



## Article

# Stationary Wavelet Singular Entropy and Kernel Extreme Learning for Bearing Multi-Fault Diagnosis

Nibaldo Rodriguez <sup>1,\*</sup> , Guillermo Cabrera <sup>1</sup> , Carolina Lagos <sup>1</sup> and Enrique Cabrera <sup>2</sup>

<sup>1</sup> Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso, Valparaíso 234 0000, Chile; guillermo.cabrera@pucv.cl (G.C.); carolina.lagos.c@mail.pucv.cl (C.L.)

<sup>2</sup> Centro de Investigación y Modelamiento de Fenómenos Aleatorios de Valparaíso (CIMFAV), Universidad de Valparaíso, Valparaíso 234 0000, Chile; enrique.cabrera@uv.cl

\* Correspondence: nibaldo.rodriguez@pucv.cl; Tel.: +56-322-273-761

Received: 7 September 2017; Accepted: 9 October 2017; Published: 13 October 2017

**Abstract:** The behavioural diagnostics of bearings play an essential role in the management of several rotation machine systems. However, current diagnostic methods do not deliver satisfactory results with respect to failures in variable speed rotational phenomena. In this paper, we consider the Shannon entropy as an important fault signature pattern. To compute the entropy, we propose combining stationary wavelet transform and singular value decomposition. The resulting feature extraction method, that we call stationary wavelet singular entropy (SWSE), aims to improve the accuracy of the diagnostics of bearing failure by finding a small number of high-quality fault signature patterns. The features extracted by the SWSE are then passed on to a kernel extreme learning machine (KELM) classifier. The proposed SWSE-KELM algorithm is evaluated using two bearing vibration signal databases obtained from Case Western Reserve University. We compare our SWSE feature extraction method to other well-known methods in the literature such as stationary wavelet packet singular entropy (SWPSE) and decimated wavelet packet singular entropy (DWPSE). The experimental results show that the SWSE-KELM consistently outperforms both the SWPSE-KELM and DWPSE-KELM methods. Further, our SWSE method requires fewer features than the other two evaluated methods, which makes our SWSE-KELM algorithm simpler and faster.

**Keywords:** stationary wavelet singular entropy; singular value decomposition; kernel extreme learning machine

## 1. Introduction

Early diagnosis of failures of rolling element bearings is very important for improving both the reliability and safety of the rotating machinery that is widely used in the industry. In order to achieve early diagnosis, we need to identify those hidden patterns that provide us with high-quality information regarding the bearing fault features. Unfortunately, extracting those features from non-stationary and non-linear vibration signals under time-varying speed conditions is not an easy task, and commonly used techniques for feature extraction are not accurate enough. Because of this, in the last decade several time-frequency analysis methods have been applied to the feature extraction problem for bearing fault diagnosis. Among them, we can find empirical mode decomposition (EMD) [1] and wavelet transform (WT) [2,3]. The EMD method can decompose a signal into a sum of intrinsic mode functions (IMFs) according to the oscillatory nature of the signal [4]. On the other hand, the WT decomposes a signal into several low-frequency components and high-frequency components, which can show features of hidden failures [5–8]. From signal decomposition methods, such as those above, different features can be calculated, such as energy entropy [9], permutation entropy [10], kurtosis value [11,12], relative energy [13],

envelope spectrum [14], and singular values [15]. These features are generally passed on to some classification methods such as support vector machines (SVMs) [9,10,14] or artificial neural networks (ANNs) [12,13]. In particular, the authors of [9] use IMF 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. The vector of singular values is obtained by means of singular value decomposition (SVD) of the IMF matrix. The authors in [10] propose a hybrid model for the bearing failure detection problem. This hybrid model uses permutation entropy (PE) to determine whether there is a failure or not. If a failure exists, then the PE of a subset of selected IMFs is computed and used as the input of an SVM in order to classify the type of the failure as well as its severity. A wavelet neural network (WNN) model combined with ensemble empirical mode decomposition (EEMD) for bearing fault diagnosis is proposed in [12]. Here, the more effective IMFs are selected based on the kurtosis value of each IMF. A subset of ten features from both the time domain and the frequency domain is used as input of the WNN for failure classification. In [16] authors combine WT and EMD to create a new time-frequency analysis method, namely empirical wavelet transform (EWT). A comparative study between EWT and EMD for bearing failure diagnosis using acoustic signals is presented in [11]. In that study, authors create an index based on the kurtosis value to select more effective IMFs. Results in [11] demonstrate that EWT performs better than EMD in terms of the accuracy of the diagnosis. Further, it is shown that EWT is able to efficiently find the frequency and the harmonic components corresponding to the bearing fault characteristic frequency.

Support vector machines and artificial neural networks are widely used to classify different kinds of failures in rotatory machines. However, one drawback of these techniques is that they are quite time consuming during the training stages. In [17–19] authors use a new method called the extreme learning machine (ELM) that aims to improve tuning time in a single-hidden layer feed-forward neural networks (SLFNs). Since then, ELM has been used in several studies mainly because of its efficiency. For instance, in [15] ELM is combined with local mode decomposition (LMD) and SVD for the diagnosis of bearing failure. Here, singular values obtained from the product function matrix are used as input of ELM. It has also been shown that the models combining LMD-SVD-ELM perform better than EMD-SVD-ELM models [15]. An ELM model combined with a real-valued gravitational search algorithm and the EEMD method for ball bearing fault diagnosis is proposed in [20]. Here, time-frequency features, energy features, and singular value features were calculated based on the EEMD method and it was shown to be effective in bearing fault diagnosis.

Besides vibration signals, other signals such as current signals [21–23], acoustic signals [24,25] and stray flux signals [26,27] have also been used for fault diagnosis. Several surveys on detection and diagnostic techniques considering vibration, current, acoustic, and stray flux signals can be found in [26,28–30].

