*, 165–205. [CrossRef]
10. Zheng, J.; Cheng, J.; Yang, Y.; Luo, S. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mech. Mach. Theory* **2014**, *78*, 187–200. [CrossRef]
11. Zhang, X.; Liang, Y.; Zhou, J.; Zang, Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* **2015**, *69*, 164–179. [CrossRef]
12. Han, M.; Pan, J. A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings. *Measurement* **2015**, *76*, 7–19. [CrossRef]
13. Wang, Y.; Xu, G.; Liang, L.; Jiang, K. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **2015**, *54–55*, 259–276. [CrossRef]
14. Yao, B.; Su, J.; Wu, L.; Guan, Y. Modified local linear embedding algorithm for rolling element bearing fault diagnosis. *Appl. Sci.* **2017**, *7*, 1178. [CrossRef]
15. Li, Y.; Xu, M.; Zhao, H.; Huang, W. Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis. *Mech. Mach. Theory* **2016**, *98*, 114–132. [CrossRef]
16. Xiong, Z.; Ramchandran, K.; Herley, C.; Orchard, M.T. Flexible Tree-structured Signal Expansions Using Time-varying Wavelet Packets. *IEEE Trans. Signal Process.* **1997**, *45*, 333–345. [CrossRef]
17. Chen, Y.; Xu, Y.; Yang, J.; Shi, Z.; Jiang, S.; Wang, Q. Fault detection, isolation, and diagnosis of status self-validating gas sensor arrays. *Rev. Sci. Instrum.* **2016**, *87*, 045001. [CrossRef]
18. Shi, Z.; Song, W.; Taheri, S. Improved LMD, permutation entropy and optimized K-means to fault diagnosis for roller bearings. *Entropy* **2016**, *18*, 70. [CrossRef]
19. Yan, R.; Gao, R.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]
20. Tiwari, R.; Gupta, V.K.; Kankar, P.K. Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier. *J. Vib. Control* **2015**, *21*, 461–467. [CrossRef]

21. Arnaiz-González, Á.; Fernández-Valdivielso, A.; Bustillo, A.; de Lacalle, L.N.L. Using artificial neural networks for the prediction of dimensional error on inclined surfaces manufactured by ball-end milling. *Int. J. Adv. Manuf. Technol.* **2016**, *83*, 847–859. [CrossRef]
22. Liu, X.; Tang, J. Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. *IEEE Syst. J.* **2014**, *8*, 910–920. [CrossRef]
23. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
24. Ghosh, S.; Mujumdar, P.P. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Adv. Water Resour.* **2008**, *31*, 132–146. [CrossRef]
25. Widodo, A.; Yang, B.S. Application of relevance vector machine and survival probability to machine degradation assessment. *Expert Syst. Appl.* **2011**, *38*, 2592–2599. [CrossRef]
26. Psorakis, I.; Damoulas, T.; Girolami, M.A. Multiclass relevance vector machines: Sparsity and accuracy. *IEEE Trans. Neural Netw.* **2010**, *21*, 1588–1598. [CrossRef] [PubMed]
27. Lei, Y.; Liu, Z.; Wu, X.; Li, N.; Chen, W.; Lin, J. Health condition identification of multi-stage planetary gearboxes using a mRVM-based method. *Mech. Syst. Signal Process.* **2015**, *60*, 289–300. [CrossRef]
28. Lei, Z. A multivariate relevance vector machine based algorithm for on-line fault prognostic application with multiple fault features. In Proceedings of the 2012 Fifth International Conference on Intelligent Computation Technology and Automation (ICICTA), Zhangjiajie, China, 12–14 January 2012; pp. 26–32.
29. Xu, H.; Tang, T.; Wang, T.; Benbouzid, M.E.H. A PCA-mRVM fault diagnosis strategy and its Application in CHMLIS. In Proceedings of the IECON 2014—40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, TX, USA, 29 October–1 November 2014; pp. 1124–1130.
30. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart C* **2000**, *278*, H2039–H2049. [CrossRef] [PubMed]
31. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [CrossRef] [PubMed]
32. Pan, Y.H.; Lin, W.Y.; Wang, Y.H.; Lee, K.T. Computing multiscale entropy with orthogonal range search. *J. Mar. Sci. Technol.* **2011**, *19*, 107–113.
33. Dan, S.; Plumbley, M.D. Fast Multidimensional Entropy Estimation by k-d Partitioning. *IEEE Signal Process. Lett.* **2009**, *16*, 537–540.
34. Pan, Y.; Wang, Y.; Liang, S.; Lee, K.T. Fast computation of sample entropy and approximate entropy in biomedicine. *Comput. Methods Programs Biomed.* **2011**, *104*, 382–396. [CrossRef]
35. Bearing Data Center, Case Western Reserve University. Available online: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file (accessed on 31 January 2010).
36. 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. [CrossRef]
37. Zeng, M.; Yang, Y.; Zheng, J.; Cheng, J. Maximum margin classification based on flexible convex hulls for fault diagnosis of roller bearings. *Mech. Syst. Signal Process.* **2016**, *66*, 533–545. [CrossRef]
38. Wang, F.; Zhang, Y.; Zhang, B.; Su, W. Application of wavelet packet sample entropy in the forecast of rolling element bearing fault trend. In Proceedings of the International Conference on Multimedia & Signal Processing, Guilin, China, 14–15 May 2011.
39. De Lacalle, L.N.L.; Lamikiz, A.; Sanchez, J.A.; de Bustos, I.F. Simultaneous measurement of forces and machine tool position for diagnostic of machining tests. *IEEE Trans. Instrum. Meas.* **2005**, *54*, 2329–2335.