Article

Weighted Kernel Entropy Component Analysis for Fault Diagnosis of Rolling Bearings

Hongdi Zhou [1], Tielin Shi [1], Guanglan Liao [1],*, Jianping Xuan [1], Jie Duan [1], Lei Su [2], Zhenzhi He [3] and Wuxing Lai [1],*

[1]   State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China; zh_hongdi@hust.edu.cn (H.Z.); tlshi@hust.edu.cn (T.S.); jpxuan@hust.edu.cn (J.X.); jduan@hust.edu.cn (J.D.)
[2]   School of Mechanical Engineering, Jiangnan University, Wuxi 214122, China; lei_su2015@jiangnan.edu.cn
[3]   School of Mechanical & Electrical Engineering, Jiangsu Normal University, Xuzhou 221116, China; hezz82@163.com
*   Correspondence: guanglan.liao@hust.edu.cn (G.L.); wxlai@hust.edu.cn (W.L.); Tel.: +86-27-8779-3103 (G.L.); +86-27-8755-6835 (W.L.)

Abstract: This paper presents a supervised feature extraction method called weighted kernel entropy component analysis (WKECA) for fault diagnosis of rolling bearings. The method is developed based on kernel entropy component analysis (KECA) which attempts to preserve the Renyi entropy of the data set after dimension reduction. It makes full use of the labeled information and introduces a weight strategy in the feature extraction. The class-related weights are introduced to denote differences among the samples from different patterns, and genetic algorithm (GA) is implemented to seek out appropriate weights for optimizing the classification results. The features based on wavelet packet decomposition are derived from the original signals. Then the intrinsic geometric features extracted by WKECA are fed into the support vector machine (SVM) classifier to recognize different operating conditions of bearings, and we obtain the overall accuracy (97%) for the experimental samples. The experimental results demonstrated the feasibility and effectiveness of the proposed method.

Keywords: fault diagnosis; weighted kernel entropy component analysis; dimensional reduction; Renyi entropy; feature extraction

1. Introduction

Rolling element bearings are widely used in rotating machines in modern industry, and bearing failure is one of the most common reasons for machine breakdown. Unexpected failures may cause huge economic losses and even lead to casualties [1–3]. Therefore, it is important to accurately diagnose bearing faults at the early stage [4,5]. Vibration-based fault diagnosis has been extensively studied to improve existing techniques toward the goal of more accurately dealing with various problems, such as varying load effect and noise contamination [3–8]. Especially, the sensitivity of diagnostic features from the vibration signals may vary with different load conditions due to nonlinear effect and non-stationary noise, of which no single-domain processing methods can comprehensively extract the fault features to reflect the condition [9]. High-dimensional feature sets constructed with mix-domain features are often used for diagnosis [10,11]. Although more features can obviously provide more information, they contain a lot of redundant and disturbed information which will increase computation time and reduce recognition accuracy. More effective feature extraction and dimensionality reduction methods are needed to obtain higher diagnostic accuracy [12,13].

Principal component analysis (PCA) is one typical method for dimensionality reduction and has been widely used for fault diagnosis [14–17], since it can extract representative features from high-dimensionality, noisy and linear correlated data. PCA is an unsupervised method that projects the original dataset onto a lower-dimensional space meanwhile minimizes the mean square error [15]. It can guarantee that the linear features can be extracted while some useful nonlinear features may be lost, as the most of industrial systems are non-linearity and non-stationary. Therefore, nonlinear methods are required to handle the nonlinear data, among which kernel principal component analysis (KPCA) [18] is the most prominent one. KPCA is an extension of traditional linear PCA by using kernel trick, implicitly mapping the original features into a high-dimensional feature space in which the mapped data are linearly separable and then the linear PCA can be conducted [15]. Both PCA and KPCA are typical spectral dimensionality reduction methods which extract features by selecting the top eigenvalues and corresponding eigenvectors of the specially constructed feature matrixes [19]. Hence, the extraction may select uninformative eigenvectors from the information theory standpoint [20].

Kernel entropy component analysis (KECA) is a newly developed information-theory-based dimensionality reduction method, first proposed and employed in pattern recognition by Robert Jenssen [21]. This method attempts to maintain the maximum estimated Renyi quadratic entropy of the input data set via a kernel-based estimator. It is fundamentally different from other methods in two ways: on the one hand, the selection of top eigenvalues and corresponding eigenvectors is not necessary; on the other hand, the dimension reduction reveals the intrinsic structure related to the Renyi entropy of the input data [21–25]. Moreover, KECA typically generates a transformed dataset with a distinct angular structure, implying that even nonlinearly related input data sets are distributed in different angular directions with respect to the high-dimensional kernel feature space [21–25]. KECA has been applied to feature extraction and pattern recognition successfully, showing superior performance over PCA and KPCA [21–24]. However, KECA is unsupervised, ignoring the label information of the input data, which may discard discriminant classification information and weaken recognition accuracy [25]. And the projections in PCA, KPCA and KECA are theoretically optimal for reconstruction from a low-dimensional basis, while they may not be optimal from the viewpoint of discrimination. Many previous studies attempt to extract discriminative features to express the original clusters [25,26], and meanwhile to find a trade-off between maximizing the testing accuracy and minimizing the training error [20,26,27].

In this study, we propose a supervised feature extraction method called weight kernel entropy component analysis (WKECA) based on KECA, in which a modified Fisher criterion is applied to represent class separability. The class-related weights are introduced to denote differences among the samples from different patterns, and genetic algorithm (GA) is applied to seek out appropriate weights for optimizing the classification results. Experimental investigation is conducted to demonstrate the feasibility and effectiveness of the proposed method for fault diagnosis.

2. The Theoretical Background of WKECA for Fault Diagnosis

2.1. Brief Review of KECA

Assuming that $p(\mathbf{x})$ is the probability density function of a given sample $\mathbf{X} = \mathbf{x_1}, \ldots, \mathbf{x_N}$, its Renyi entropy of the order α is expressed as $H_\alpha(\mathbf{X}) = \frac{1}{1-\alpha}\lg(\int p^\alpha(\mathbf{x})d\mathbf{x})$ [28], where $\alpha \geq 1$. In KECA, Renyi quadratic entropy ($\alpha = 2$) is employed, because the entropy value can be elegantly estimated by Parzen window density estimator [29]. Renyi quadratic entropy can be expressed by $H(\mathbf{X}) = -\lg(\int p^2(\mathbf{x})d\mathbf{x})$. Since the monotonicity property of logarithmic function, only the integral function $V(p) = \int p^2(\mathbf{x})dx = E\{p(\mathbf{x})\}$ needs to be considered [21,22,30]. To estimate $V(p)$, a Parzen window density estimator $\overline{p}(\mathbf{x}) = \frac{1}{N}\sum_{x_i \subset D} \mathbf{K}_\sigma(\mathbf{x}, \mathbf{x_i})$ is applied [21,29], where $\mathbf{K}_\sigma(\mathbf{x}, \mathbf{x_i})$ is the estimator or kernel function centered at $\mathbf{x_i}$ and σ is the smoothing width or the kernel size. According to the