In this article, we present a feature extraction method based on the Shannon entropy, which is computed by combining stationary wavelet transform (SWT) and SVD. Features extracted are passed on to a kernel extreme learning machine (KELM) model. We call our method stationary wavelet single entropy KELM (SWSE-KELM). The KELM model is created by replacing the ELM hidden activation function with a kernel function to improve the generalisation capacities of ELM and reduce time consumption for determining the number of hidden layer nodes [17–19]. While SWT is able to provide local features in both the time domain and frequency domain as well as it is able to distinguish sudden changes in the vibration signal, singular values are very stable, which leads to a more robust and reliable method [31,32]. Further, this extraction method requires the input of fewer features than other well-known extraction methods in the literature, with the same level of accuracy. We apply our method to two bearing vibration signal data sets under variable speed operation conditions obtained from Case Western Reserve University [33,34], and we evaluate our experimental results considering ten different bearing fault types. We compare the diagnosis accuracy obtained by our method to those obtained using stationary wavelet packet singular entropy (SWPSE) and decimated wavelet packet singular entropy (DWPSE) [35–38].

The remaining sections of this article are as follows. In Section 2 wavelet analysis is described. Section 3 describes the bearing multi-fault diagnosis algorithm and the experimental setup that is used in this paper. A discussion on the results obtained by the SWSE-KELM, SWPSE-KELM, and DWPSE-KELM methods is presented in Section 4. Finally, in Section 5 some conclusions are drawn.

## 2. Wavelet Analysis Techniques

### 2.1. Stationary Wavelet Transform

Stationary wavelet transform (SWT) [39–41] is a wavelet analysis method. It can be seen as an alternative to discrete wavelet transform (DWT) [42,43]. SWT and DWT share some similarities; the most important being that both high-pass and low-pass filters are applied at each level of the input signal. At the first level of SWT, an input signal  $\{x(n) = w_{0,0}(n), n = 1, \dots, N\}$  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:

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

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

Since no sub-sampling is performed, the obtained sub-bands  $w_{1,1}(n)$  and  $w_{1,2}(n)$  have the same number of elements as the input signal  $w_{0,0}(n)$ . The general process of the SWT recursively continues for  $j = 2, \dots, J$  and is given as follows:

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

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

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 SWT strategy is then completely 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, mainly due to its low complexity [34,42], whereas the decomposition at the  $J$ -th level is given as follows:

$$J \leq \log_2 \left( \frac{0.5f_s}{FCF} \right) \quad (3)$$

where  $f_s$  represents the frequency sampling of the vibration signal and FCF denotes the bearing fault characteristic frequency [44].

### 2.2. Stationary Wavelet Packet Transform

The stationary wavelet packet transform (SWPT) is similar to the SWT in that the high-pass and low-pass filters are applied to the input signal at each level. At the first level of SWPT, an input signal  $w_{0,0}(n)$  is convolved with a low-pass filter  $h_1$  and with a high-pass filter  $g_1$  to obtain both the

approximation coefficient  $w_{1,1}$  and the detail coefficient  $w_{1,2}$ , respectively, which are calculated using Equation (1a,b). The general process of the SWPT is continued recursively for  $j = 2, \dots, J$  as follows:

$$w_{j,2i-1}(n) = \sum_{k=0}^{r-1} h_j(k) w_{j-1,i}(n-k) \quad (4a)$$

$$w_{j,2i}(n) = \sum_{k=0}^{r-1} g_j(k) w_{j-1,i}(n-k) \quad (4b)$$

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  $i = 1, \dots, 2^{j-1}$ .

### 2.3. Decimated Wavelet Packet Transform

While very similar to the SWPT, the DWPT is different in that it includes a downsampling operator by a factor of two. Just as in the SWPT, the DWPT process computes recursively the  $j$ -th level for  $j = 2, \dots, J$ , as follows:

$$w_{j,2i-1}(n) = \sum_{k=0}^{r-1} h_j(k) w_{j-1,i}(2n-k) \quad (5a)$$

$$w_{j,2i}(n) = \sum_{k=0}^{r-1} g_j(k) w_{j-1,i}(2n-k) \quad (5b)$$

where again, 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  $i = 1, \dots, 2^{j-1}$ . Note that, unlike the SWPT, the length of each decomposition coefficient in DWPT,  $w_{j,i}$ , is equal to  $N_j = \frac{N}{2^j}$ .

## 3. Bearing Multi-Fault Diagnosis Algorithm

The failure diagnosis algorithm proposed in this study is based on both the feature extraction phase and the classification phase. On the one hand, the feature extraction phase is carried out by integrating wavelet analysis and singular value decomposition. On the other hand, the multi-fault classification phase is constructed using a KELM model based on the Gaussian kernel function. These phases are described in the following sections.

### 3.1. Feature-Extraction Algorithm

Below we present the algorithm for the wavelet singular entropy (WSE) based on both wavelet analysis and the SVD method.

1. Calculate the envelope signal from the raw vibration signal using the Hilbert transform as follows:

$$x(n) = \sqrt{x(n) + HT^2[x(n)]} \quad n = 1, \dots, L \quad (6)$$

where  $HT[\cdot]$  denotes the Hilbert transform, and  $L$  represents the length of the vibration signal.

2. Divide the envelope signal into non-overlapping sub-signals of  $N$  data points.
3. Decompose the envelope sub-signal  $x(n)$ ,  $n = 1, \dots, N$  into  $J$  levels by using wavelet analysis (SWT, SWPT, and DWPT).
4. Decompose the wavelet coefficients matrix  $W$  using the SVD method. The SVD method decomposes the wavelet matrix  $W$  into a series of mutually orthogonal, unit-rank, and elementary matrices, whose representation is given as follows [45]:

$$W = \sum_{k=1}^K s_k u_k v_k^T = USV^T \quad (7)$$

where

$$K = \begin{cases} J+1 & \text{for SWT} \\ 2^J & \text{for SWPT and DWPT} \end{cases} \quad (8)$$

where  $U \in R^{K \times K}$ ,  $V \in R^{N \times N}$ ,  $S$  is the  $K \times N$  diagonal matrix, and  $s_k$  represents the  $k$ -th singular value of matrix  $W$ .

