Article

Combining Multi-Scale Wavelet Entropy and Kernelized Classification for Bearing Multi-Fault Diagnosis

Nibaldo Rodriguez [1,*] **, Pablo Alvarez** [1] **, Lida Barba** [2] **and Guillermo Cabrera-Guerrero** [1]

[1] Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso, 2374631 Valparaíso, Chile; pablo.alvarez.a01@mail.pucv.cl (P.A.); guillermo.cabrera@pucv.cl (G.C.-G.)

[2] Facultad de Ingeniería, Universidad Nacional de Chimborazo, 060102 Riobamba, Ecuador; lbarba@unach.edu.ec

* Correspondence: nibaldo.rodriguez@pucv.cl; Tel.: +56-32-227-3761

Abstract: Discriminative feature extraction and rolling element bearing failure diagnostics are very important to ensure the reliability of rotating machines. Therefore, in this paper, we propose multi-scale wavelet Shannon entropy as a discriminative fault feature to improve the diagnosis accuracy of bearing fault under variable work conditions. To compute the multi-scale wavelet entropy, we consider integrating stationary wavelet packet transform with both dispersion (SWPDE) and permutation (SWPPE) entropies. The multi-scale entropy features extracted by our proposed methods are then passed on to the kernel extreme learning machine (KELM) classifier to diagnose bearing failure types with different severities. In the end, both the SWPDE–KELM and the SWPPE–KELM methods are evaluated on two bearing vibration signal databases. We compare these two feature extraction methods to a recently proposed method called stationary wavelet packet singular value entropy (SWPSVE). Based on our results, we can say that the diagnosis accuracy obtained by the SWPDE–KELM method is slightly better than the SWPPE–KELM method and they both significantly outperform the SWPSVE–KELM method.

Keywords: stationary wavelet transform; multi-scale entropy; Kernel Extreme Learning Machine

1. Introduction

Early diagnosis of failures of bearings is a key factor to improve both safety and reliability of rotating machinery, intensively used in industrial environments. During the last years, several vibration signal analysis methods has been used to achieve early bearing fault diagnosis. Among them, we can find the empirical mode decomposition (EMD) [1], the local mode decomposition (LMD) [2,3] and the wavelet transform (WT) [4]. While the EMD method can self-adaptively decompose a signal into some intrinsic mode functions (IMFs) based on the local characteristic time scale of the signal [5], the LMD method also self-adaptively decomposes a signal into a series of product functions (PFs), each of which is exactly a mono-component signal [2]. Unlike the EMD and LMD, the WT decomposes a signal into several scales using a wavelet base function. One notable feature of this function is that it can show features of hidden failures [6–9]. Based on these time-frequency methods for signal decomposition, different entropy features have been used such as Wiener-Shannon's entropy [10,11], energy entropy [12,13], wavelet energy entropy [14], samples entropy [15], multiscale entropy [16,17], permutation entropy (PE) [18–21], multi-scale permutation entropy [22,23], generalized composite multiscale permutation entropy [24], multi-scale fuzzy entropy [25], composite multi-scale fuzzy entropy [26], dispersion entropy (DE) [27], multiscale dispersion entropy [28], and improved multiscale

dispersion entropy [29]. These entropy features are, in turn, passed on to classifiers such as artificial neural networks (ANN) [3,30–32] or support vector machines (SVM) [12,17,18,24,26,33,34].

In particular, authors in [12] used IMFs' energy entropy to determine whether a failure exists or not. In case of failure, a vector of singular values is passed on to an SVM in order to determine the type of failure. In order to determine whether there is a failure or not, the authors in [18] proposed a hybrid model based on permutation entropy (PE). In case a failure actually exists, the PE of a subset of selected IMFs is calculated and used as the input of an SVM. The SVM will, then, classify the type and severeity of the failure. Yongbo Li [35] investigated the LMD method combined with an improved multi-scale fuzzy entropy. They also used the SVM for the fault diagnosis of the rolling bearings. Authors in [26] studied composite multiscale fuzzy entropy (CMFE) to extract the hidden nonlinear features from vibration signals and then the CMFE features were used as the input of an ensemble SVM to improve rolling bearing fault diagnosis. As we can see, both SVM and ANN are commonly used for the classification of different types of failures in rotary machines. Unfortunately, during training stages, these methods are quite time consuming, which makes them not very efficient.

To overcome the weaknesses of the vector support machine and the neural network, Huang et al. [36–38] proposed a new learning algorithm called extreme learning machine (ELM), which aims to improve tuning time in single-hidden layer feed-forward neural networks. Since then, many researchers have adopted ELM in their works mainly because of its efficiency. For instance, authors in [3] combined LMD and ELM and singular value decomposition (SVD) for bearing failure diagnosis. Here, SV are obtained from the product function matrix are passed on to the ELM as its input. The authors in [3] also demonstrated that LMD–SVD–ELM models performed better than EMD–SVD–ELM models. Authors in [32], proposed an ELM model that is combined with a real-valued gravitational search algorithm. They proposed to use the ensambled EMD method as their classifier. Here, energy features, time–frequency features and SV features are computed using the the ensambled EMD method obtaining very good results when applied on bearing fault diagnosis. In a previous work [39], we proposed an ELM classifier based on a combination of stationary wavelet transform (SWT) and SVD. The SWT is used to separate the vibration signals into a series of wavelet component signals. Then, the obtained wavelet component matrix is decomposed by means of a SVD method to obtain a set of wavelet singular values. Finally, the wavelet singular values are used as input to the ELM for classification among ten different bearing failure types. More recently, in [40] we modified the strategy proposed in [39] by replacing the ELM classifier by the Kernel–ELM (KELM) classifier. Including the KELM classifier led us to better results compared to the ELM classifier. This is mainly because the KELM classifier includes two extra features called the wavelet singular value entropy and the Shannon entropy of the raw vibration signal.

Based on our previous results [39,40], including extensions of the Shannon entropy seems to be an efficient strategy to improve the accuracy of the bearing fault diagnosis. Thus, in this article, we consider integrating stationary wavelet packet (SWP) transform with both dispersion (SWPDE) and permutation (SWPPE) entropies. The SWP transform is an extension of the wavelet transform [41–45]. It has a more flexible decomposition capacity in time and frequency, especially in the high frequency region, and also it is able to distinguish sudden changes in the bearing vibration signal. After the entropy features extraction, the KELM classifier is used to perform automatic fault diagnosis. The KELM classifier is created by replacing the ELM's hidden activation function with a Gaussian kernel function and so improves the generalisation performance of ELM and reduces time consumption for determining the number of hidden layer nodes [36–38]. We choose to use KELM as it has been shown to be very efficient in both classification accuracy and tuning time [37]. Using the proposed extraction methods we can create discriminative fault features by calculating the entropy value (either DE or PE) of each wavelet sub-band signal obtained from the raw vibration signal. Furthermore, discriminative fault features obtained using the proposed methods are more effective than the ones obtained by using both the multi-scale DE [28,29] and multi-scale PE itself [22,23].