convolution theorem, the convolution of two Gaussian functions is another Gaussian function with $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Substituted $K_\sigma(\mathbf{x}, \mathbf{x_i})$ and $\overline{p}(\mathbf{x})$ into $V(p)$, the following estimation can be obtained:

$$
\begin{aligned}
\overline{V}(p) \quad &= \int p^2(\mathbf{x}) d\mathbf{x} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int K_\sigma(\mathbf{x}, \mathbf{x_i}) K_\sigma(\mathbf{x}, \mathbf{x_j}) d\mathbf{x} \\
&= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{\sqrt{2}\sigma}(\mathbf{x_i}, \mathbf{x_j}) = \frac{1}{N^2} \mathbf{1K1^T}
\end{aligned}
\tag{1}
$$

where $\mathbf{K}$ is a N × N kernel matrix, the element (i, j) of $\mathbf{K}$ is $K_\sigma(\mathbf{x_i}, \mathbf{x_j})$, and $\mathbf{1}$ is a N × 1 vector (all elements are one). Therefore, the Renyi entropy can be estimated by the corresponding kernel matrix that can be decomposed as $\mathbf{K} = \mathbf{EDE^T}$, where $\mathbf{D} = diag(\lambda_1, \lambda_2,..., \lambda_N)$ and $\mathbf{E} = [\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2},...,\boldsymbol{\alpha_N}]$. Here λ_i and $\boldsymbol{\alpha_i}$ are the eigenvalues and corresponding eigenvectors, respectively. Then:

$$
\overline{\mathbf{V}}(p) = \frac{1}{N^2} \mathbf{1K1^T} = \frac{1}{N^2} \mathbf{1EDE^T1^T} = \frac{1}{N^2} \sum_{i=1}^{N} \left(\sqrt{\lambda_i} \boldsymbol{\alpha_i}^\mathbf{T} \mathbf{1} \right)^2
\tag{2}
$$

This expression is the so-called entropy values, and each term $\sqrt{\lambda_i} \boldsymbol{\alpha_i}^\mathbf{T}$ contributes to the entropy estimation. The eigenvectors and corresponding eigenvalues are ranked in decreasing order of the entropies. KECA selects certain eigenvalues and corresponding eigenvectors according to the *d* largest entropies [21], different from PCA and KPCA that select largest eigenvalues. Therefore, the resulting KECA expression is $\boldsymbol{\Phi}_{keca} = \mathbf{D_d}^{\frac{1}{2}} \mathbf{E_d}^\mathbf{T}$, where $\mathbf{D_d}$ and $\mathbf{E_d}$ store the top d eigenvalues and corresponding eigenvectors.

2.2. Introduction of WKECA

Given a set of *c*-class training sample patterns $\mathbf{x_i} \in R^N$ (i = 1, 2, ..., N), and each sample $\mathbf{x_i}$ belongs to one of *c*-class. Defined that the weight vector is $[u_1, u_2, ..., u_N]$ and the label values are $\{l_1, l_2, ..., l_c\}$. Each sample has the corresponding label value based on its own class properties. Thus, $u_i = l_j$ if $\mathbf{x_i} \in$ j-th class, where i = 1, 2, ..., N and j = 1,2, ..., *c*. Here the weights are depended on the class so that they can represent the class information. The weighted matrix that has the same dimension as the original kernel matrix $K(\mathbf{x_i}, \mathbf{x_j})$ is defined as:

$$
W(\mathrm{i}, \mathrm{j}) = \langle \boldsymbol{\Phi}(u_\mathrm{i}), \boldsymbol{\Phi}(u_\mathrm{j}) \rangle = \exp\left(-\frac{\|u_\mathrm{i} - u_\mathrm{j}\|^2}{2\sigma^2} \right)
\tag{3}
$$

We constructed the new weighted kernel matrix $\mathbf{K_w}$ with $\mathbf{K_W}(\mathrm{i}, \mathrm{j}) = \mathbf{K}(\mathrm{i}, \mathrm{j})^{W(\mathrm{i},\mathrm{j})}$ as:

$$
\mathbf{K_W}(\mathrm{i}, \mathrm{j}) = \mathbf{K}(\mathrm{i}, \mathrm{j})^{W(\mathrm{i},\mathrm{j})} = \exp\left(-\frac{\|\mathbf{x_i} - \mathbf{x_j}\|^2 \exp\left(-\left(\|u_\mathrm{i} - u_\mathrm{j}\|^2\right)/2\sigma^2 \right)}{2\sigma^2} \right)
\tag{4}
$$

The effects of the weights under two conditions can be analyzed: (1) If $u_\mathrm{i} = u_\mathrm{j}$, the samples $\mathbf{x_i}$ and $\mathbf{x_j}$ belong to the same class and $W(\mathrm{i}, \mathrm{j}) = 1$. As observed, the weighted kernel matrix $\mathbf{K_W}$ will be equal to the original kernel matrix $\mathbf{K}$. (2) If $u_\mathrm{i} \neq u_\mathrm{j}$, the $W(\mathrm{i}, \mathrm{j})$ will be a positive value, in which the label information can be embedded in the weighted kernel matrix.

Eigen-decomposed $\mathbf{K_W}$: $\mathbf{K_W} = \mathbf{E_W D_W E_W}^\mathbf{T}$, the eigenvalues $\lambda_{w1}, \lambda_{w2}, ..., \lambda_{wN}$ of the weighted kernel matrix are ranked in decreasing order of the entropies, and $\boldsymbol{\alpha_{w1}}, \boldsymbol{\alpha_{w2}}, ..., \boldsymbol{\alpha_{wN}}$ are the corresponding eigenvectors. The subspace is defined as $\mathbf{U_W}$ spanned by the principal axes that contribute most to the Renyi entropy estimation. Requiring $||\mathbf{u_{wi}}||^2 = 1$, thus $\mathbf{u_{wi}} = \lambda_{wi} \boldsymbol{\Phi} \boldsymbol{\alpha_{wi}}$ can be

obtained. We can project both training and testing samples onto $\mathbf{U_W}$ to extract the intrinsic features. For the out-of-sample data set $\mathbf{x_t}$, the extracted features can be calculated:

$$
\begin{aligned}
\mathbf{y}(\mathbf{x_t}) &= \mathbf{u_{wi}}^T \Phi(\mathbf{x_t}) = \langle \mathbf{U_W}, \Phi(\mathbf{x_t}) \rangle = \left\langle \lambda_{\mathrm{wi}}^{-\frac{1}{2}} \sum_{i=1}^{N} \boldsymbol{\alpha_{wi}} \Phi(\mathbf{x_i}), \Phi(\mathbf{x_t}) \right\rangle \\
&= \lambda_{\mathrm{wi}}^{-\frac{1}{2}} \sum_{i=1}^{N} \boldsymbol{\alpha_{wi}} k_\sigma(\mathbf{x_i}, \mathbf{x_t}) = \mathbf{D_W}^{-\frac{1}{2}} \mathbf{E_W}^{\mathbf{T}} \mathbf{K'}^{\mathbf{T}}
\end{aligned}
\tag{5}
$$

