

Article

# A Hybrid Feature Model and Deep-Learning-Based Bearing Fault Diagnosis

Muhammad Sohaib <sup>1</sup>, Cheol-Hong Kim <sup>2</sup> and Jong-Myon Kim <sup>1,\*</sup>

<sup>1</sup> Department of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan 44610, Korea; md.sohaibdurrani@gmail.com

<sup>2</sup> School of Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea; chkim22@chonnam.ac.kr

\* Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Received: 2 November 2017; Accepted: 7 December 2017; Published: 11 December 2017

**Abstract:** Bearing fault diagnosis is imperative for the maintenance, reliability, and durability of rotary machines. It can reduce economical losses by eliminating unexpected downtime in industry due to failure of rotary machines. Though widely investigated in the past couple of decades, continued advancement is still desirable to improve upon existing fault diagnosis techniques. Vibration acceleration signals collected from machine bearings exhibit nonstationary behavior due to variable working conditions and multiple fault severities. In the current work, a two-layered bearing fault diagnosis scheme is proposed for the identification of fault pattern and crack size for a given fault type. A hybrid feature pool is used in combination with sparse stacked autoencoder (SAE)-based deep neural networks (DNNs) to perform effective diagnosis of bearing faults of multiple severities. The hybrid feature pool can extract more discriminating information from the raw vibration signals, to overcome the nonstationary behavior of the signals caused by multiple crack sizes. More discriminating information helps the subsequent classifier to effectively classify data into the respective classes. The results indicate that the proposed scheme provides satisfactory performance in diagnosing bearing defects of multiple severities. Moreover, the results also demonstrate that the proposed model outperforms other state-of-the-art algorithms, i.e., support vector machines (SVMs) and backpropagation neural networks (BPNNs).

**Keywords:** autoencoders; bearing fault diagnosis; fault diagnosis; fault severity; hybrid features; multi crack size; stacked autoencoders

---

## 1. Introduction

In the case of rotating machines, bearings are vital and common parts of the machine systems that are used in a variety of industries [1]. Because these parts are extensively used, bearings are prone to health degradation, which contributes to approximately 50% of the failures in electrical machines [2–4]. The health degradation of bearings results in unexpected failures of machines, which can lead to long downtimes, large economic losses, and human injuries [5–7]. Such issues can be mitigated with the help of fault diagnosis that assures the smooth operation of the systems by predicting their health states [8–11]. Bearing fault diagnosis, with the help of data obtained via vibration signals, acoustic emissions, electric currents, and temperature monitoring, has been a key area of research over the last few decades [12–14]. Bearing fault diagnosis is helpful in reducing the operational and maintenance costs and enhancing the reliability of a machine [15–21]. Vibration acceleration signals, which can be collected with an accelerometer, are extensively used in bearing fault diagnosis. Defective bearings add weak fault signatures to vibration signals whenever a rolling element strikes the fault location and can be explored via suitable signal processing techniques such as envelope analysis [22]. In general, a fault diagnosis pipeline has three stages: data acquisition, feature extraction, and fault