We apply our diagnosis methods on two bearing vibration signal databases under variable work conditions obtained in [46,47]. Using these datasets, a comparison among the accuracy obtained by our methods and results obtained by the stationary wavelet packet singular value entropy (SWPSVE)–KELM in [40] is performed.

This work is organized as follows; in Section 2 we present a short description of stationary wavelet packet transform and three different measurements of Shannon entropy. In Section 3 we describe the bearing multi-fault diagnosis algorithm implemented in this paper and the setup we consider for our experiments. In Section 4, we analyse the results obtained by our algorithms. We draw some conclusions in Section 5.

2. Wavelet Analysis and Entropy Measures

In this section we briefly introduce stationary wavelet packet transform (SWPT) and three Shannon entropy measures, namely SWPDE, SWPPE, and SWPSVE.

2.1. Stationary Wavelet Packet Transform

The SWPT is similar to both the stationary wavelet transform [41–43] and discrete wavelet transform (DWT) [44,45]. At the first level of wavelet decomposition, an input signal $\{x(n) = w_{0,0}(n), n = 1, \ldots, N\}$ is convolved with a low-pass filter h_1 defined by a sequence $h_1(n)$ of length r and a high-pass filter g_1 defined by a sequence $g_1(n)$ of length r. Both, the approximation coefficient $w_{1,1}$ and the detail coefficient $w_{1,2}$ are obtained as follows:

$$w_{1,1}(n) = \sum_{k=0}^{r-1} h_1(k)w_{0,0}(n-k) \tag{1a}$$

$$w_{1,2}(n) = \sum_{k=0}^{r-1} g_1(k)w_{0,0}(n-k). \tag{1b}$$

Since no sub-sampling is performed, the obtained sub-band signals $w_{1,1}(n)$ and $w_{1,2}(n)$ have the same number of elements as the input signal $w_{0,0}(n)$. Filters h_j and g_j are computed by using an operator called dyadic up-sampling. Using this operator, zero values are inserted between each pair of elements in the filter that are adjacent. Thus, the SWPT is defined by the pair of filters (low- and high-pass filters) that is chosen and the number of decomposition steps J. For this paper, a pair of Db2 wavelet filters has been chosen [44,47]. In the literature, wave filters with order greater than two have also been proposed [14]. Although, wave filters with order greater than two have better discriminatory potential both in time and frequency domains [14], we found that increasing the order of the wave filters does not lead to better diagnosis accuracy levels. Thus, we chose to use the simplest mother wavelet filter, i.e., Db2.

The general process of the SWPT is continued recursively for $j = 2, \ldots, J$ as follows:

$$w_{j,2i-1}(n) = \sum_{k=0}^{r-1} h_j(k)w_{j-1,i}(n-k) \tag{2a}$$

$$w_{j,2i}(n) = \sum_{k=0}^{r-1} g_j(k)w_{j-1,i}(n-k), \tag{2b}$$

where the i value denotes the i-th sub-band at the $(j-1)$-th level and the number of sub-bands at the $(j-1)$-th level is equal to $i = 1, \ldots, 2^{j-1}$.

2.2. Stationary Wavelet Packet Dispersion Entropy

Let $w(n)$ be a signal corresponding to one of the $D = 2^J$ wavelet sub-band components, then its stationary wavelet packet dispersion entropy (SWPDE) is calculated through the following steps [48]:

Step 1: The wavelet sub-band signal $\{w(n)\}$ is normalized between 0 and 1 using the normal cumulative distribution function as follows:

$$y(n) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{w(n)} exp\left[\frac{-(t-\mu)^2}{2\sigma^2}\right] dt, \tag{3}$$

where μ and σ are the mean and standard deviation of the raw vibration signal of N data points.

Step 2: The normalized signal $y(n)$ is mapped into c classes with integer indices from 1 to c using the following equation:

$$z^c(n) = round\,(c \cdot y(n) + 0.5) \quad n = 1, 2, \ldots, N, \tag{4}$$

where $round(\cdot)$ denotes the rounding operation.

Step 3: Create multiples m-dimensional vector $z_i^{c,m}$ as follows:

$$z_i^{c,m} = [z^c(i), z^c(i+1), \ldots, z^c(i+m-1)], \quad i = 1, 2, \ldots, N-m+1. \tag{5}$$

Step 4: Each embedding vector $z_i^{c,m}$ is mapped into a dispersion pattern $\pi_{v_0, v_1, \ldots, v_{m-1}}$, where $z^c(i) = v_0, z^c(i+1) = v_1, \ldots, z^c(i+(m-1) = v_{m-1}$. Thus, the number of possible dispersion patterns is equal to c^m.

Step 5: Calculate the probability of occurrence for each permutation pattern $\pi_{v_0, v_1, \ldots, v_{m-1}}$ as follows:

$$p(\pi_{v_0, v_1, \ldots, v_{m-1}}) = \frac{\mathbf{Number}\left\{i | i = 1, 2, \ldots, N-m+1; z_i^{c,m} \quad \mathbf{has\ type} \quad \pi_{v_0, v_1, \ldots, v_{m-1}}\right\}}{N-m+1}, \tag{6}$$

where $N-m+1$ denotes the total of embedding vectors.

Step 6: Calculate the normalized SWPDE of the i-th wavelet sub-band signal $w(n)$ using Equation (7):

$$SWPDE[w(n)] = \frac{-1}{log\,(c^m)} \sum_{\pi=1}^{c^m} p(\pi_{v_0, v_1, \ldots, v_{m-1}}) log\,\left(p(\pi_{v_0, v_1, \ldots, v_{m-1}})\right). \tag{7}$$

Here, for all the experimental examples, the embedding dimension is set to $m = 2$ and the number of classes is in the range $c = [5, 8]$ [27,29,48,49].

2.3. Stationary Wavelet Packet Permutation Entropy

The stationary wavelet packet permutation entropy (SWPPE) of a wavelet sub-band signal $\{w(n) = w_i(n), i = 2^J, n = 1, \ldots, N\}$ obtained by using Equations (1) and (2) is calculated through the following steps [50]:

Step 1: Create a set of m-dimensional vectors W_i^m as follows:

$$W_i^m = [w(i), w(i+1), \ldots, w(i+m-1)], \quad i = 1, 2, \ldots, N-m+1, \tag{8}$$

where m is the embedding dimension of the vector W_i^m.

Step 2: Each vector W_i^m is sorted in ascending order with permutation pattern π as follows:

$$W_i^m = [w(i+j_1-1) \le w(i+j_2-1), \le \ldots \le, w(i+j_m-1)] \tag{9a}$$
$$\pi = [j_1, j_2, \ldots, j_m], \tag{9b}$$

where each vector W_i^m in m-dimensional space can be mapped to one of the $m!$ ordinal patters π.