5. Create the  $D$ -dimensional feature vector as follows:

$$z_i = [s_1, \dots, s_K, En_1, En_2] \quad i = 1, 2, \dots, L/N \quad \text{and} \quad D = K + 2 \quad (9)$$

where  $En_1$  and  $En_2$  represent the wavelet singular entropy value, and it is computed as

$$En_1 = -\frac{1}{\log_2(K)} \sum_{k=1}^K p_k \log_2(p_k) \quad (10a)$$

$$p_k = \frac{z_{i,k}^2}{\sum_{k=1}^D z_{i,k}^2} \quad (10b)$$

$$En_2 = -\frac{1}{\log_2(N)} \sum_{n=1}^N q_n \log_2(q_n) \quad (10c)$$

$$q_n = \frac{x^2(n)}{\sum_{n=1}^N x^2(n)} \quad (10d)$$

6. Normalise the features matrix  $Z$  as follows:

$$Z_j = \frac{Z_j}{\max(Z_j)} \quad j = 1, 2, \dots, D \quad (11)$$

where  $Z_j$  represents the  $j$ -th column of the features matrix  $Z$ .

7. Randomly select 80% of the features matrix  $Z$  for training data and the remaining 20% for testing data.

### 3.2. Kernel-ELM Classifier

In this section a brief description of ELM and KELM is presented. For more details on this topic see [17–19,46]. The output of the ELM algorithm is obtained as follows:

$$\hat{y}_j(z) = \sum_{i=1}^{N_h} \beta_{j,i} \phi_i[z_k, a_{i,k}, b_i] \quad j = 1, 2, \dots, 10 \quad (12)$$

where  $N_h$  denotes the number of hidden nodes,  $z_k \in R^D$  represents the input vector containing  $D$ -features,  $a_{i,k}$  are the weights of the hidden layer,  $b_i$  are bias units,  $\beta_{j,i}$  are output weights of the output layer, and  $\phi(\cdot)$  represents the hidden nodes activation functions. The Moore–Penrose generalised inverse method (M-P) [47] is used to estimate the output weights  $\beta_{j,i}$ , whereas the weights  $a_{i,k}$  and  $b_i$  are randomly assigned. The optimal values of the linear weights of the output layer, for any given representation of the hidden weights, are obtained as follows:

$$\beta = H^T \left( \frac{I}{C} + HH^T \right)^{\dagger} Y \quad (13)$$

where  $(\cdot)^T$  denotes transposed operator,  $H$  is the hidden layer output matrix,  $Y$  represents the desired output pattern matrix,  $I$  is the identity matrix, and  $C$  denotes a regularisation parameter. The expression  $(\cdot)^\dagger$  represents the Moore–Penrose generalised inverse matrix [47].

In the ELM algorithm, if the mapping function  $\phi(\cdot)$  is unknown, then as proposed by Huang [18], we can use the Mercer's conditions on ELM to calculate a kernel matrix, which is given as follows:

$$\Omega = \Phi\Phi^T : \Omega_{i,j} = \phi(z_i)\phi(z_j) = \ker(z_i, z_j) \quad i, j = 1, \dots, M \quad (14)$$

where  $M$  represents the sample number and  $\ker(\cdot)$  denotes a Kernel function. In this paper we will use a Gaussian kernel function to construct the KELM model because of its superior performance [18]. This Gaussian kernel function is given as

$$\ker(z_i, z_j) = \exp\left(-\frac{\|z_i - z_j\|^2}{2\sigma^2}\right) \quad (15)$$

where  $\sigma$  represents the value of the Kernel width parameter, and here the  $\sigma$  parameter is selected as follows:

$$\sigma^2 = D^{-1} \quad (16)$$

where the  $D$  value represents the dimension of the input features vector to the KELM model (see Equation (9)).

Finally, by placing Equations (13) and (15) in Equation (12), we can obtain the output values of KELM classifier as follows:

$$\hat{y}(z) = \begin{pmatrix} \ker(z, z_1) \\ \ker(z, z_2) \\ \vdots \\ \ker(z, z_N) \end{pmatrix} \left( \frac{I}{C} + Ker \right)^\dagger Y \quad (17)$$

where  $\hat{y}_j$  denoted the output value of the  $j$ -th output node and the predicted class label of sample  $z$  is obtained as:

$$Label(\hat{y}(z)) = \max\{\hat{y}_1(z), \dots, \hat{y}_{10}(z)\} \quad (18)$$

To obtain a good generalisation performance in the KELM model, the regularisation parameter  $C$  is selected from  $\{10^1, 10^2, \dots, 10^{10}\}$  by using the 5-fold cross-validation (CV) method [48,49].

#### 4. Experimental Setup

The experimental raw data used in this paper corresponds to 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. Both data sets can be obtained from the CWRU bearing data centre [33]. This data set were generated using an experimental setup that considered a 2 hp Reliance Electric motor, a torque transducer/encoder, and a dynamometer. During the experiments, the bearing holds the motor shaft. An accelerometer mounted on the motor housing (as shown in Figure 1), is used to collect vibration signals. Single point failures with different failure diameters of 0.007, 0.014, and 0.021 inches are introduced to both the driving-end and the fan-end bearings using the electro-discharge machining method, with the motor speed varied at 1730, 1750, 1772, and 1797 r/min with loads of 3, 2, 1, and 0 hp, respectively. Digital data is produced at 12,000 samples per second for normal bearing (NB) samples and failure samples: inner race fault (IRF), outer race fault (ORF), and ball fault (BF). Further details on the experimental setup can be found in [33].



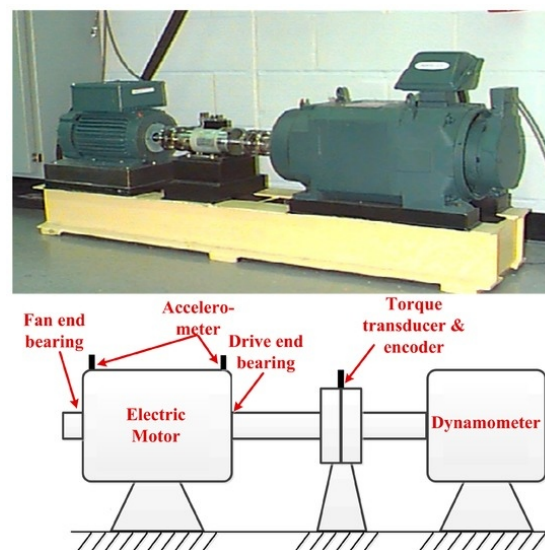


Figure 1. Experimental setup [50].