type classification. Most recent studies related to bearing fault diagnosis have focused on identifying appropriate features of the raw vibration signals. The signals measured from the operational bearings are nonstationary and nonlinear in nature due to the variable operating conditions and multiple fault severities. Therefore, in such conditions, analysis of the measured signals by means of classical signal processing techniques alone, like the fast Fourier transform, is considered to be insufficient because they provide a global transformation that is unable to properly capture the local time–frequency properties of a signal [23]. The nonstationary behavior can be explored by various time–frequency analysis techniques, including the Wigner Ville distribution (WVD) [24], short time Fourier transform (STFT) [25–27], and wavelet packet transform (WPT) [28,29]. The WPT is more practical in fault diagnosis schemes because of its better time–frequency resolution. Numerous studies investigating the time domain, frequency domain, and time–frequency domain features have been carried out to design fault diagnosis schemes using vibration signals in collaboration with machine learning (ML) methods (e.g., regression models, support vector machines, and artificial neural networks (ANNs)) [30–36]. Huo et al. [37] presented a multi-speed fault diagnosis scheme with the help of self-adaptive wavelet transform components. Particle swarm optimization (PSO) and Broyden–Fletcher–Goldfarb–Shanno (BFGS)-based quasi-Newton minimization algorithms were considered in their scheme. The aim of their work was to determine the optimal parameters for impulse modeling the continuous wavelet transform (IMCWT). The scheme could discriminate signatures for four different health conditions. In [38], time-domain (TD) statistical features were preprocessed instead of preprocessing the vibration signal prior to implementing a classifier. Preprocessing the features helped in removing the effects of possible fluctuation and random impulses in the vibration signals. An advantage of feature preprocessing, in contrast to the traditional approach where the raw vibrational signal is preprocessed, is its computational efficiency. To achieve enhanced dimensionality reduction and improve the fault diagnosis performance, an improved manifold learning scheme based on the Mahalanobis distance (MD) was proposed in [39]. Time and frequency domain analyses were performed in the scheme to construct a high-dimensional feature set. The results of the proposed scheme were found to be better than those of the traditional manifold algorithms. The authors in [40] presented a frequency domain analysis of low-speed bearings by employing time varying and multiresolution envelope analysis (TVMREA) in combination with genetic algorithm (GA)-based discriminative feature analysis (GADFA). The proposed method effectively identified combined faults in low-speed bearings.

In recent years, deep learning has made a remarkable impact on pattern recognition, image processing, and natural language processing. Deep learning mimics the learning process of the human brain in artificial networks and has displayed superior ability in capturing useful information from the input data via non-linear transformations. In contrast to conventional machine learning algorithms, deep networks can extract highly representative features via multiple layered architectures, simplifying the learning task. In addition, deep networks keep only the most representative information in each layer and discard the rest, thereby reducing the dimensionality. Hence, due to the simplified learning capability and built-in feature reduction mechanisms, deep networks can be used for fault diagnosis of complex rotary machine bearings.

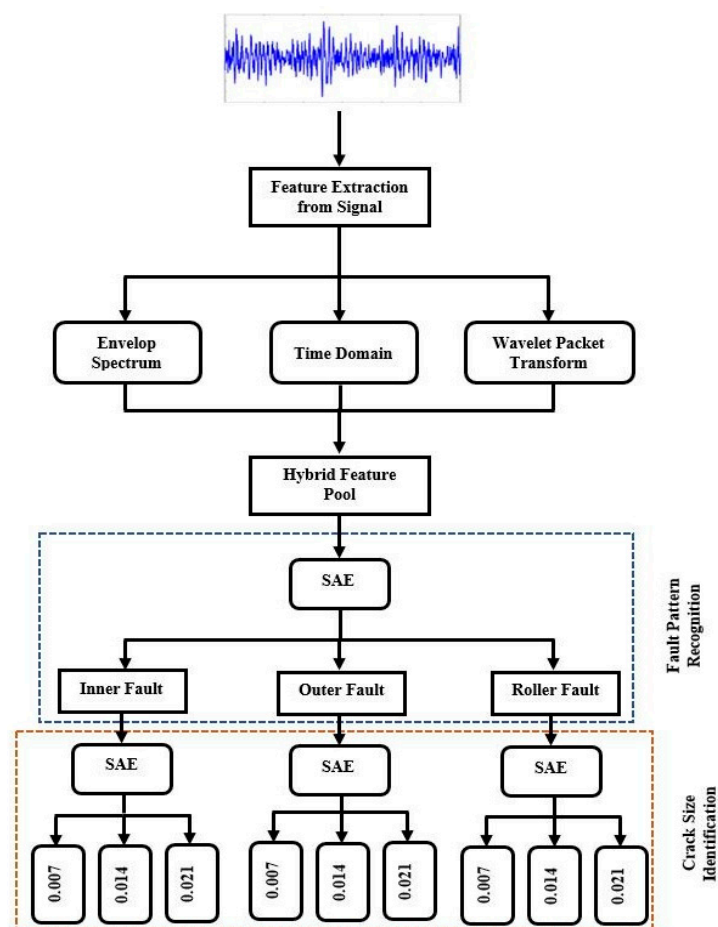
Despite the existence of several state-of-the-art bearing fault diagnosis schemes, there is still room for improvement in machine fault diagnosis; for instance, dealing with the bearing signals that exhibit nonstationary behavior due to variable working conditions and multiple fault severities. Fault pattern identification and crack size identification are two key aspects of bearing fault diagnosis. Fault pattern identification is essential as it can allow the localization of faults on a given component, whereas determining the fault severity is vital because it can highlight the urgency of repairing or replacing a damaged component. A fault diagnosis scheme that can perform both fault pattern identification and fault severity classification can be very challenging; it requires better feature representation and a strong classifier. Existing fault diagnosis schemes are vulnerable to fault misidentification due to the presence of fluctuations and random amplitudes in the vibration signals caused by multiple crack sizes.