Let $\boldsymbol{\Phi}'$ refer to a collection of the out-of-sample data sets, $\mathbf{K}' = \boldsymbol{\Phi}'^T \boldsymbol{\Phi}$ is the inner product matrix. Then we can extract the first d nonlinear principal components which contribute most to Renyi entropies of the input data by using the weighted kernel matrix. The number, d, of the projection vectors is determined in terms of $\sum_{i=1}^{d} \left(\sqrt{\lambda_i} \boldsymbol{\alpha_i}^{\mathbf{T}} \mathbf{1} \right)^2 / \sum_{j=1}^{N} \left(\sqrt{\lambda_j} \boldsymbol{\alpha_j}^{\mathbf{T}} \mathbf{1} \right)^2 \geq \alpha$ (set to 0.95 here for both KECA and WKECA).

2.3. Selecting Optimal Weights for Weighted Kernel Entropy Component Analysis by Genetic Algorithm

The relevance of different classes leads to diversified generalization performances. Therefore, weights are important to the recognition system, and determination of weights can be considered as an optimization problem. GA is a search and optimization process inspired by the laws of nature evolution and selection [31], which is a powerful intelligent optimization tool based on a group of independent computations controlled by the probabilistic strategy. GA has been widely used in various applications because of its excellent global search ability [31,32]. In this study, we use GA to find the most suitable weights for WKECA where the optimality is defined regarding the recognition accuracy and class separability. The main optimization process can be described as follows:

(1)　Individual encoding: defined the individual is a set of weights l_1, l_2, ..., l_c, the encoding method based on binary for each weight is used.

(2)　Population initialization: an initial population with n_r individuals (set to 20) is randomly created.

(3)　Fitness calculation: the individual selection for the next generation is done based on the fitness. Taking advantage of Liu and Wang's work [19], the fitness function is defined as $f(X) = CA + kR_{BW}$, where CA is the training accuracy which can represent the performance of extracted features, k is a positive constant, and R_{BW} is the Fisher criterion which can indicate the class separability. R_{BW} is the ratio of between-class distance S_b and within-class distance S_w [33]. High classification accuracy and large class separability can be obtained by maximizing the fitness function, which results in evolving more discriminate information than KECA with a proper k. Therefore, good generalization performance for WKECA is possible to be acquired on both training and testing samples.

(4)　Genetic operators: new chromosomes are generated to update and optimize population continuously by genetic operators including selection, cross-over and mutation. The crossover probability and mutation probability are set to 0.7 and 0.01, respectively. The selected probability of every individual is $p_m = \dfrac{f(w_m)}{\sum\limits_{m}^{n_r} f(w_m)}$, m = 1,... , n_r, where $f(w_m)$ is the individual' fitness value.

(5)　Terminating conditions: when the value of fitness does not change again during the iteration procedure or the number of iterations has reached the maximum value (50 in this study) the program will terminate.

2.4. Fault Diagnosis Based on WKECA

The high-dimensional feature set, which can represent well the operating condition of machines, should be first extracted from the raw vibration signals. Generally, the vibration signals of fault bearings are non-stationary, and wavelet packet decomposition (WPD) that can provide a more meticulous analysis is a powerful tool in dealing with non-stationary signals [34]. WPD is effective

for decomposing both high- and mid-frequency information from a signal into the corresponding frequency regions, widely used for fault diagnosis of bearings now [34–38]. In this study, WPD is performed to extract the fault features including the relative energy in a wavelet packet node (REWPN) and the entropy in a wavelet packet node (EWPN). The REWPN indicates the normalized energy of the wavelet packets node, and the EWPN represents the uncertainty of the normalized coefficients of the wavelet packets node [39]. For a given sample $x(n)$, the jth wavelet packet coefficients of the i-th wavelet packet node is defined as $C_i{}^j$, and then REWPN and EWPN can be expressed as follows:

$$\text{REWPN}(i) = \frac{\sum\limits_{j=1}^{K} \left(C_i{}^j \right)^2}{\sum\limits_{m=1}^{N} \sum\limits_{j=1}^{K} \left(C_m{}^j \right)^2} \tag{6}$$

$$\text{EWPN}(i) = -\sum_{j=1}^{K} p_i{}^j \log_2 \left(p_i{}^j \right) \tag{7}$$

where $p_i{}^j = \left(C_i{}^j \right)^2 / \sum_{j=1}^{K} \left(C_i{}^j \right)^2$, N is the total number of wavelet packet nodes, and K is the total number of wavelet packet coefficients in each wavelet packet node.

The REWPNs and EWPNs can truly reflect the diversity among different fault patterns of bearings. They are used as the high-dimensional input vector to WKECA for dimensionality reduction, which can be written as $\mathbf{x_i} = [\text{REWPN}(1), ..., \text{REWPN}(p), \text{EWPN}(1), ..., \text{EWPN}(p)]^T$. Here, p is the number of wavelet packet node. The implementation process of the proposed fault diagnosis method using WKECA for bearings is detailed as shown in Figure 1:

(1) Decomposing the vibration signals into different frequency bands by using WPD, and then we can acquire the high dimensional feature set $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]^T$ including REWPNs and EWPNs, where N is the number of the signal samples.

(2) Carrying out feature extraction to the high-dimensional dataset obtained from vibration signals with WKECA algorithm, capturing their intrinsic manifold structure, and then we can obtain the low-dimensional features by projecting the original high-dimensional observed space into low-dimensional feature space. Meanwhile, the optimal mapping direction can be acquired so that new testing samples can be mapped into the low-dimensional feature space.

(3) Implementing pattern classification of the datasets in the low-dimensional feature space with support vector machine (SVM) classifier.