Step 3: Calculate the probability of occurrence for each permutation pattern π as follows:

$$p(\pi) = \frac{\mathbf{Number}\left\{i | i = 1, 2, \ldots, N - m + 1; W_i^m \quad \mathbf{has\ type} \quad \pi\right\}}{N - m + 1}, \tag{10}$$

where $N - m + 1$ denotes the total of embedding vectors.

Step 4: Calculate the normalized SWPPE of the *i*-th wavelet sub-band signal $w(n)$ using Equation (11):

$$SWPPE[w(n)] = \frac{-1}{log\,(m!)} \sum_{j=1}^{m!} p\,(\pi_j)\, log\left(p\,(\pi_j)\right). \tag{11}$$

Here, for all the experimental examples, the embedding dimension is in the range $m = [4, 7]$ [50].

2.4. Stationary Wavelet Packet Singular Value Entropy

The stationary wavelet packet singular value entropy (SWPSVE) of the wavelet coefficients matrix W is calculated as follows:

$$SWPSVE(W) = -\frac{1}{log_2(K)} \sum_{k=1}^{K} p(k)log_2(p(k)) \qquad K = 2^j \tag{12a}$$

$$p(k) = \frac{s_k^2}{\sum_{k=1}^{K} s_k^2}, \tag{12b}$$

where s_k corresponds to the *k*-th SV of wavelet packet coefficients matrix W, which are obtained using the singular value decomposition method as follows [51]:

$$W = \sum_{k=1}^{K} s_k u_k v_k^T = USV^T, \tag{13}$$

where $U \in R^{K \times K}$, $V \in R^{N \times N}$ represent mutually orthogonal elementary matrices and S denotes the $K \times N$ diagonal singular values matrix.

3. Bearing Fault Diagnosis Algorithm

The algorithm for failure diagnosis presented in this study consists of two phases; the entropy features extraction phase and the classification phase. While the discriminative features extraction phase is carried out by integrating stationary wavelet packet transform and both the dispersion and permutation entropy, the multi-fault classification is performed by means of a KELM model based on the Gaussian kernel function and the *k*-fold cross validation method. We describe this phases in the next sections.

3.1. Proposed Diagnosis Algorithm

The steps of the bearing fault diagnosis algorithms proposed in this paper are as follows:

Step 1: Divide the discrete time raw vibration signal into multiple non-overlapped signals of N data points.

Step 2: Decompose the non-overlapping signals $x(n), n = 1, \ldots, N$ into $D = 2^J$ sub-band signals by using SWPT given as Equations (1) and (2).

Step 3: Create a D-dimensional features vector based on multi-scale wavelet Shannon entropy as follows:

$$u_k = [1/E_1, 1/E_2, \ldots, 1/E_i, \ldots, 1/E_D], \tag{14}$$

where E_i represents one of the SWPDE/SWPPE/SWPSVE value of the *i*-th wavelet sub-band signal and *k* corresponds to the *k*-th non-overlapping raw vibration signal.

Step 4: Normalize the features matrix *Z* as follows:

$$z_i = \frac{u_i - u_{i,min}}{u_{i,max} - u_{i,min}} \quad i = 1, 2, \ldots, D, \tag{15}$$

where z_i corresponds to the *i*-th column of the feature matrix *Z*, $u_{i,min}$ and $u_{i,max}$ denote the minimum value and maximum value of the z_i vector, respectively.

Step 5: Create the KELM classifier based on both the feature matrix *Z* and *k*-fold cross-validation method.

3.2. Kernel-ELM Classifier

In this section we present a brief description of KELM and its main characteristics, based on our previous work on ELM [39] and KELM [40]. For more details on this topic see [36–38,52].

The KELM classifier output is obtained as follows:

$$\hat{Y}(z) = \begin{pmatrix} \ker(\tilde{z}, z_1) \\ \ker(\tilde{z}, z_2) \\ \vdots \\ \ker(\tilde{z}, z_{M2}) \end{pmatrix} \beta \tag{16a}$$

$$\beta = \left(\frac{I_{M1}}{C} + Ker(\tilde{z}, \tilde{z}) \right)^{\dagger} Y, \tag{16b}$$

where $\tilde{z} \in R^{D \times M1}$ represent the set of input vectors to train, $z \in R^{D \times M2}$ denotes the set of input vectors to test, *M*1 and *M*2 represent the samples number of training and testing, respectively. The function $\ker(\cdot)$ denotes the Gaussian kernel given as:

$$\ker(\tilde{z}_i, \tilde{z}_j) = exp\left(- \frac{\|\tilde{z}_i - \tilde{z}_j\|^2}{2\sigma^2} \right), \tag{17}$$

where the σ parameter corresponds to the kernel width and the σ parameter is set to $\sigma^2 = log_{10}(D)$, the *D* value corresponds to the dimensionality of the input features vector to the KELM classifier (see Equation (14)). The I_{M1} is the identity matrix, the β values are output weights of the KELM classifier and the *C* parameter corresponds to the regularisation value. The $(\cdot)^{\dagger}$ expression corresponds to the Moore–Penrose generalized inverse matrix [53] and *Y* corresponds to the desired output pattern matrix.

Finally, the class label predicted for sample *z* is computed as follows:

$$Label\left(\hat{y}(z) \right) = max\{\hat{y}_1(z), \ldots, \hat{y}_{10}(z), \hat{y}_{11}(z), \hat{y}_{12}(z)\}. \tag{18}$$

Using the 5-Fold Cross-Validation (CV) method [54,55], the regularisation parameter *C* is chosen from the range $\{10^1, \ldots, 10^6\}$.

3.3. Experimental Setup

In this paper we considered experimental raw data obtained from vibration signals coming from two bearings; the drive-end (6205-2RS JEM SKF, deep groove ball bearing) and the fan-end (6203-2RS JEM SKF, deep groove ball bearing) bearings. These two datasets were obtained from [46]. An experimental setup as the one shown in Figure 1 was used to generate this dataset. This setup consisted of a 2 hp Reliance electric motor, a dynamometer and a torque transducer/encoder. The bearing held the motor shaft during the experiments. In order to collect vibration signals, an accelerometer mounted on the motor housing, as the one shown in Figure 1, was used. Single point failures with different failure diameters of 7, 14, 21 and 28 mils (1 mils = 0.001 inch) were introduced to both the driving-end