## 5. Discussion Results

For both data sets (drive-end and fan-end bearing), we consider one normal bearing condition and nine faulty bearing conditions that correspond to all possible combinations of the three failure locations over the three different fault severity levels, giving 10 class labels. For each class, we have four vibration signals corresponding to the rotatory shaft speeds of 1797, 1772, 1750, and 1730 r/min with loads of 0, 1, 2 and 3 hp, respectively, leaving a total of 40 vibration signals. The lengths of these raw vibration signals are set to 120,000 data points (obtained in 10 s). Each of these 40 signals is divided into 150 segments. The size of each segment is set to 800 data points ( $\approx$ two times the rotation shaft period), i.e., we use the all 120,000 data points for our experiments.

As explained before, we set parameter  $J$  of our feature extraction method to values in  $\{1, 2\}$  for the drive-end bearing data set and values in  $\{2, 3, 4\}$  for the fan-end bearing data set. Parameter  $C$  of our KELM classifier is then adjusted using a 5-fold cross validation method for each value of  $J$ . To this end, we use 150 segments for each of the 40 signals considered in our data set, that is, 6000 samples in total for the fault diagnosis. Then, 80% of these 6000 samples (i.e., 4800 samples) are used in the training process, while the remaining 20% (1200 samples) are used during the testing process (see Feature Extraction Algorithm in Section 3.1). Table 1 shows these values.

**Table 1.** Structure of both data sets considered in this paper. NB: normal bearing; IRF: inner race fault; ORF: outer race fault; BF: ball fault.

Fault Types	Speed (r/min)	Load (hp)	Fault Diameter (in)	Training/Test Samples	Class Label
NB	1797–1730	0–3	0	480/120	1
IRF	1797–1730	0–3	0.007	480/120	2
			0.014	480/120	3
			0.021	480/120	4
ORF	1797–1730	0–3	0.007	480/120	5
			0.014	480/120	6
			0.021	480/120	7
BF	1797–1730	0–3	0.007	480/120	8
			0.014	480/120	9
			0.021	480/120	10

The training database, which consists of 4800 samples, is divided into five folds (960 samples each). Four out of the five folds are used to adjust the parameters  $J$  and  $C$  of our fault diagnosis model. The remaining fold is used during the validation stage. During the training process, the output weights of the output layer are obtained as explained in Equations (13) and (15). To evaluate the performance of both the training and the testing process, we use two performance measures given as follows:

$$\text{Accuracy} = \frac{1}{M} \sum_{j=1}^{10} CM_{j,j} \quad (19)$$

where the  $M$  value is the total number of samples for all classes combined,  $CM$  denotes the confusion matrix, and  $CM_{j,j}$  represents the number of samples in class  $y_j$  that are correctly classified as class  $y_j$  [51,52]. The second measures are for the F-scores and they are obtained for every class label as follows:

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

$$\text{Precision}(j) = \frac{CM_{j,j}}{\sum_{i=1}^{10} CM_{j,i}} \quad (20b)$$

$$\text{Recall}(j) = \frac{CM_{j,j}}{\sum_{i=1}^{10} CM_{i,j}} \quad (20c)$$

where  $\text{Precision}(j)$ ,  $\text{Recall}(j)$  and  $\text{F-scores}(j)$  represent the precision, recall, and F-scores measures of the  $j$ -th predicted class; respectively [51,52].

We try the proposed bearing multi-fault diagnosis method on both the drive-end and the fan-end bearing data set to validate the efficiency of our approach. We compare the SWSE-KELM method proposed in this paper to two fault diagnosis methods that combine a wavelet packet singular entropy and a KELM classifier, namely stationary wavelet packet singular entropy (SWPSE) and decimated wavelet packet singular entropy (DWPSE). On the one hand, we have the DWPSE, a widely used technique that has been shown to be very effective in the context of feature extraction [35,37,38]. On the other hand, we have the SWPSE method, which is an extension of the SWSE method we propose in this paper.

### 5.1. Data Set 1: Drive-End Bearing

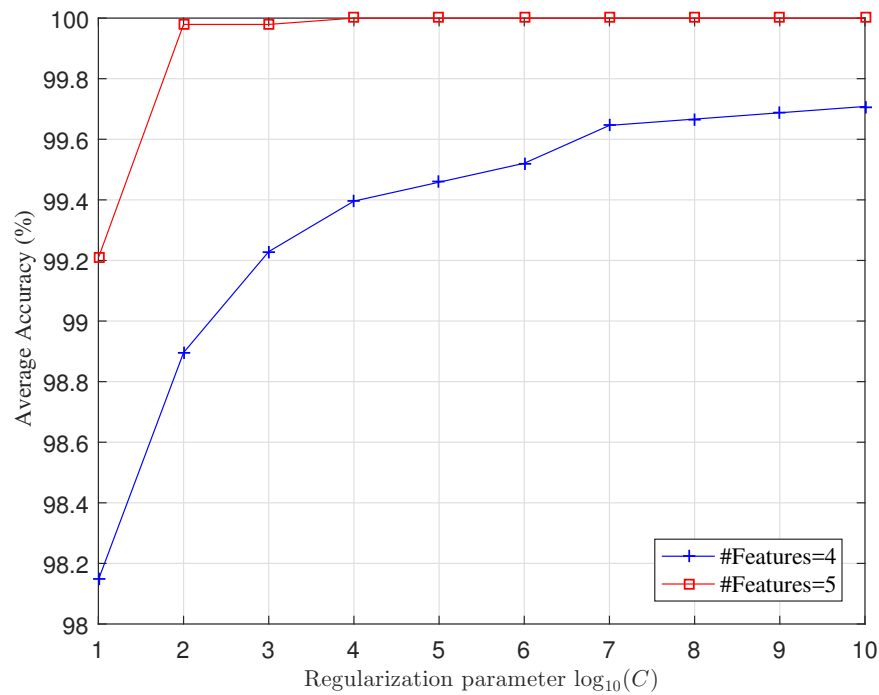
We first use the training data to find the best possible parameters for the fault diagnosis. In order to calculate the diagnostic accuracy of the proposed method, first, the feature extraction algorithm is carried out to obtain fault signatures based on stationary wavelet singular entropy, which are later used as input patterns in the KELM-classifier.