(4) Determining the type of failures by the classification results, and we can put forward the corresponding decisions or control measures.
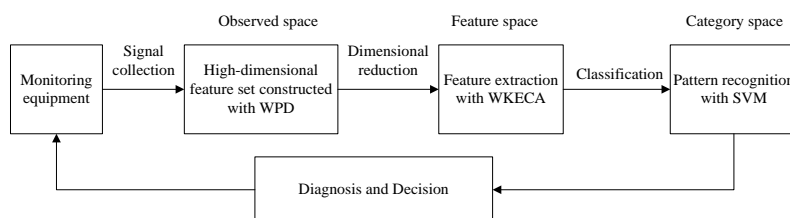

Figure 1. Implementation process of the proposed fault diagnosis method.

3. Experimental Results and Analysis

3.1. Experimental Description

To evaluate the effectiveness of the WKECA, an experimental study on fault diagnosis of rolling bearings was performed. As shown in Figure 2, the tested bearings were delivered through the

automatic machinery system which contained the preset mechanism, the measuring mechanism, the sorting mechanism, and the feeding mechanism [40,41]. The radial vibration signals on one point of the tested bearings were detected by a piezoelectric acceleration sensor (YD-1, Far East Vibration (Beijing) System Engineering Technology Co., Ltd., Beijing, China) located on the top of the bearings, and amplified by a charge amplifier (DHF-2, same company as the sensor). The charge sensitivity and frequency response of the sensor are 6–10 pC/ms^{-2} and 1–10,000 Hz $\pm$ 1 dB, respectively, and the frequency range of the amplifier is 0.3 Hz–100 kHz. Then the signals were converted to voltage signals by an A/D converter (PCI-9114) (ADLINK Technology, Inc., Taiwan) and sent to a computer for further processing. The sampling frequency was 25 kHz, and the rotational speed of the driving motor was set to 1500 rpm.
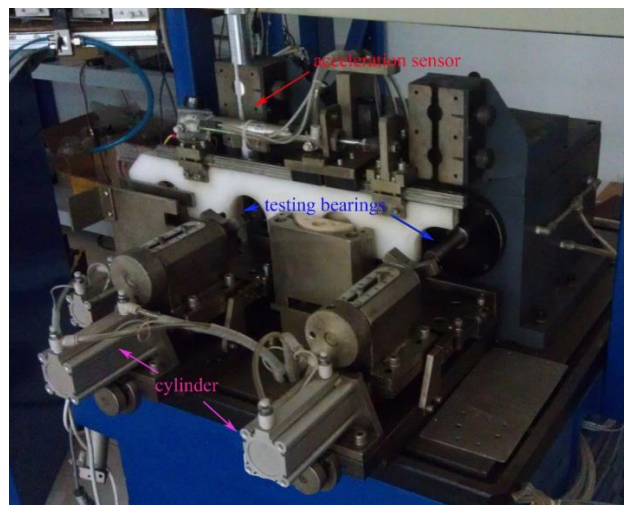
Figure 2. The test rig.

Deep groove ball bearings (6328-2RZ) (Changjiang bearing co., LTD, Chongqing, China) were used as the tested bearings, and four different operating conditions (i.e., inner race fault, outer race fault, ball fault, and normal condition) were simulated in this experiment. Single point defects were introduced to the tested bearings by electric engraving pen, where the widths of the scratch defects were 65 $\pm$ 22 μm, 70 $\pm$ 20 μm, and 70 $\pm$ 20 μm for the inner race, outer race and ball, respectively, and the depths of the scratch defects were 0.2 $\pm$ 0.05 mm. The characteristic bearing defect frequencies can be calculated by [42]:

$$\text{Defect on inner race (BPI)} = \frac{Zf_r}{2}\left(1 + \frac{d}{D}\cos\alpha\right) \tag{8}$$

$$\text{Defect on outer race (BPO)} = \frac{Zf_r}{2}\left(1 - \frac{d}{D}\cos\alpha\right) \tag{9}$$

$$\text{Defect on ball (BS)} = \frac{f_r D}{2d}\left(1 - \frac{d^2}{D^2}\cos^2\alpha\right) \tag{10}$$

where Z is the number of rolling elements, f_r is the rotational frequency, d is the diameter of the rolling element, D is the pitch diameter, and α is the contact angle. According to the kinematic parameters of the tested bearings and the rotational speed, the characteristic bearing defect frequencies of the inner race, outer race and ball are 121.75 Hz, 78.25 Hz and 55 Hz, respectively. Figure 3 indicates the four different vibration signal waveforms in the time-domain together with the amplitude spectrums. The peak values of the accelerations are obtained at 24.42 Hz which is closed to the rotational frequency 25 Hz. As observed, it is difficult to distinguish different faults only from Figure 3 due to the effects of