and the fan-end bearings using the electro-discharge machining method, with the motor speed varied at 1730 r/min, 1750 r/min, 1772 r/min, and 1797 r/min for loads of 3, 2, 1, and 0 hp, respectively. Digital data was produced at 12,000 samples per second during 10 seconds for normal bearing (NB) condition samples and failure condition samples; outer race fault (ORF), inner race fault (IRF), and ball fault (BF). Further details on the experimental setup can be found in [46].
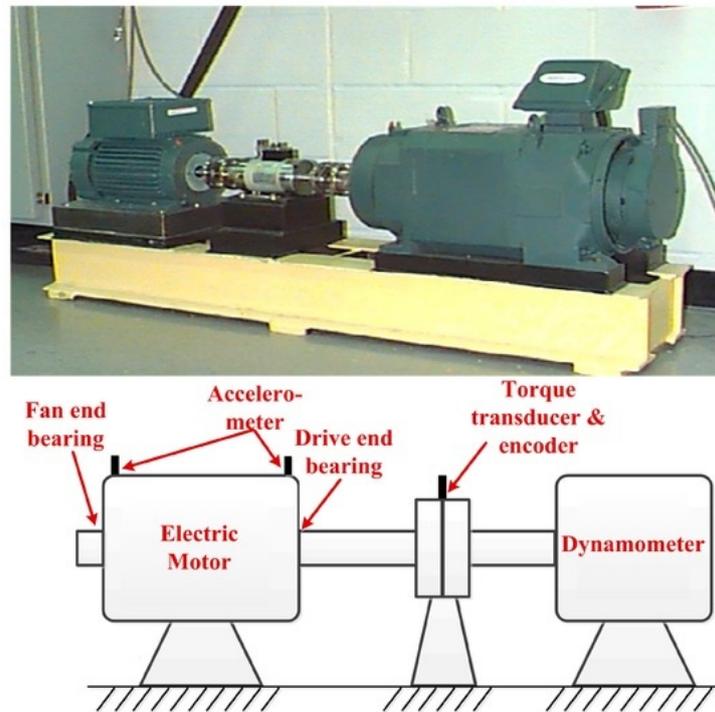

Figure 1. Experimental Setup [56].

4. Experimental Results

We performed experiments on the two datasets presented above. With these experiments, we aimed to evaluate the performance of the diagnostic methods proposed in this paper. Firstly, we applied the J-levels SWP transform to decompose the non-overlap signal into $D = 2^J$ sub-band signals. Secondly, the Shannon entropy value was computed using the corresponding Equations (7), (11) or (12) for each wavelet sub-band raw signal. Thirdly, KELM model was applied to diagnose the bearing fault types with different severities. Equations (13) and (15) were used to compute the output weights of the KELM classifier, and a 5-fold cross validation method for each value of J was used to adjust the regularization Parameter C. The bearing vibration signal database was split into five folds. In order to adjust the parameters J and C, four out of the five folds were used. The remaining fold was used for the testing stage. To evaluate the performance of the KELM model during the testing stage we considered the following measures of performance:

$$\text{Accuracy} = \frac{1}{M_2} \sum_{j=1}^{12} CM_{j,j}, \tag{19}$$

where the M_2 value corresponds to the number of testing samples for all classes combined, CM represents the confusion matrix and $CM_{j,j}$ corresponds to the number of samples in class y_j that are correctly classified as class y_j [57,58]. The second measure called F-scores was computed for every class label and it is calculated as follows:

$$\text{F-scores}(j) = 2 \times \frac{\text{Precision}(j) \times \text{Recall}(j)}{\text{Precision}(j) + \text{Recall}(j)} \quad j = 1, 2, \ldots, 10, 11, 12 \tag{20a}$$

$$\text{Precision}(j) = \frac{CM_{j,j}}{\sum_{i=1}^{12} CM_{j,i}} \tag{20b}$$

$$\text{Recall}(j) = \frac{CM_{j,j}}{\sum_{i=1}^{12} CM_{i,j}}, \tag{20c}$$

where the precision, recall and F-score measures of the *j*-th predicted class are represented by Precision(j), Recall(j), and F-scores(j), respectively [57,58].

4.1. Case 1: Drive-End Bearing

The collected dataset considered one normal bearing (NB) condition and 11 faulty bearing conditions that represented all possible combinations of the three possible failure locations (ORF, IRF and BF) over the four different fault severity levels (7, 14, 21 and 28 mils), giving a 12-class identification problem. For each class, there were four vibration signals corresponding to the rotatory shaft speeds of 1797 r/min, 1772 r/min, 1750 r/min and 1730 r/min with loads of 0, 1, 2 and 3 hp, respectively, leaving a total of 48 vibration signals. The length of these raw vibration signal was set to 120,000 data points (obtained in 10 s). Each of these 48 signals was divided into 60 segments. The size of each segment was set to 2000 data points (≈five times the rotation shaft period). Table 1 shows these values.

Table 1. Structure of bearing datasets.

Fault Types	Speed (r/min)	Load (hp)	Fault Diameter (mils)	Samples Numbers	Class Label [1]	Class Label [2]
NB	1797–1730	0–3	0	240	1	1
ORF	1797–1730	0–3	7	240	2	2
			14	240	3	3
			21	240	4	4
IRF	1797–1730	0–3	7	240	5	5
			14	240	6	6
			21	240	7	7
			28	240	8	–
BF	1797–1730	0–3	7	240	9	8
			14	240	10	9
			21	240	11	10
			28	240	12	–

[1] drive-end bearing; [2] fan-end bearing.

We used the 5-fold cross validation method to find the regularisation parameter *C* and the number of discriminative features. Figure 2 illustrates the average accuracy values and F-score values of the proposed SWPDE–KELM method during the testing stage considering *c* (number of states of the dispersion entropy) equal to 5. The effect on the diagnosis accuracy of other values of *c* can be seen in Table 2. As we can see from Figure 2a, an average accuracy of 100% is achieved considering eight features and the regularization parameter set to $C = 10^4$, whereas Figure 2b shows the F-score results achieved for twelve different types of faults. As can be seen, there were no misclassified testing samples and the F-score value was 100% for each of the twelve classes, which validated the effectiveness of the proposed SWPDE–KELM method.
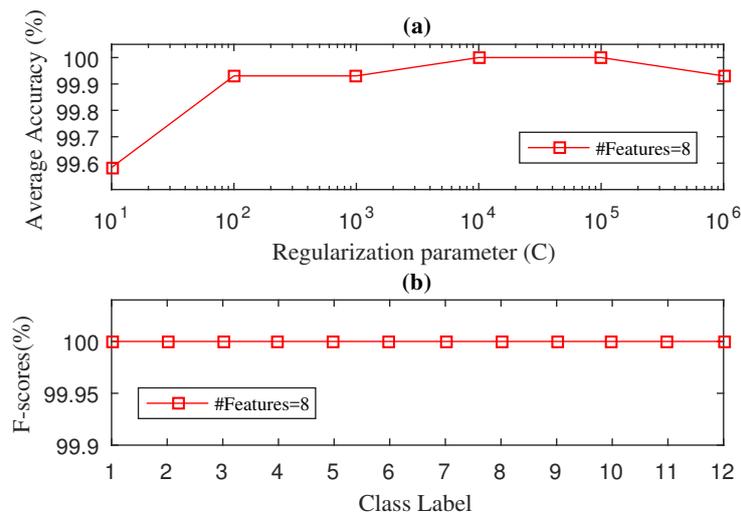
Figure 2. Diagnosis Accuracy (**a**) and F-score (**b**) values obtained by the SWPDE-KELM diagnosis with 5-Fold-CV during testing phase for drive end bearing.