To evaluate the effect of the number of features and the regularisation parameter  $C$  on the diagnosis accuracy level, we use a 5-fold cross validation method. While results of the 5-fold cross validation method during the validation phase for the SWSE method are shown in Figure 2, results for the SWPSE and the DWPSE are shown in Figure 3a,b, respectively.

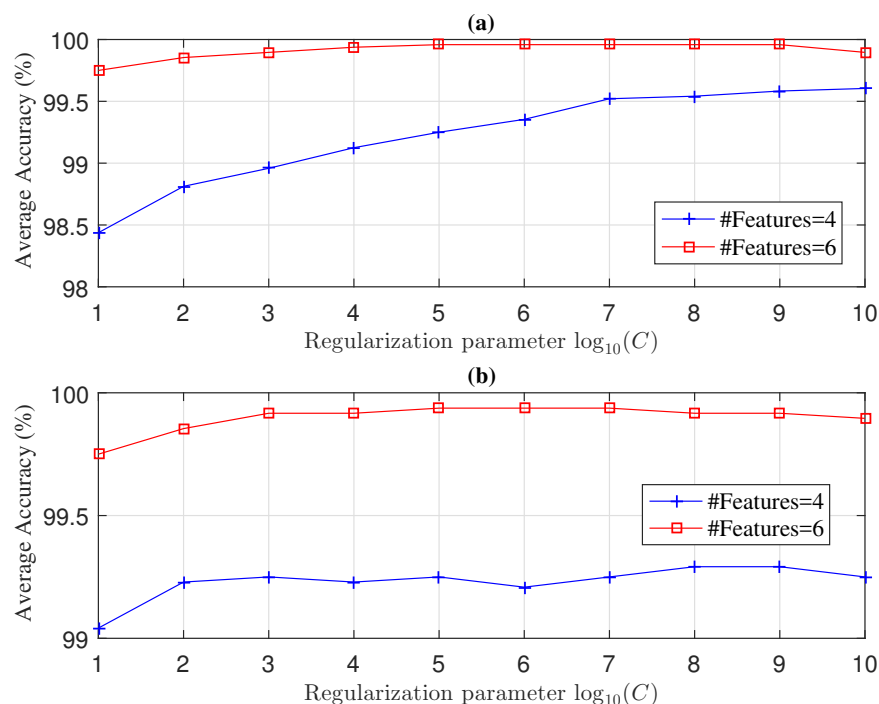
As we can see in Figure 2, for four features, the best value for the average accuracy level (99.7%) is obtained when the parameter  $C$  is equal to  $10^{10}$ . When five features are considered, a 100% average accuracy level is reached for values of  $C = \{10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}\}$ . Thus, the model we shall use in the testing phase considers five features and the regularisation parameter  $C = 10^4$ .

We then adjust the parameter  $C$  and the number of features for both the SWPSE and DWPSE methods. We do this by using the same 5-fold cross validation procedure. As we can see in Figure 3a, the best average accuracy level for the SWPSE is reached for six features (i.e.,  $J = 2$ ) and the regularisation parameter  $C = 10^5$ . The same values are obtained for the DWPSE method (see Figure 3b).





**Figure 2.** Stationary wavelet single entropy kernel extreme learning machine (SWSE-KELM) selection results with 5-fold cross validation (CV) during the validation phase for drive end bearing.

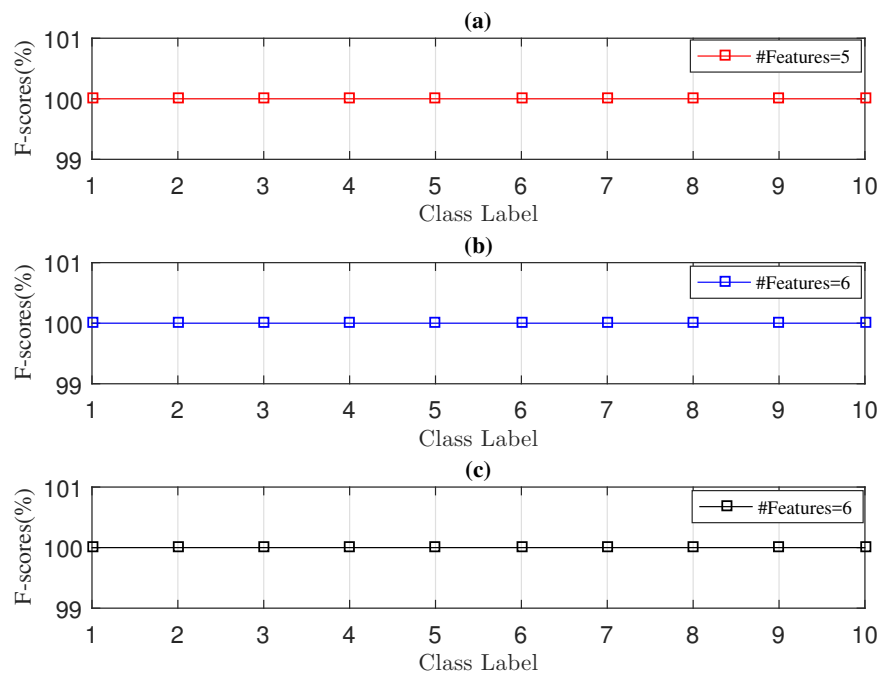


**Figure 3.** Model selection results with 5-fold CV during validation phase for drive end bearing: (a) Stationary wavelet packet singular entropy (SWPSE)-KELM; (b) Decimated wavelet packet singular entropy (DWPSE)-KELM.