the noise. The vibration signals under those four conditions are selected as samples, and 100 bearings for each state were tested. Thus, 400 data can be obtained, and the length of each data set is 25,000. The training data set is half samples of the original data set in the experiment.
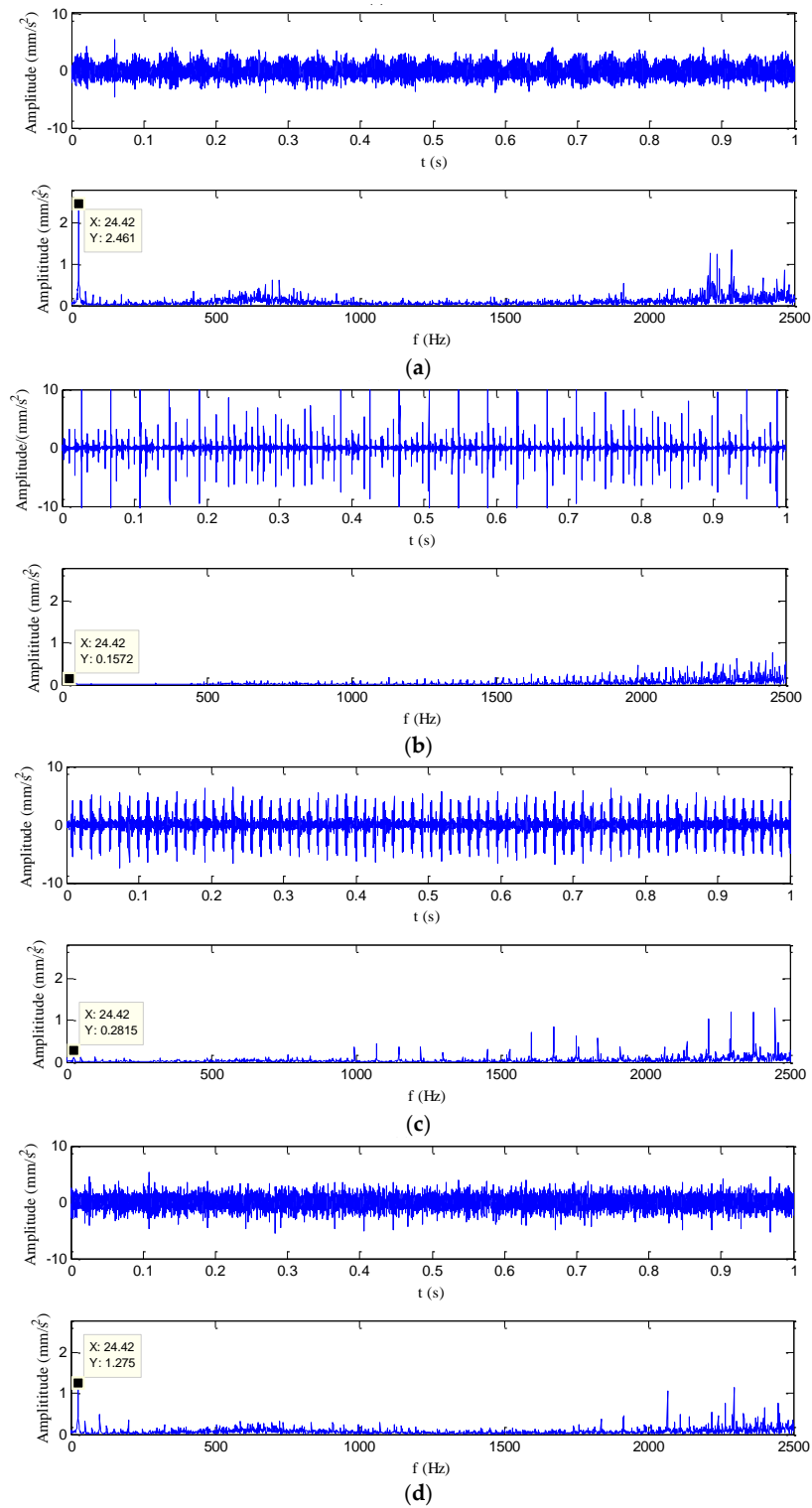

Figure 3. The time domain and frequency domain figures of vibration signals for the four bearing conditions: (**a**) normal condition, (**b**) inner race fault, (**c**) outer race fault, and (**d**) ball fault.

3.2. Dimensionality Reduction and Pattern Classification

The high dimensional feature set containing REWNs and EWPNs are first constructed. The wavelet packet node energy features obtained by Daubechies2 (db2) wavelet packet decomposition were found to achieve the best classification performance for bearing fault diagnosis after many experiments on a serials of Daubechies wavelets [43]. Here the Daubechies2 (db2) is selected as the mother wavelet function to implement binary WPD for vibration signals, where the maximum decomposition level is set to 4. The normalized wavelet packet energy and wavelet packets node entropy spectrums of the bearing vibration signals are shown in Figure 4. Obviously, different bearing faults have different amplitude in different frequency bands. 32 fault features in total including 16 REWPNs and 16 EWPNs are used for fault diagnosis of bearings.
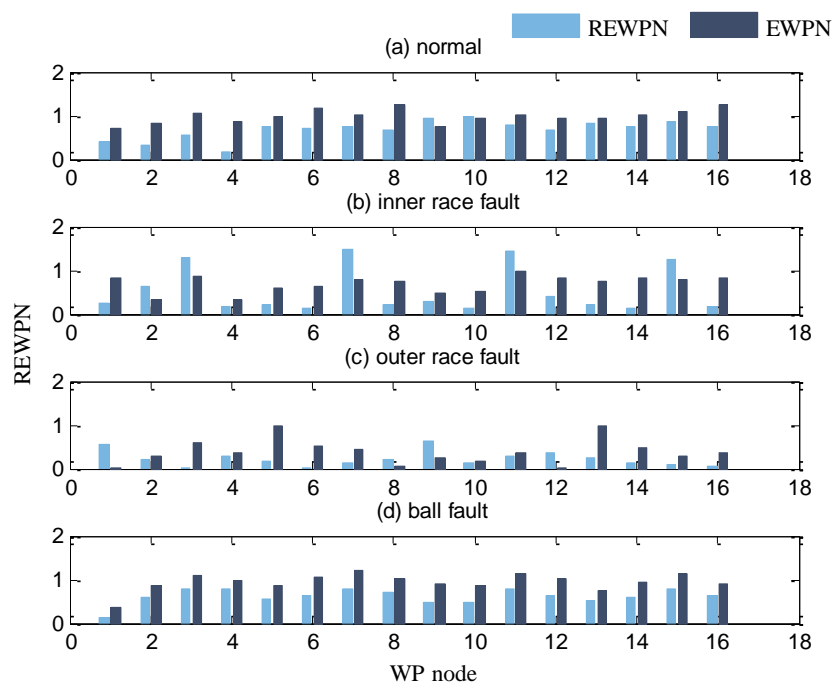

Figure 4. The normalized wavelet packet energy and entropy spectrums of the bearing vibration signals under four conditions: (**a**) normal condition, (**b**) inner race fault, (**c**) ball fault, (**d**) outer race fault.