Table 2. Entropy's parameter for drive-end bearing.

Method	Embedding (m)	Classes (c)	Avg. Accuracy
3-level SWPDE	2	5	100
		6	100
		7	100
		8	100
4-level SWPPE	4		99.97
	5	——	99.97
	6		100
	7		100

We then tried our SWPPE–KELM method on the drive-end bearing signals. We adjusted both the number of features and the C parameter using the same 5-fold cross validation procedure. Figure 3 shows the performance evaluation during the testing phase for the average accuracy values and the F-score values for the SWPPE–KELM method using the embedding dimension $m = 6$. The effect on the diagnosis accuracy of other values of m can be seen in Table 2. We can see in Figure 3a, that the SWPPE–KELM method reached an average accuracy of 100% with sixteen features (i.e., 4-level wavelet decomposition) and the regularization parameter set to $C = 10$, while the average accuracy of the SWPPE–KELM method constructed with eight features decreased, and the best average accuracy achieved with eight features was 99.97% for a regularization parameter set equal to $C = 10^4$. The F-score results for the SWPPE–KELM method with eight and sixteen features are shown in Figure 3b. As can be seen, there were no misclassified test samples when the SWPPE–KELM method was constructed with sixteen features. On the contrary, the method built with eight features achieved an F-score of 100% in only ten classes, since for class 4 (ORF-21 mils) and class 10 (BF-14 mils) it reached an F-score of 99.91% and 99.751%, respectively. Therefore, even though the SWPPE–KELM method achieved an average accuracy of 100%, it needed eight features more than the SWPDE–KELM method, making the SWPDE–KELM slightly more efficient than the SWPPE–KELM method.
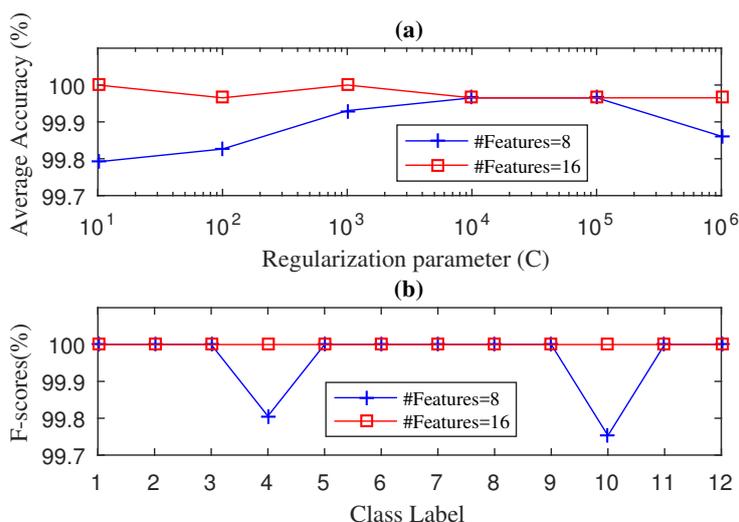
Figure 3. Diagnosis Accuracy (**a**) and F-score (**b**) values obtained by the SWPPE-KELM diagnosis with 5-Fold-CV during testing phase for drive end bearing.

The average accuracy of the SWPSVE–KELM method during testing phase with $2^J + 2$ features with $J = \{3, 4, 5\}$, where the two extra features corresponded to the Shannon entropy of the raw signal and the Shannon entropy of the singular values, are illustrated in Figure 4a. As we can see in Figure 4a, an average accuracy of 99.98% was achieved for 34 features and the parameter $C = 10^3$. On the contrary, the average accuracy for 10 and 18 features was generally lower than for 34 features. In Figure 4b we present the F-score results obtained during the testing phase. Overall, the best results were obtained with 34 features. It obtains an F-score of 100% for 10 classes, whereas for class 10 (BF-14 mils) and class 11 (BF-21 mils), the F-score achieved was of 99.87% and 99.89%, respectively. In addition, from Figure 4b it can be observed that the F-score results worsen with 10 and 18 features.
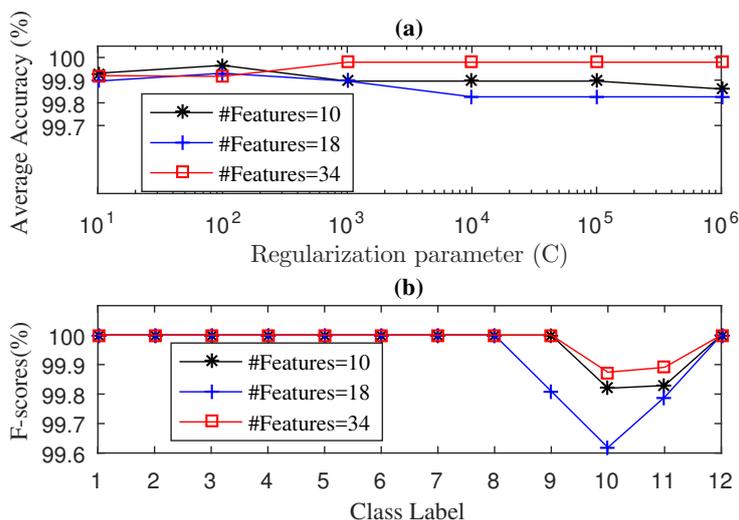

Figure 4. Diagnosis Accuracy (**a**) and F-score (**b**) values obtained by the SWPSVE-KELM diagnosis with 5-Fold-CV during testing phase for drive end bearing.

Therefore, from these results, diagnosis accuracy obtained by the SWPDE–KELM method was slightly better than the SWPPE–KELM method and they both significantly outperformed the SWPSVE–KELM method when applied to the drive-end dataset.

4.2. Case 2: Fan-End Bearing

For the fan-end bearing dataset we used in this paper, collected data consisted of nine faulty bearing conditions with three failure diameters (7, 14 and 21 mils) and a normal bearing condition, giving a 10-class recognition problem. For each class, there are 240 samples and a total of 2400 samples. We use the same 5-fold cross validation method to find parameters C and J.