Once the number of features and the regularisation parameter  $C$  have been chosen, the output weights matrix for the KELM classifier is selected from the best fold computed during the validation phase. We do this for the all three methods considered in our experiments. We then compare our SWSE method to the SWPSE and DWPSE methods. As mentioned before, all three methods are applied to

the data set of the testing phase and the same KELM classifier is used. To the best of our knowledge, the SWSE method has not been applied to the bearing fault diagnosis problem.

Figure 4 shows the F-score values obtained during the testing phase. As we can see, all three methods are able to reach, for each class, a 100% F-score value. Further, the accuracy level for all methods is also 100%. Although both the SWPSE-KELM and the DWPSE-KELM methods are also able to reach 100% F-Score value for the all 10 classes, they need one more feature than the SWSE-KELM method, which increases the method complexity.



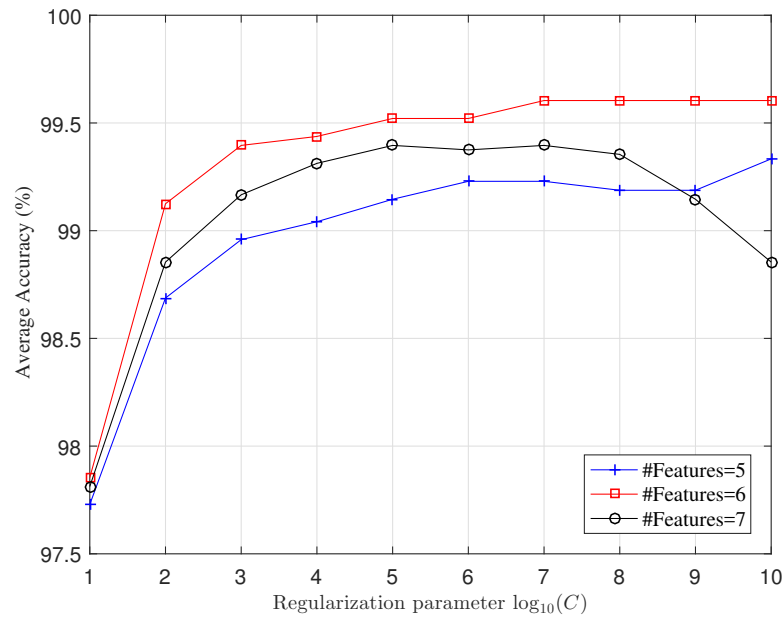
**Figure 4.** Bearing fault diagnosis results during testing phase for the drive-end bearing: (a) SWSE-KELM with five features; (b) SWPSE-KELM with six features; (c) DWPSE-KELM with six features.

## 5.2. Data Set 2: Fan-End Bearing

For the second data set that we use in this paper (fan-end bearing), we first use the training data to find the best possible parameters for the fault diagnosis. Just as we did for the drive-end bearing data set, we calculate the diagnostic accuracy of the proposed method by computing fault signatures using the SWSE feature extraction algorithm. The obtained signatures will be then used as input patterns in the KELM-classifier.

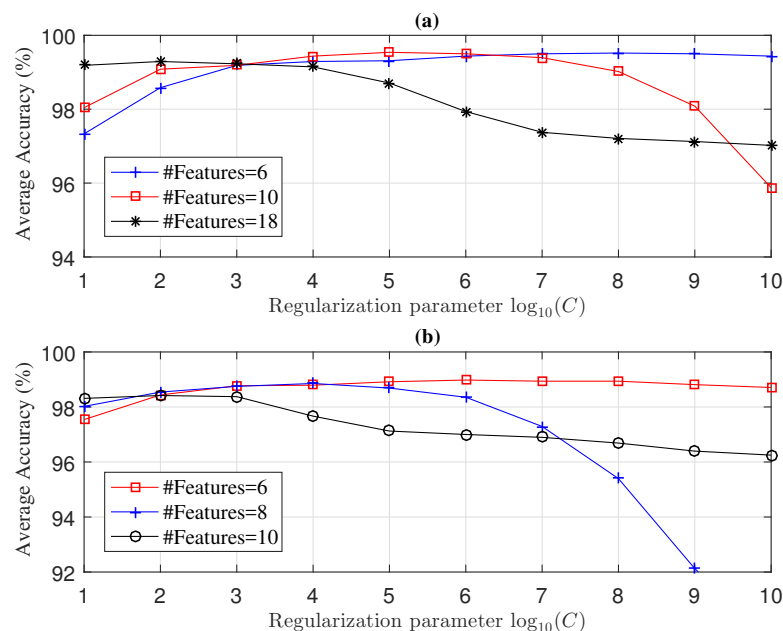
As in the first data set, we need to evaluate the effect of the number of features and the regularisation parameter  $C$  on the diagnosis accuracy level. To this end, we use again a 5-fold cross validation method.

Figure 5 shows the average accuracy with 5, 6, and 7 features, obtained after the training process. In Figure 5, can see that the SWSE-KELM method reaches its highest value when considering  $C = 10^7$  and six features. Moreover, results obtained for five and seven features are consistently below the ones obtained when using six features. Thus, the model we shall use during the testing phase considers  $C = 10^7$  and six features. It is interesting to note that the SWSE-KELM method needs, for the fan-end bearing data set, one more feature than for the drive-end bearing data set. Thus, we can say that the fan-end bearing is more complex than the drive-end bearing data set. This might be caused because of the location of the bearing in the motor, which makes the fan-end bearing data set harder to work with.



**Figure 5.** SWSE-KELM selection results with 5-fold CV during the validation phase for fan-end bearing.

We then need to evaluate the effect of the regularisation parameter  $C$  on the diagnosis accuracy level for the SWPSE-KELM and the DWPSE-KELM methods. As before, we implement the 5-fold cross validation method again with 6, 10, and 18 features. As we can see in Figure 6a, the best average accuracy level for the SWPSE-KELM is reached for 10 features (i.e.,  $J = 3$ ) and regularisation parameter  $C = 10^5$ . The best average accuracy level for the DWPSE-KELM is reached for six features (i.e.,  $J = 2$ ) and regularisation parameter  $C = 10^6$  (see Figure 6b). We need to note at this point that, as the number of features gets larger, both the kernel width ( $\sigma^2$  in Equation (16)) and the value of  $C$  become smaller.