After the high-dimensional feature set is constructed, it is input into WKECA for non-linear dimension reduction, where the parameter k of the fitness function is set to 0.001. The first d most significant component vectors contributing most to the Renyi entropy are extracted by WKECA, and similar methods including PCA, KPCA and KECA are conducted for comparison. The target dimensionality for every method is set to a certain number so that the cumulative variance contribution rate is more than 95%. For visualization, the plots of the first three principal components of their projection results are shown in Figures 5–8, where Figures 5a, 6a, 7a and 8a represent the training results, and Figure 5b, Figure 6b, Figure 7b, and Figure 8b represent the testing results. It is evident that PCA, KPCA and KECA are not well separated those four classes because some samples are overlapped, which will lead to low recognition accuracy. By contrast, WKECA has little misjudgment samples: the testing points are consistent with the training points in WKECA, and the WKECA algorithm can obviously identify different classes both for the training samples and the testing samples. It proves that WKECA has better clustering performance than PCA, KPCA and KECA, because WKECA introduces the fault class label information and a weight strategy into feature extraction, which is conductive to pattern recognition.
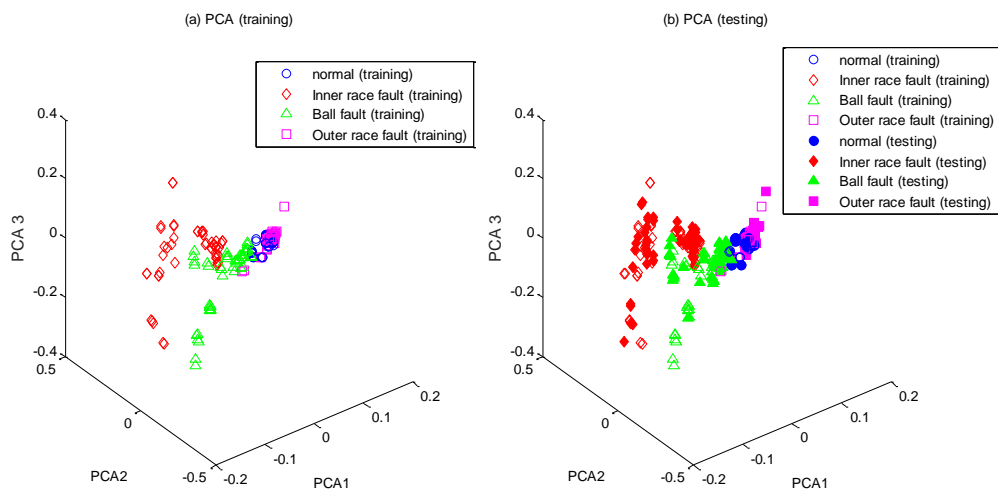
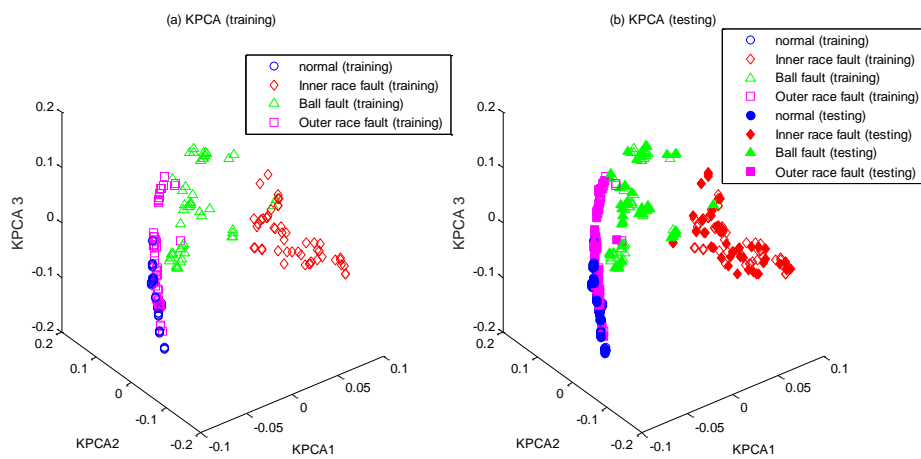**Figure 5.** Feature extraction with PCA: (**a**) training samples, (**b**) testing samples.

Figure 6. Feature extraction with KPCA: (**a**) training samples, (**b**) testing samples.
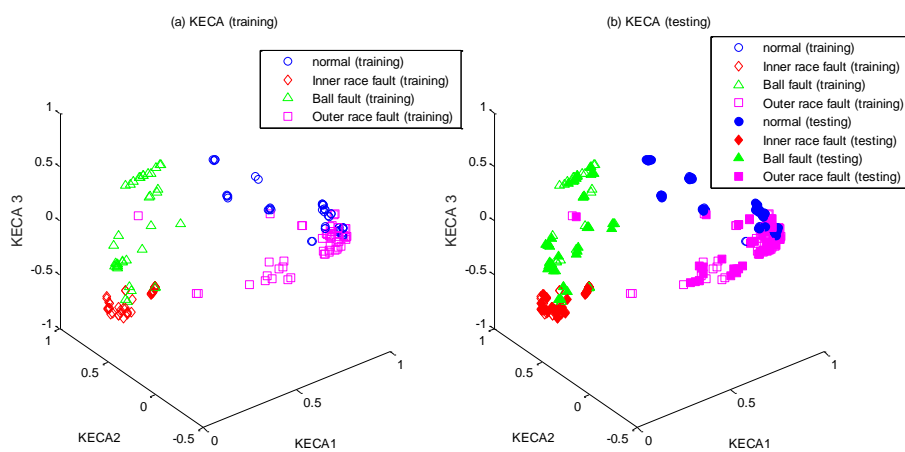
Figure 7. Feature extraction with KECA: (**a**) training samples, (**b**) testing samples.
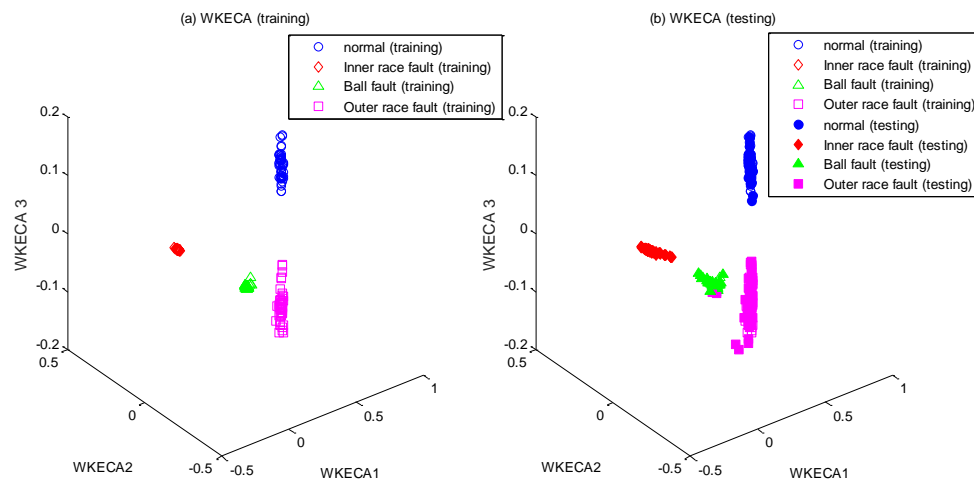
Figure 8. Feature extraction with WKECA: (**a**) training samples, (**b**) testing samples.

3.3. Results and Discussion

Within the fault diagnosis related to pattern recognition in conjunction with feature extraction techniques that find low-dimensional representation for samples, classifiers are needed to identify those different bearing faults. Support vector machine (SVM) is adopted for its well-developed statistical learning theory. 50 data from inner race fault, outer race fault, ball fault, and normal condition were selected randomly for SVM training and the others were used for testing. The quantitative evaluation procedure for SVM, PCA-SVM, KPCA-SVM, KECA-SVM, and WKECA-SVM were repeated for 10 times. In order to highlight the effectiveness of the proposed WKECA-SVM method, the fault detection rate of the method was compared with the results of the other four methods. The testing average results are summarized in Table 1, and the classification accuracies are 77.5%, 83%, 89.5%, 93% and 97%. The results demonstrate that satisfactory overall classification results have been achieved by means of the dimension reduction, and the classification accuracy is significantly improved by introducing WKECA. WKECA performs better than the other methods in terms of extracting discriminative features which can lead to high classification rates. Therefore, WKECA is suitable as a feature extraction step prior to classification, and functions well for fault patterns recognition.