To solve this issue, we present a two-layered fault diagnosis scheme that uses a set of hybrid features and a sparse stacked autoencoder (SSAE)-based deep neural network (DNN). The fluctuations and random amplitudes caused by multiple crack sizes cannot be overcome by analyzing the signal in just the time or frequency domain. However, a hybrid feature pool that is constructed after analyzing the vibration signal in different domains can provide sufficient information to effectively segregate bearings of different health conditions. Sparse stacked autoencoders (SSAEs) are deep neural networks (DNNs) that can extract intrinsic information from the input hybrid feature pool effectively, due to the highly nonlinear activation function used in the hidden layers. The first layer of the proposed scheme is for fault pattern identification, whereas the second layer identifies the crack size in each fault type.

The rest of the paper is organized as follows: Section 2 presents the proposed methodology. Section 3 describes the data set used for the experiments. Section 4 details the experimental results of the proposed scheme, and Section 5 concludes the paper.

## 2. Methodology

The workflow of the proposed scheme is presented in Figure 1. The scheme can be divided into three phases. The first phase consists of hybrid feature pool generation, which involves combining time domain features, envelope power spectrum features, and wavelet energy features. In the next phase, the hybrid feature pool is provided as input to the stacked autoencoders to perform fault pattern identification (i.e., identifying inner raceway, outer raceway, and roller element faults). The last phase of the pipeline is to predict the crack size for a given fault.



**Figure 1.** The proposed hierarchical fault diagnosis model, where SAE stands for stacked autoencoder.

The main idea of the work is to utilize a hybrid feature pool in combination with an SSAE-based DNN to extract high-level representative features, which would enhance the performance of the fault diagnosis model in the presence of multiple crack sizes. The hybrid feature pool provides more discriminating information about the raw vibrational signals and can overcome the nonstationary behavior of the input signal to boost the performance of the subsequent SSAE-based DNN. To create the hybrid feature pool, various feature extraction paradigms, including envelope power spectrum analysis, time domain analysis, and wavelet packet energy features, are used together.

### 2.1. Statistical Features

The representative set of statistical time domain features used in [41] was adopted in our study. The time domain statistical features that are included in the hybrid feature pool are the root mean square value (RMS), kurtosis value (KV), square root of the magnitude (SRM), peak-to-peak value (PPV), standard deviation (SD), skewness value (SV), margin factor (MF), crest factor (CF), impulse factor (IF), kurtosis factor (KF), and mean value (MV). The given representative features are listed in Table 1 with their respective mathematical formulations.