**Figure 6.** Model selection results with 5-fold CV during the validation phase for fan-end bearing: (a) SWPSE-KELM; (b) DWPSE-KELM.

We then compare our SWSE-KELM method to the SWPSE-KELM and DWPSE-KELM methods. Results of this comparison are illustrated in Figure 7. As we can see, for the first seven classes,

all three methods reach a 100% F-score value. Unlike this, for the ball faults (classes 8, 9, and 10 according to Table 1), our method outperforms both the SWPSE-KELM and the DWPSE-KELM methods for the severity levels of 0.007 inches and 0.021 inches. For the severity level of 0.014 inches, both our method and the SWPSE-KELM method reach a 100% F-Score value, although our method needs only six features. The DWPSE-KELM only reaches a 98.76% F-score value for the severity level of 0.014 inches. Finally, the best accuracy reached by our model is 99.83%, while the best accuracy levels for the SWPSE-KELM and DWPSE-KELM methods are 99.75% and 98.75%, respectively.

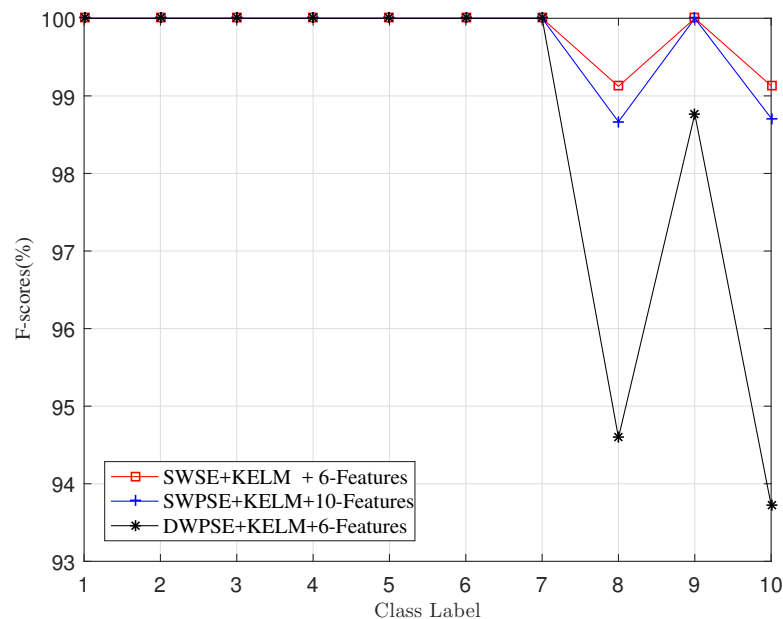


Figure 7. Bearing fault diagnosis results during the testing phase for the fan-end bearing.

## 6. Conclusions

This article presents a method of feature extraction for bearing failure diagnosis. Our proposed method uses the Shannon entropy, which is computed by combining SWT and SVD, to improve the accuracy of the classifier. Two data sets, namely drive-end and fan-end bearing, are used to validate our proposal and the obtained results are compared to those obtained by other two well-known feature extraction methods previously proposed in the literature, namely SWPSE and DWPSE. For the first data set (drive-end bearing), we found that our SWSE-KELM method reaches a 100% accuracy level and 100% F-score value for the 10 bearing operation conditions using only five features. Although both the SWPSE and the DWPSE also reach 100% accuracy level and 100% F-score value, they need one more feature to do so, and, thus, our method result is simpler and faster. For the second data set (fan-end), all methods reach a 100% F-score value for the first seven bearing operation conditions. However, for the last three bearing operation conditions (ball failure) our method reaches better F-score values than the other two methods for severity levels of 0.007 inches and 0.021 inches. As our SWSE-KELM method, the SWPSE-KELM method also reaches a 100% F-Score value for the severity level of 0.014 inches. However, the SWPSE-KELM needs 10 features to do so. Further, our method exhibits the best accuracy level (99.83%) when compared to both the SWPSE-KELM and the DWPSE-KELM methods.

Based on these results, we can state that the stationary wavelet transform allows us to extract, in an effective way, single-point bearing fault signatures using fewer features and improving significantly the accuracy of the diagnosis.

As future work, we aim to apply our feature extraction method to bearing failure diagnosis considering a run-to-failure data set. Although this problem is more representative of the damage propagation during the lifetime of the bearing under real operation conditions, it has been shown

that this kind of signal is much harder to classify due to the (highly) nonlinear damage propagation. An application of this kind of signals can be found in the transmission systems of mining machinery. In the near future, we expect to apply our algorithms to run-to-failure data sets obtained from a mining company in Chile.

**Acknowledgments:** Authors wish to thank the Bearing Data Centre at Case Western Reserve University for providing vibration signals data set and some of the images used in this paper.

**Author Contributions:** Nibaldo Rodríguez conceived the idea and designed the experiments. Guillermo Cabrera and Nibaldo Rodríguez wrote the paper. Nibaldo Rodríguez and Carolina Lagos implemented the algorithms. Nibaldo Rodríguez, Guillermo Cabrera, and Enrique Cabrera analyzed the data and the obtained results. All the authors have read and approved the final manuscript.