Table 1. The classification accuracies of different methods to the bearing sets with support vector machine (SVM) classifier.

Operating Condition	Normal (%)	Inner Race Fault (%)	Outer Race Fault (%)	Ball Fault (%)	Average Accuracy (%)
Original	68	86	76	80	77.5
PCA	72	90	88	82	83
KPCA	92	92	84	90	89.5
KECA	96	98	82	96	93
WKECA	100	100	92	96	97

To obtain discriminative representations through GA, a suitable fitness function is important to the whole recognition procedure. Therefore, it is necessary to know the effects of the parameter *k* in fitness function. Table 2 presents the results of evolutionary process with different *k*, where CA_{test} is the testing accuracy. It is obvious that R_{BW} increases with the raising of *k* while CA_{test} decreases accordingly. This observation reflects that *k* can adjust the contribution of class separability to the fitness function, and a proper *k* can lead to larger R_{BW} as well as good classification performance.

Table 2. The results of evolutionary process with different values of parameter k.

Performance	k = 0.001	k = 0.01	k = 0.1	k = 1
$f(X)$	0.9702	0.9939	1.0236	1.2328
R_{BW}	1.4506	1.4875	1.7913	2.0828
CA_{test}	0.97	0.965	0.935	0.905

In order to investigate the performance of WKECA in handling the Small Sample Size (SSS) problem with different training sample sizes, PCA, KPCA and KECA were conducted for comparison. Figure 9 presents the recognition rates of the four feature extraction methods and the original features with different numbers of labeled samples. It is obvious that the classification accuracy increases with the raising of training sample sizes. This reveals that the feature extraction based on manifold learning can improve the recognition performance, and WKECA performs better than other methods in achieving high classification accuracy. The effects of SSS problem are obvious in other methods when only ten samples are used for training, while WKECA is less sensitive to the training sample size. This proves that WKECA can capture the intrinsic geometric structure embedded in the data and achieve efficient performance in feature extraction and classification.
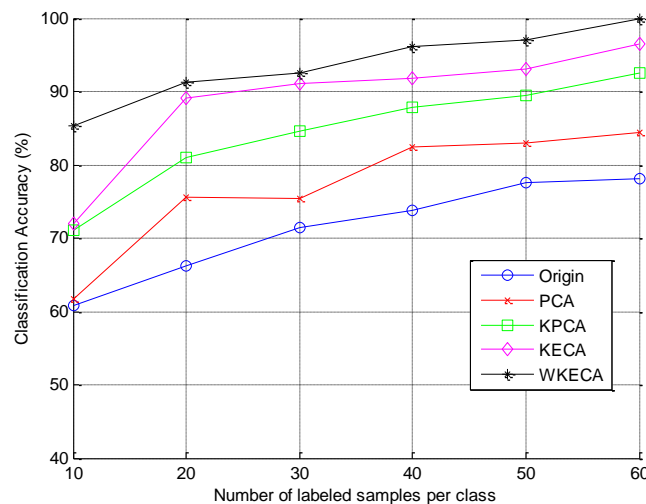
Figure 9. Classification accuracy of SVM based on different feature extraction methods for different labeled samples.

4. Conclusions

In this study, a new feature extraction method called weighted entropy component analysis (WKECA) is proposed for fault diagnosis of rolling bearings. It makes the most of the labeled information and introduces a weight strategy in feature extraction, and GA is performed to find optimal weights for achieving high training classification results. The original high-dimensional feature sets are first constructed based on WPD which can provide a more meticulous analysis for signals. WKECA is then used to extract the intrinsic independent features among the multiple manifolds to reflect the states of the rolling bearings. Finally, the extracted intrinsic geometric features are fed into SVM to recognize different operating conditions of bearings. WKECA outperforms PCA, KPCA and KECA in terms of achieving higher testing accuracies. The results demonstrate the feasibility and effectiveness of the proposed method for fault diagnosis of rolling bearings. Next, we are trying to extend our approach to diagnose different faults magnitudes in different machines. The challenge is the great time consumption for training, which is inevitable confronted by almost all evolutionary processes for pattern recognition. Therefore, fast optimal strategies are deserved for further investigation.

Author Contributions: Each author contributed extensively to the preparation of this manuscript. Lei Su and Zhenzhi He developed the testing hardware system; Hongdi Zhou and Zhenzhi He performed the experiments; Hongdi Zhou, Tielin Shi, Guanglan Liao, Jianping Xuan and Duan Jie analyzed the data; Hongdi Zhou, Wuxing Lai and Guanglan Liao wrote the paper.

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

References

1. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [CrossRef]

2. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Proc.* **2015**, *64*, 100–131. [CrossRef]

3. Cong, F.; Zhong, W.; Tong, S.; Tang, N.; Chen, J. State Space Formulation of Nonlinear Vibration Responses Collected from a Dynamic Rotor-Bearing System: An Extension of Bearing Diagnostics to Bearing Prognostics. *Sensors* **2017**, *17*, 369.

4. Henao, H.; Capolino, G.A.; Fernandez-Cabanas, M.; Filippetti, F. Trends in fault diagnosis for electrical machines: a review of diagnostic techniques. *IEEE Ind. Electron. Mag.* **2014**, *8*, 31–42. [CrossRef]

5. Frosini, L.; HarlişCa, C.; Szabó, L. Induction machine bearing fault detection by means of statistical processing of the stray flux measurement. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1846–1854. [CrossRef]

6. Li, Y.; Xu, M.; Wang, R.; Huang, W. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **2016**, *360*, 277–299. [CrossRef]

7. Yang, Y.; Dong, X.J.; Peng, Z.K.; Zhang, W.M.; Meng, G. Vibration signal analysis using parameterized time–frequency method for features extraction of varying-speed rotary machinery. *J. Sound Vib.* **2015**, *335*, 350–366. [CrossRef]

8. Immovilli, F.; Bellini, A.; Rubini, R.; Tassoni, C. Diagnosis of bearing faults in induction machines by vibration or current signals: a critical comparison. *IEEE Trans. Ind. Appl.* **2008**, *46*, 1350–1359. [CrossRef]