**Table 1.** Time domain statistical features ( $x$  is the vibrational signal).

Features	Equations	Features	Equations	Features	Equations
Mean value (MV)	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$	Standard deviation (SD)	$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$	Root mean square (RMS)	$RMS = \left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{\frac{1}{2}}$
Peak-to-peak value (PPV)	$PPV = \max(x_i) - \min(x_i)$	Skewness value (SV)	$SV = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^3$	Margin factor (MF)	$MF = \frac{\max( x_i )}{\left( \frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^{\frac{1}{2}}}$
Crest factor (CF)	$CF = \frac{\max( x_i )}{\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{\frac{1}{2}}}$	Impulse factor (IF)	$IF = \frac{\max( x_i )}{\frac{1}{N} \sum_{i=1}^N  x_i }$	Square root of the magnitude (SRM)	$SRM = \left( \frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$
Kurtosis value (KV)	$KV = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^4$	Kurtosis factor (KF)	$KF = \frac{\frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \bar{x}}{\sigma} \right)^4}{\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^2}$		

### 2.2. Envelope Power Spectrum

A typical bearing found in a motor has four components: the outer raceway (OR), inner raceway (IR), cage (C), and the rolling elements (RE). At a constant speed, when a bearing has a defect on any of these components, periodic vibrations are generated. There are four fundamental defect frequencies: the ball spin frequency ( $B_{SF}$ ), the ball pass outer raceway frequency ( $B_{PFO}$ ), the ball pass inner raceway frequency ( $B_{PFI}$ ), and the cage frequency ( $F_C$ ). According to [42], the  $B_{PFI}$ ,  $B_{PFO}$ , and  $B_{SF}$  can be mathematically formulated as shown in Equations (1)–(3), respectively:

$$B_{PFI} = \frac{N_b S}{2} \left( 1 + \frac{B_d}{P_d} \cos \phi \right), \quad (1)$$

$$B_{PFO} = \frac{N_b S}{2} \left( 1 - \frac{B_d}{P_d} \cos \phi \right), \quad (2)$$

$$B_{SF} = \frac{P_d}{2B_d} S \left[ 1 - \left( \frac{B_d}{P_d} \cos \phi \right)^2 \right]. \quad (3)$$

Theoretically, if the defect is on the inner or outer raceway, an impulse is added to the vibration signal whenever the rolling element strikes the defective component. These impulses can be visualized from the associated defect frequencies, i.e.,  $B_{PFI}$  and  $B_{PFO}$ , respectively. If the defect is on a rolling ball, each time it strikes the inner raceway or outer raceway an impulse will be generated; theoretically, this will be twice the  $B_{SF}$ . These fundamental defect frequencies can be useful for identifying faults

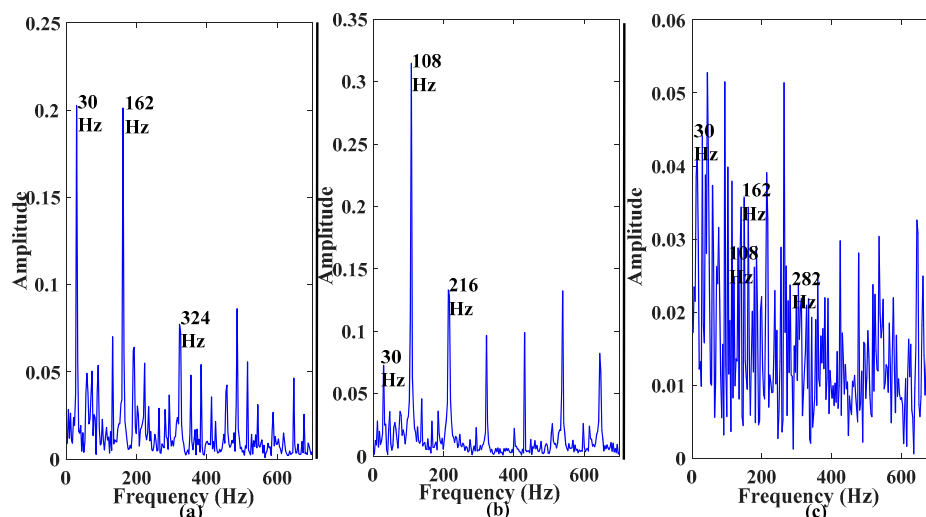
on the inner raceway, outer raceway, and the rolling element. These impulses at the associated defect frequencies can be explored via envelope spectrum analysis.