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

## References

1. Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126.
2. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15.
3. Chen, J.; Li, Z.; Pan, J.; 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.
4. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995.
5. Mishra, C.; Samantaray, A.K.; Chakraborty, G. Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising. *Measurement* **2017**, *103*, 77–86.
6. Purushotham, V.; Narayanan, S.; Prasad, S.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.
7. Abbasion, S.; Rafsanjani, A.; Farshidianfar, A.; Irani, N. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mech. Syst. Signal Process.* **2007**, *21*, 2933–2945.
8. 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.
9. 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.
10. 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.
11. Kedadouch, M.; Thomas, M.; Tahan, A. A comparative study between empirical wavelet transforms and empirical mode decomposition methods: Application to bearing defect diagnosis. *Mech. Syst. Signal Process.* **2016**, *81*, 88–107.
12. Lei, Y.; He, Z.; Zi, Y. EEMD method and WNN for fault diagnosis of locomotive roller bearings. *Expert Syst. Appl.* **2011**, *38*, 7334–7341.
13. Yang, Y.; Yu, D.; Cheng, J. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* **2006**, *294*, 269–277.
14. 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.
15. Tian, Y.; Ma, J.; Lu, C.; Wang, Z. Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mech. Machine Theory* **2015**, *90*, 175–186.
16. Gilles, J. Empirical wavelet transform. *IEEE Trans. Signal Process.* **2013**, *61*, 399–4010.
17. Huang, G.B. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Netw.* **2006**, *17*, 879–892.
18. 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.
19. Zhang, R.; Lan, Y.; Huang, G.-B.; Xu, Z.-B. Universal approximation of extreme learning machine with adaptive growth of hidden nodes. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 365–371.

20. 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.
21. Abd-el-Malek, M.; Abdelsalam, A.K.; Hassan, O.E. Induction motor broken rotor bar fault location detection through envelope analysis of start-up current using hilbert transform. *Mech. Syst. Signal Process.* **2017**, *93*, 332–350.
22. Gangsar, P.; Tiwari, R. Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. *Mech. Syst. Signal Process.* **2017**, *94*, 464–481.
23. 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.* **2010**, *46*, 1350–1359.
24. Henriquez, P.; Alonso, J.B.; Ferrer, M.A.; Travieso, C.M. Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 642–652.
25. Tang, J.; Qiao, J.; Wu, Z.; Chai, T.; Zhang, J.; Yu, W. Vibration and acoustic frequency spectra for industrial process modeling using selective fusion multi-condition samples and multi-source features. *Mech. Syst. Signal Process.* **2018**, *99*, 142–168.
26. Frosini, L.; Harlişca, C.; Szabo, L. Induction machine bearing fault detection by means of statistical processing of the stray flux measurement. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1846–1854.
27. Salah, M.; Bacha, K.; Chaari, A. An improved spectral analysis of the stray flux component for the detection of air-gap irregularities in squirrel cage motors. *ISA Trans.* **2014**, *53*, 816–826.
28. Cerrada, M.; Sánchez, R.-V.; Li, C.; Pacheco, F.; Cabrera, D.; de Oliveira, J.V.; Vásquez, R.E. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *99*, 169–196.
29. Delvecchio, S.; Bonfiglio, P.; Pompoli, F. Vibro-acoustic condition monitoring of internal combustion engines: A critical review of existing techniques. *Mech. Syst. Signal Process.* **2018**, *99*, 661–683.
30. Giantomassi, A.; Ferracuti, F.; Iarlori, S.; Ippoliti, G.; Longhi, S. *Signal Based Fault Detection and Diagnosis for Rotating Electrical Machines: Issues and Solutions*; Springer International Publishing: Cham (ZG), Switzerland, 2014; pp. 275–309.
31. Kang, M.; Kim, J.-M. Singular value decomposition based feature extraction approaches for classifying faults of induction motors. *Mech. Syst. Signal Process.* **2013**, *41*, 348–356.
32. Gan, M.; Wang, C.; Zhu, C. Multiple-domain manifold for feature extraction in machinery fault diagnosis. *Measurement* **2015**, *75*, 76–91.
33. Bearing Data Center. Case Western Reserve University. Technical Report. 2017. Available online: <https://csegroups.case.edu/bearingdatacenter/home> (accessed on 11 October 2017).
34. Lou, X.; Loparo, K.A. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal Process.* **2004**, *18*, 1077–1095.
35. Chen, J.; Dou, Y.; Li, Y.; Li, J. Application of shannon wavelet entropy and shannon wavelet packet entropy in analysis of power system transient signals. *Entropy* **2016**, *18*, 1–17.
36. Samui, A.; Samantaray, S.R. Wavelet singular entropy-based islanding detection in distributed generation. *IEEE Trans. Power Deliv.* **2013**, *28*, 411–418.
37. Taiyong, L.; Min, Z. Ecg classification using wavelet packet entropy and random forests. *Entropy* **2016**, *18*, 1–16.
38. Zhigang, L.; Cui, Y.; Li, W. Combined power quality disturbances recognition using wavelet packet entropies and s-transform. *Entropy* **2015**, *17*, 5811–5828.
39. Coifman, R.; Donoho, D. Translation-invariant de-noising. *Wavelets Stat. Lect. Notes Stat.* **1995**, *102*, 125–150.
40. Nason, G.; Silverman, B. The stationary wavelet transform and some statistical applications. *Wavelets Stat. Lect. Notes Stat.* **1995**, *103*, 281–300.
41. Pesquet, J.-C.; Krim, H.; Carfantan, H. Time-invariant orthonormal wavelet representations. *IEEE Trans. Signal Process.* **1996**, *44*, 1964–1970.
42. Daubechies, I. *Ten Lectures on Wavelet*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992.
43. Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press: San Diego, CA, USA, 1999.
44. Randall, R.B.; Antoni, J. Rolling element bearing diagnostics: A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520.



45. Klema, V.C.; Laub, A.J. The singular value decomposition: Its computation and some applications. *IEEE Trans. Autom. Control* **1980**, *25*, 164–176.
46. Huang, G.-B.; Chen, L. Convex incremental extreme learning machine. *Neurocomputing* **2007**, *70*, 3056–3062.
47. Serre, D. *Matrices: Theory and Applications*; Springer: New York, NY, USA, 2002.
48. Geisser, S. The predictive sample reuse method with applications. *J. Am. Stat. Assoc.* **1975**, *70*, 320–328.
49. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 111–142.
50. Sun, P.; Liao, Y.; Lin, J. The shock pulse index and its application in the fault diagnosis of rolling element bearings. *Sensors* **2017**, *17*, doi:10.3390/s17030535.
51. Ferri, C.; Hernández-Orallo, J.; Modroiu, R. An experimental comparison of performance measures for classification. *Pattern Recognit. Lett.* **2009**, *30*, 27–38.
52. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).