9. Tang, B.; Song, T.; Li, F.; Deng, L. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. *Renew. Energ.* **2014**, *62*, 1–9. [CrossRef]

10. Lei, Y.; He, Z.; Zi, Y.; Hu, Q. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mech. Syst. Signal Proc.* **2007**, *21*, 2280–2294. [CrossRef]

11. Yu, J. Machinery fault diagnosis using joint global and local/nonlocal discriminant analysis with selective ensemble learning. *J. Sound Vib.* **2016**, *382*, 340–356. [CrossRef]

12. Li, B.; Zhang, Y. Supervised locally linear embedding projection (SLLEP) for machinery fault diagnosis. *Mech. Syst. Signal Proc.* **2011**, *25*, 3125–3134. [CrossRef]

13. Liao, G.; Liu, S.; Shi, T.; Zhang, G. Gearbox condition monitoring using self-organizing feature maps. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **2004**, *218*, 119–129. [CrossRef]

14. Su, L.; Shi, T.; Liu, Z.; Zhou, H.; Du, L.; Liao, G. Nondestructive diagnosis of flip chips based on vibration analysis using PCA-RBF. *Mech. Syst. Signal Proc.* **2017**, *85*, 849–856. [CrossRef]

15. Shao, R.; Hu, W.; Wang, Y.; Qi, X. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement* **2014**, *54*, 118–132. [CrossRef]

16. Liu, H.; Zhang, J.; Cheng, Y.; Lu, C. Fault diagnosis of gearbox using empirical mode decomposition and multi-fractal detrended cross-correlation analysis. *J. Sound Vib.* **2016**, *385*, 350–371. [CrossRef]

17. Trendafilova, I.; Cartmell, M.P.; Ostachowicz, W. Vibration-based damage detection in an aircraft wing scaled model using principal component analysis and pattern recognition. *J. Sound Vib.* **2008**, *313*, 560–566. [CrossRef]

18. Schölkopf, B.; Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [CrossRef]

19. Liu, N.; Wang, H. Weighted principal component extraction with genetic algorithms. *Appl. Soft. Comput.* **2012**, *12*, 961–974. [CrossRef]

20. Zhang, Z.H.; Hancock, E.R. Kernel entropy-based unsupervised spectral feature selection. *Int. J. Pattern Recognit. Artif. Intell.* **2012**, *26*, 1260002. [CrossRef]

21. Jenssen, R. Kernel Entropy Component Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 847–860. [CrossRef] [PubMed]

22. Shi, J.; Jiang, Q.; Zhang, Q.; Huang, Q.; Li, X. Sparse kernel entropy component analysis for dimensionality reduction of biomedical data. *Neurocomputing* **2015**, *168*, 930–940. [CrossRef]

23. Yang, Y.; Li, X.; Liu, X.; Chen, X. Wavelet kernel entropy component analysis with application to industrial process monitoring. *Neurocomputing* **2015**, *147*, 395–402. [CrossRef]

24. Gómez-Chova, L.; Jenssen, R.; Camps-Valls, G. Kernel entropy component analysis for remote sensing image clustering. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 312–316. [CrossRef]

25. Jenssen, R. Kernel Entropy Component Analysis: New Theory and Semi-Supervised Learning. In Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing, Beijing, China, 18–21 September 2011; pp. 1–6.

26. Fukunaga, K. *Introduction to Statistical Pattern Recognition*, 2nd ed.; Rheinboldt, W., Ed.; Academic press: Cambridge, MA, USA, 1990.

27. Sierra, A.; Echeverría, A. Evolutionary discriminant analysis. *IEEE Trans. Evol. Comput.* **2006**, *10*, 81–92. [CrossRef]

28. Renyi, A. On measures of entropy and information. *Fourth Berkeley Symp. Math. Statist. Prob.* **1961**, *1*, 547–561.

29. Jenssen, R. Information Theoretic Learning and Kernel Methods. In *Information Theory and Statistical Learning*; Emmert-Streib, F., Dehmer, M., Eds.; Springer: New York, NY, USA, 2009; pp. 209–230.

30. Gao, L.; Qi, L.; Chen, E.; Guan, L. A fisher discriminant framework based on Kernel Entropy Component Analysis for feature extraction and emotion recognition. In Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Chengdu, China, 14–18 July 2014; pp. 1–6.

31. Mitchell, M. *An Introduction to Genetic Algorithms*; MIT Press: London, UK, 1998.

32. Tang, K.S.; Man, K.F.; Kwong, S.; He, Q. Genetic algorithms and their applications. *IEEE Signal Process. Mag.* **1996**, *13*, 22–37. [CrossRef]

33. He, Q.; Kong, F.; Yan, R. Subspace-based gearbox condition monitoring by kernel principal component analysis. *Mech. Syst. Signal Proc.* **2007**, *21*, 1755–1772. [CrossRef]

34. Chen, X.; Liu, D.; Xu, G.; Jiang, K.; Liang, L. Application of wavelet packet entropy flow manifold learning in bearing factory inspection using the ultrasonic technique. *Sensors* **2015**, *15*, 341–351. [CrossRef] [PubMed]

35. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]

36. Lei, Y.; He, Z.; Zi, Y. Application of an intelligent classification method to mechanical fault diagnosis. *Expert Syst. Appl.* **2009**, *36*, 9941–9948. [CrossRef]

37. Hu, Q.; He, Z.; Zhang, Z.; Zi, Y. Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mech. Syst. Signal Proc.* **2007**, *21*, 688–705. [CrossRef]

38. Li, B.; Chen, X. Wavelet-based numerical analysis: A review and classification. *Finite Elem. Anal. Des.* **2014**, *81*, 14–31. [CrossRef]

39. Feng, Y.; Schlindwein, F.S. Normalized wavelet packets quantifiers for condition monitoring. *Mech. Syst. Signal Proc.* **2009**, *23*, 712–723. [CrossRef]

40. Chen, Y.; He, Z.; Yang, S. Research on on-line automatic diagnostic technology for scratch defect of rolling element bearings. *Int. J. Precis. Eng. Manuf.* **2012**, *13*, 357–362. [CrossRef]

41. Zhou, H.; Shi, T.; Liao, G.; Xuan, J.; Su, L.; He, Z.; Lai, W. Using supervised kernel entropy component analysis for fault diagnosis of rolling bearings. *J. Vib. Control* **2015**. [CrossRef]

42. Tandon, N.; Choudhury, A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribol. Int.* **1999**, *32*, 469–480. [CrossRef]

43. Jiang, L.; Shi, T.; Xuan, J. Fault diagnosis of rolling bearings based on marginal fisher analysis. *J. Vib. Control* **2014**, *20*, 470–480. [CrossRef]