The envelope of a vibration signal  $s(t)$  can be calculated by using the Hilbert transform. The Hilbert transform is a convolution between the Hilbert transform operator  $\frac{1}{\pi t}$  and the original signal  $s(t)$  [43]. It can be represented as

$$H[s(t)] = s(t) \bullet \frac{1}{\pi t}, \quad (4)$$

$$H[s(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(t)}{t - \tau} dt, \quad (5)$$

where  $\bullet$  is the convolution operator in (4) and  $H[s(t)]$  is an analytical signal of the original signal  $s(t)$ . By taking the square of the fast Fourier transform of  $abs(H[s(t)])$ , a one-sided spectrum in the frequency domain can be obtained; this is the desired envelope power spectrum. The envelope power spectra of three fault types can be seen in Figure 2. Features extracted from the envelope spectra of the given example are presented in Table 2.



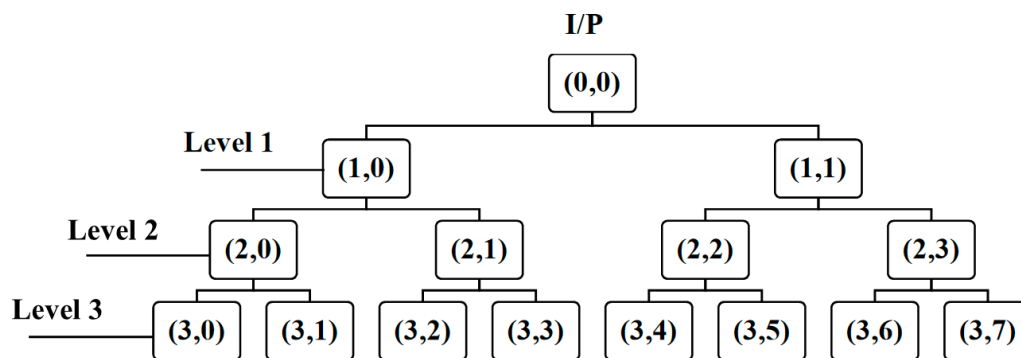
**Figure 2.** Envelope power spectrum: (a) inner raceway fault; (b) outer raceway fault, and (c) roller element fault.

**Table 2.** Statistical features extracted from the envelope power spectrum.

Feature	Equation
RMS frequency	$RMS_f = \left( \frac{1}{K} \sum_{i=1}^K y_K^2 \right)^{\frac{1}{2}}$
Frequency center	$FC = \frac{1}{K} \sum_{i=1}^K y_K$
Standard deviation	$\sigma_f^2 = \frac{1}{K} \sum_{i=1}^K (y_K - FC)^2$
Root variance frequency	$RVF = \left( \frac{1}{K} \sum_{i=1}^K (y_K - FC)^2 \right)^{\frac{1}{2}}$
Spectral kurtosis	$K_f = \frac{1}{K} \frac{\sum_{i=1}^K (y_K - FC)^4}{(\sigma_f^2)^2}$

### 2.3. Wavelet Packet Transform (WPT)

The wavelet packet transform (WPT) is a variation of the basic wavelet transform (WT) that decomposes the input signal into  $j$  levels. The WPT splits both the high-pass and low-pass filters, creating  $2^j$  nodes at each level. The WPT overcomes the poor resolution of the WT by providing comprehensive time–frequency analysis of the signal at both low and high frequencies. Each level of the WPT provides a frequency range that is half as wide as the preceding level and twice as wide as the preceding level. A three-level WPT tree structure can be seen in Figure 3.