Figure 5 shows the average accuracy and the F-score values obtained by the SWPDE–KELM method considering c (number of states of the dispertion entropy) equal to 5. The effect on the diagnosis accuracy of other values of c can be seen in Table 3. As we can see in Figure 5a, when 16 features are considered ($J = 4$), the method cannot reach 100% average accuracy. If we increase the number of features to 32 ($J = 5$), the 100% of average accuracy is only reached for $C = \{10^1, 10^2\}$ values. It was interesting to note that for both values of J, as the parameter C increased ($C > 10^2$), the average accuracy was heavily impaired. We then compute the F-score values using 16 and 32 features with $C = 10^2$ and $C = 10^1$, respectively. As we can see in Figure 5b, when we used 32 features the SWPDE–KELM method reached the 100% F-score value for the all 10 failure classes. However, when we used 16 features, the method reached the the 100% F-score value for only eight out of the 10 failure classes. It was also interesting to note that the two failure classes for which our method was not able to reach the 100% F-score value corresponded to ball faults (7 mils and 21 mils). This might mean that ball faults are more difficult to identify. We can also note that failures in this dataset (fan-end bearing) were harder to identify than the ones in the drive-end bearing dataset.
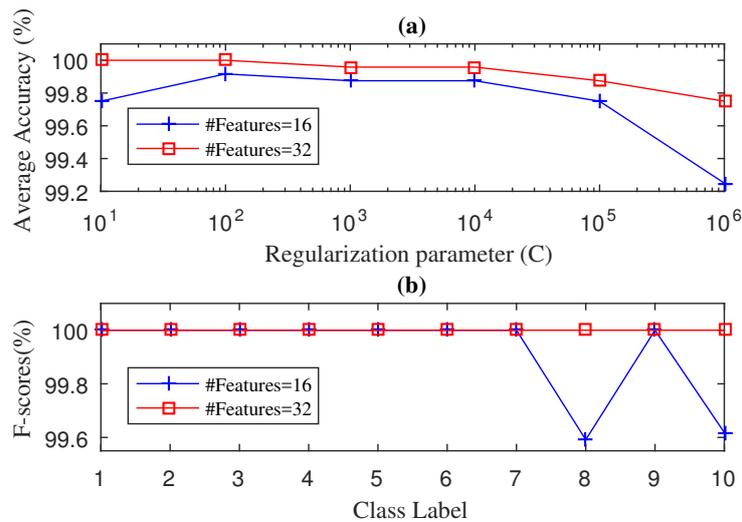

Figure 5. Diagnosis Accuracy (**a**) and F-score (**b**) values obtained by the SWPDE-KELM diagnosis with 5-Fold-CV during testing phase for fan-end bearing.

Table 3. Entropy's parameter for Fan-end Bearing.

Method	Embedding (m)	Clases (c)	Avg. Accuracy
3-level SWPDE	2	5	100
		6	100
		7	100
		8	100
4-level SWPPE	4		99.93
	5		99.97
	6	——	100
	7		100

Figure 6 shows the average accuracy and the F-score values obtained by the SWPPE–KELM method using the embedding dimension $m = 6$. The effect on the diagnosis accuracy of other values

of *m* can be seen in Table 3. Just as for the SWPDE–KELM method, when 16 features are considered ($J = 4$), the method cannot reach the 100% of average accuracy and, again, if we increase the number of features to 32 ($J = 5$), the 100% of average accuracy is reached for all values of *C*. We then computed the F-score values using 16 and 32 features with $C = 10^3$ and $C = 10^1$, respectively. As we can see in Figure 6b, when we used 32 features the SWPPE–KELM method reached the 100% F-score values for the all 10 failure classes. Again, when we used 16 features, the method reached the the 100% F-score values for only eight out of the 10 failure classes. Ball faults were, again, the only two failure classes for which our method was not able to reach the 100% F-score value. It is important to note that, although the number of times both the SWPDE–KELM and the SWPPE–KELM methods reach the 100% F-score value was the same, the SWPPE–KELM method was slightly better than the the SWPDE–KELM method for those fault classes for which the 100% F-score value was not reached.
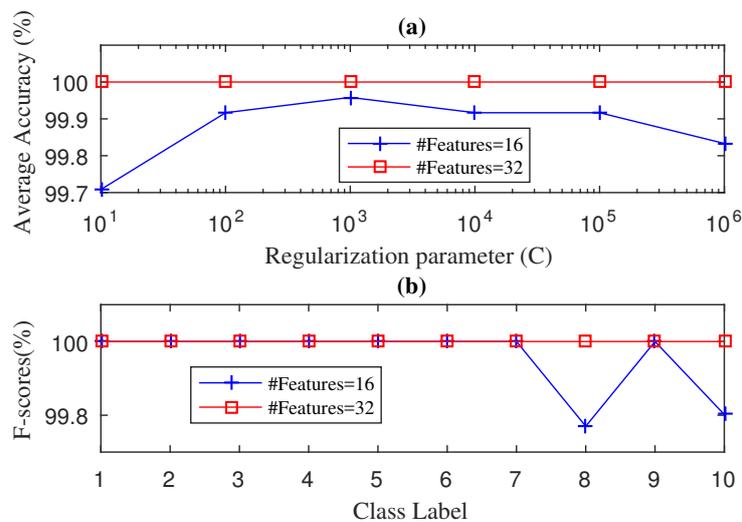

Figure 6. Diagnosis Accuracy (**a**) and F-score (**b**) values obtained by the SWPPE-KELM diagnosis with 5-Fold-CV during testing phase for fan-end bearing.

Finally, in Figure 7 we present the results obtained by the SWPSVE–KELM algorithm introduced in [40]. For this method we considered, again, $2^J + 2$ features with $J = \{3, 4, 5\}$. Figure 7a shows the average accuracy for different values of *C* and *J*. Unlike the methods proposed here, the SWPSVE–KELM method never reached the 100% of average accuracy. The best average accuracy value (99.88%) was obtained for $C = 10^3$ and 10 features. For 18 and 34 features, the best average accuracy value (99.83%) was obtained when $C = 10^4$. Once we set values of *C* for each value of *J* we computed the F-score. As we can see in Figure 7b, when we used 10 features, the SWPSVE–KELM method reached the 100% F-score value in eight out of the 10 failure classes. When we used 18 and 34 features, the 100% F-score value was only reached in seven out of the 10 failure classes. We need to point out that, just as for our methods, ball faults are much harder to identify, however, results obtained by our methods clearly outperformed the ones obtained by the SWPSVE–KELM method.
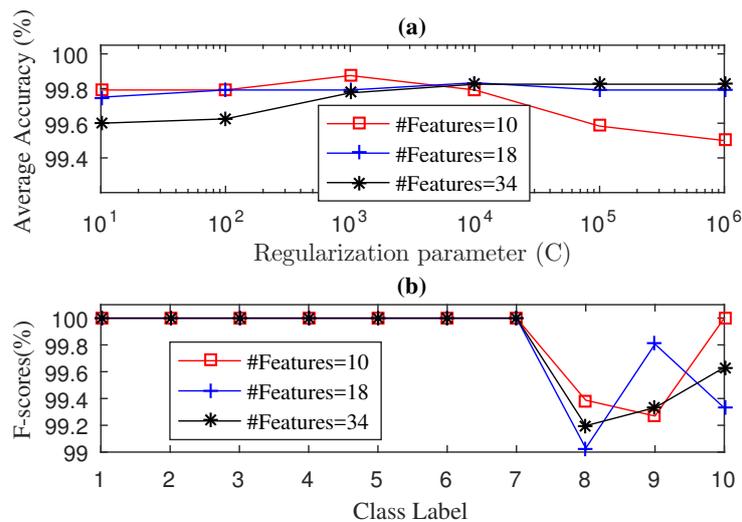
Figure 7. Diagnosis Accuracy (**a**) and F-score (**b**) values obtained by the SWPSVE-KELM diagnosis with 5-Fold-CV during testing phase for fan-end bearing.