**Figure 3.** An illustration of a three-level wavelet packet tree decomposition, where I/P stands for input.

The WPT coefficients can be formulated as

$$c_{j+1}^{2k}(n) = c_j^k \times h(-2n), \quad 0 < k < 2^j - 1, \quad (6)$$

$$d_{j+1}^{2k+1}(n) = d_j^k(n) \times g(-2n), \quad 0 < k < 2^j - 1, \quad (7)$$

where  $h$  and  $g$  are the low-pass and high-pass filters associated with the mother wavelet, respectively. These are predefined scaling factors. In the WPT, the scale parameter (level) is represented by  $j$ , and the frequency parameters (nodes) are represented by  $2k$  and  $2k + 1$ .

Existing methods based on the WPT for bearing fault diagnosis consider the entropy, standard variation, and energy as input features to the subsequent classifier. Among these, using the wavelet packet energy is an intuitive approach to differentiating the fault types. The WPT nodes contain an abundance of information about the fault types and energy fluctuations in a specific node and can be useful in specifying the fault type.

In the current work, signals are decomposed up to  $j = 4$  levels, as described in [44], which results in  $2^j = 2^4 = 16$  nodes. After decomposition of the signals into different sub-bands, the WPT energy is computed by

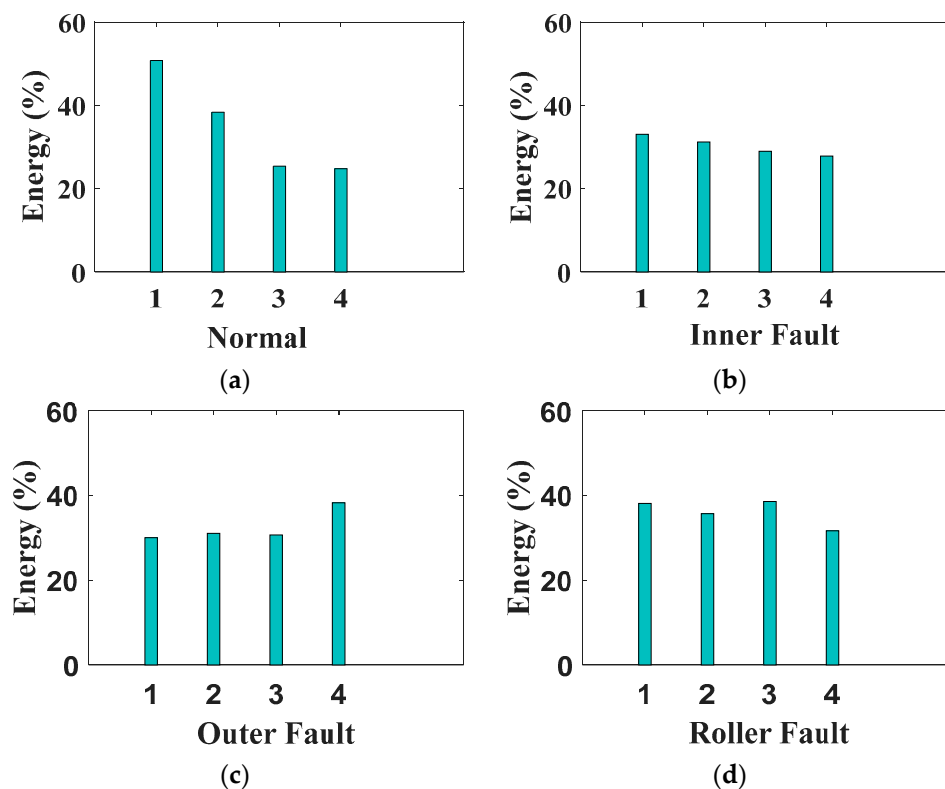
$$E = \left( \frac{\sum_{p=1}^M (c_j^k(p)^2)^{\frac{1}{2}}}{M} \right). \quad (8)$$