5. Conclusions

This article presents two methods, called SWPDE and SWPPE, for feature extraction for bearing failure diagnosis. Our proposed methods combine SWP transform and Shannon entropy to improve the accuracy of the classifier. More specifically, the probability distribution function of the entropy is computed using either differential or permutation entropy.

Drive-end and a fan-end bearing datasets are considered in this study. We apply our algorithms on these datasets and compare our results to those obtained by a recently reported method called SWPSVE–KELM.

For the drive-end bearing dataset, we found that our SWPDE–KELM method reached a 100% accuracy level and 100% F-score value for the 12 bearing work conditions using 8 features. Although the SWPPE–KELM method also reached a 100% accuracy level and 100% F-score value, it needed 16 features to do so. These values are still very good when compared to the 34 features that the SWPSVE–KELM needs to reach similar values.

For the fan-end bearing dataset, our proposed methods reach the 100% F-score value for the all 10 failure classes. Since failures in the fan-end dataset are much harder to identify, our methods need 32 features to reach the 100% F-score value. Although the SWPSVE–KELM method only reached the 100% F-score value in eight out of the 10 failure classes, it only needed 10 features.

It is also interesting to note that, when considering only 16 features, our method was not able to reach the 100% F-score value for two classes corresponding to ball faults (7 mils and 21 mils). This might mean that ball faults are more difficult to identify. Furthermore, these failures in the fan-end bearing are harder to identify than the ones in the drive-end bearing dataset. However, although ball faults are much harder to identify, results obtained by our methods for these classes of failures are much better than the ones obtained by the SWPSVE–KELM method.

Although only experimental data has been considered in our experiments, it is important to note that, as has been shown in the literature previously, diagnosis algorithms tuned using experimental data can be succesfully used in more realistic environments for fan-end bearings [14].

Overall, our methods clearly outperform the SWPSVE–KELM method w.r.t. F-score and accuracy. Further, as shown in Table 4, our method is able to reach the same diagnosis accuracy levels as other previously proposed methods with a lower complexity of parameter tuning. As future work, we aim to combine Shannon entropy and different time–frequency analysis methods to rotationary machine failure diagnosis under variable work conditions. Also, we intend to apply our algorithms to more realistic datasets such as the ones from planetary bearings and gearboxes.

Table 4. Comparison between the proposed method and some previous work for bearing fault diagnosis.

Reference	Feature Extraction	Classification Method	Classes Number	Average Accuracy (%)
Brkovic et al. [14]	Wavelet energy entropy	Quadratic Classifier	4	100
Li et al. [59]	MPE from LMD	SVM with Binary Tree	4	100
Zheng et al. [60]	FE from LCD	ANFIS	7	100
Yan et al. [61]	IED-PE from IVMD	KNN	8	98.38
Rodriguez et al. [40]	Singular entropy from stationary wavelet	KELM	10	100
Mao et al. [62]	Fourier amplitude	Deep-ELM	10	100
Yan and Jia [63]	Multi-domain features with Laplace score	SVM with PSO	12	100
This work	DE and PE from stationary wavelet	KELM	12	100

References

1. Lei, Y.; Lin, J.; He, Z.; Zuo, M. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [CrossRef]

2. Smith, J.S. The local mean decomposition and its application to EEG perception data. *J. R. Soc. Interface* **2005**, *2*, 443–454. [CrossRef] [PubMed]

3. Tian, Y.; Jian, M.; Chen, L.; Wang, Z. Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mech. Mach. Theory* **2015**, *90*, 175–186. [CrossRef]

4. Chen, J.; Li, Z.; Chen, G.; Zi, Y.; Yuan, J.; Chen, B.; He, Z. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2016**, *70–71*, 1–35. [CrossRef]

5. Huang, N.E.; Zheng, S.; Long, S.; Wu, M.; Shih, H.; Zheng, Q.; Yen, N.; Tung, C.; Liu, H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]

6. Abbasion, S.; Rafsanjani, A.; Irani, A.F.A.; Rafsanjani, A. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mech. Syst. Signal Process.* **2007**, *21*, 2933–2945. [CrossRef]

7. Mishra, C.; Samantaray, A.; Chakraborty, G. Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising. *Measurement* **2017**, *103*, 77–86. [CrossRef]

8. Purushotham, V.; Narayanan, S.; Suryanarayana, S.; Prasad, A.N. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT E Int.* **2005**, *38*, 654–664. [CrossRef]

9. Su, W.; Wang, F.; Zhu, H.; Zhang, Z.; Guo, Z. Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. *Mech. Syst. Signal Process.* **2010**, *41*, 127–140. [CrossRef]

10. Villecco, F.; Pellegrino, A. Entropic Measure of Epistemic Uncertainties in Multibody System Models by Axiomatic Design. *Entropy* **2017**, *19*, 291. [CrossRef]

11. Villecco, F.; Pellegrino, A. Evaluation of Uncertainties in the Design Process of Complex Mechanical Systems. *Entropy* **2017**, *19*, 475. [CrossRef]

12. Zhang, X.; Zhou, J. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. *Mech. Syst. Signal Process.* **2013**, *41*, 127–140. [CrossRef]

13. Gligorijevic, J.; Gajic, D.; Brkovic, A.; Savic-Gajic, I.; Georgieva, O.; Di Gennaro, S. Online Condition Monitoring of Bearings to Support Total Productive Maintenance in the Packaging Materials Industry. *Sensors* **2016**, *16*, 316. [CrossRef]

14. Brkovic, A.; Gajic, D.; Gligorijevic, J.; Savic-Gajic, I.; Georgieva, O.; Gennaro, S.D. Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery. *Energy* **2017**, *136*, 63–71. [CrossRef]

15. 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]

16. Liu, H.; Han, M. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. *Mech. Mach. Theory* **2014**, *75*, 67–78. [CrossRef]

17. Gao, Q.; Liu, W.; Tang, B.; Li, G. A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine. *Renew. Energy* **2018**, *116*, 169–175. [CrossRef]

18. 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]

19. Yi, C.; Lv, Y.; Ge, M.; Xiao, H.; Yu, X. Tensor Singular Spectrum Decomposition Algorithm Based on Permutation Entropy for Rolling Bearing Fault Diagnosis. *Entropy* **2017**, *19*, 139. [CrossRef]

20. Gao, Y.; Villecco, F.; Li, M.; Song, W. Multi-Scale Permutation Entropy Based on Improved LMD and HMM for Rolling Bearing Diagnosis. *Entropy* **2017**, *19*, 176. [CrossRef]

21. Tian, Y.; Wang, Z.; Lu, C. Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping. *Mech. Syst. Signal Process.* **2019**, *114*, 658–673. [CrossRef]

22. Zhao, L.Y.; Wang, L.; Yan, R.Q. Rolling Bearing Fault Diagnosis Based on Wavelet Packet Decomposition and Multi-Scale Permutation Entropy. *Entropy* **2015**, *17*, 6447–6461. [CrossRef]

23. Yasir, M.N.; Koh, B.H. Data Decomposition Techniques with Multi-Scale Permutation Entropy Calculations for Bearing Fault Diagnosis. *Sensors* **2018**, *18*, 1278. [CrossRef] [PubMed]

24. Zheng, J.; Pan, H.; Yang, S.; Cheng, J. Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2018**, *99*, 229–243. [CrossRef]

25. 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]

26. Zheng, J.; Pan, H.; Cheng, J. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **2017**, *85*, 746–759. [CrossRef]

27. Rostaghi, M.; Ashory, M.R.; Azami, H. Application of dispersion entropy to status characterization of rotary machines. *J. Sound Vib.* **2019**, *438*, 291–308. [CrossRef]

28. Zhang, Y.; Tong, S.; Cong, F.; Xu, J. Research of Feature Extraction Method Based on Sparse Reconstruction and Multiscale Dispersion Entropy. *Appl. Sci.* **2018**, *8*, 888. [CrossRef]

29. Yan, X.; Jia, M. Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection. *Knowl.-Based Syst.* **2019**, *163*, 450–471. [CrossRef]

30. Lei, Y.; He, Z.; Zi, Y. EEMD method and WNN for fault diagnosis of locomotive roller bearings. *Expert Syst. Appl.* **2011**, *38*, 7334–7341. [CrossRef]

31. Yang, Y.; Dejie, Y.; Junsheng, C. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* **2006**, *294*, 269–277.

32. Luo, M.; Li, C.; Zhang, X.; Li, R.; An, X. Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings. *ISA Trans.* **2016**, *65*, 556–566. [CrossRef] [PubMed]

33. Chen, F.; Tang, B.; Chen, R. A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm. *Measurement* **2013**, *46*, 220–232. [CrossRef]

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

35. 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]

36. Huang, G.B. Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes. *IEEE Trans. Neural Netw.* **2006**, *17*, 879–892. [CrossRef] [PubMed]

37. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part B* **2012**, *42*, 513–529. [CrossRef]

38. Zhang, R.; Lan, Y.; Huang, G.; Xu, Z. Universal approximation of extreme learning machine with adaptive growth of hidden nodes. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 365–371. [CrossRef]

39. Rodriguez, N.; Lagos, C.; Cabrera, E.; Canete, L. Extreme learning machine based on stationary wavelet singular values for bearing failure diagnosis. *Stud. Inf. Control* **2017**, *26*, 287–294. [CrossRef]

40. Rodriguez, N.; Cabrera, G.; Lagos, C.; Cabrera, E. Stationary Wavelet Singular Entropy and Kernel Extreme Learning for Bearing Multi-Fault Diagnosis. *Entropy* **2017**, *19*, 541. [CrossRef]

41. Coifman, R.; Donoho, D. Translation-invariant de-noising. *Wavelets Stat. Lect. Notes Stat.* **1995**, *102*, 125–150.

42. Nason, G.; Silverman, B. The stationary wavelet transform and some statistical applications. *Wavelets Stat. Lect. Notes Stat.* **1995**, *103*, 281–300.

43. Pesquet, J.C.; Krim, H.; Carfantan, H. Time-invariant orthonormal wavelet representations. *IEEE Trans. Signal Process.* **1996**, *44*, 1964–1970. [CrossRef]

44. Daubechies, I. *Ten Lectures on Wavelet*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992.

45. Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press: San Diego, CA, USA, 1999.

46. Bearing Data Center. Case Western Reserve University. Technical Report. 2017. Available online: https://csegroups.case.edu/bearingdatacenter/home (accessed on 11 October 2017).

47. Lou, X.; Loparo, K.A. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal Process.* **2004**, *18*, 1077–1095. [CrossRef]

48. Rostaghi, M.; Azami, H. Dispersion Entropy: A Measure for Time-Series Analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [CrossRef]

49. Azami, H.; Rostaghi, M.; Abásolo, D.; Escudero, J. Refined Composite Multiscale Dispersion Entropy and Its Application to Biomedical Signals. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2872–2879.

50. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef]

51. Klema, V.C.; Laub, A.J. The singular value decomposition: Its computation and some applications. *IEEE Trans. Autom. Control* **1980**, *25*, 164–176. [CrossRef]

52. Huang, G.; Chen, L. Convex incremental extreme learning machine. *Neurocomputing* **2007**, *70*, 3056–3062. [CrossRef]

53. Serre, D. *Matrices: Theory and Applications*; Springer: New York, NY, USA, 2002.

54. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 111–142. [CrossRef]

55. Geisser, S. The predictive sample reuse method with applications. *J. Am. Stat. Assoc.* **1975**, *70*, 320–328. [CrossRef]

56. Sun, P.; Liao, Y.; Lin, J. The Shock Pulse Index and Its Application in the Fault Diagnosis of Rolling Element Bearings. *Sensors* **2017**, *17*, 535. [CrossRef] [PubMed]

57. Ferri, C.; Hernández-Orallo, J.; Modroiu, R. An experimental comparison of performance measures for classification. *Pattern Recognit. Lett.* **2009**, *30*, 27–38. [CrossRef]

58. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]

59. 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]

60. Zheng, J.; Cheng, J.; Yang, Y. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mech. Mach. Theory* **2013**, *70*, 441–453. [CrossRef]

61. Yan, X.; Jia, M.; Zhao, Z. A novel intelligent detection method for rolling bearing based on IVMD and instantaneous energy distribution-permutation entropy. *Measurement* **2018**, *130*, 435–447. [CrossRef]

62. Mao, W.; He, J.; Li, Y.; Yan, Y. Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2017**, *231*, 1560–1578. [CrossRef]

63. Yan, X.; Jia, M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing* **2018**, *313*, 47–64. [CrossRef]