In the equation above,  $M$  is the number of samples at the nodes. All the energies acquired from  $j = 4$  level nodes are combined to form the vector  $V$ , which can be given as

$$V = [E_j^1, E_j^2, \dots, E_j^{2^j}]. \quad (9)$$

The maximum value of the vector is selected for each input signal and included in the hybrid feature pool. The extracted wavelet energy features can be seen in Figure 4. In the figure, wavelet energies for four different health conditions of the bearing are given. For each health condition, four signals are available—one for each motor load and rotational speed (i.e., 1722 to 1797 r/min). We notice that there is a variation in the wavelet energy levels for different health conditions, which

can be of benefit to SSAEs in learning distinctive high-level features for a given health condition. On the other hand, there is also variation within the wavelet energy levels of a specific health condition, which can lead to confusion among the instances of different health conditions and can result in misclassification of the instances. To minimize the misclassification of the instances due to the variation in either the values of statistical features from the time domain, envelope power spectrum, or wavelet energy levels, a hybrid feature pool is formed by including the extracted time domain statistical features, envelope power spectrum, and WPT energy features. The hybrid feature pool can provide detailed intrinsic information about the nonstationary and nonlinear signals obtained from bearings with multiple fault severities. The length of the resulting hybrid features' vector is  $6 + 6 + 1 = 13$ . After creating the hybrid feature pool, it is provided as an input to the SAE-based DNN to learn high-level representative features and perform fault pattern recognition and fault severity classification.



**Figure 4.** Wavelet energy features, (a) baseline condition; (b) inner fault signals; (c) outer fault signals; and (d) roller fault signals.

#### 2.4. Sparse Stacked Autoencoders (SSAEs)

A simple autoencoder is basically a variation of an artificial neural network (ANN) with a minimum of three layers that uses an unsupervised learning process. The structure of a basic autoencoder is presented in Figure 5.

The first layer of the autoencoder is the input layer, which receives the input data. The intermediate layer tends to extract high-level representative features (i.e., latent codes) from the input data. The latent codes are, in essence, the result of principal component analysis (PCA) of the inputs and reduce the representation of the original data. The dimensionality of the latent codes depends on the number of nodes used in the hidden layer. The last layer decodes the latent codes and tries to reconstruct the original input. In short, an autoencoder performs two key tasks: to encode the input data into latent codes and then reconstruct the data from the latent codes. The resulting latent codes have

lower dimensionality than the input data. In this regard, an autoencoder contributes to dimensionality reduction. The encoding  $\partial$ , and decoding  $\beta$  processes of an autoencoder are described as follows:

$$\begin{aligned} \partial &: s \rightarrow F \\ \beta &: F \rightarrow s \\ \partial, \beta &= \operatorname{argmin}_{\partial, \beta} \|s - (\beta \bullet \alpha)s\|^2. \end{aligned} \quad (10)$$

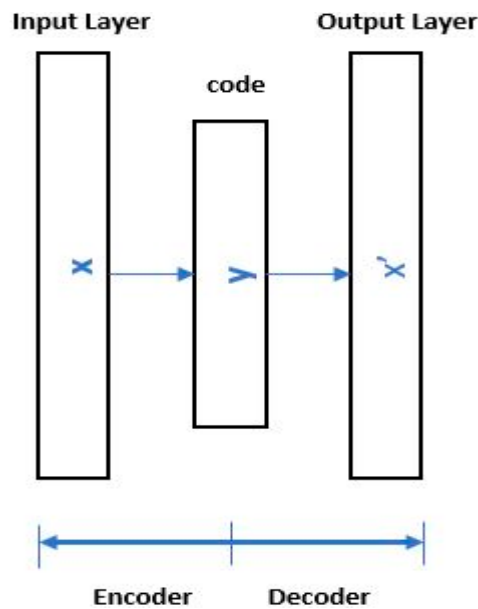


Figure 5. Structure of a basic autoencoder.

The simplest form of an autoencoder has one hidden layer. The encoder stage receives input data  $s$  with dimension  $R^m$  and maps the data to latent variables  $o$  with dimension  $R^n$ . The latent code can be given by

$$o = g(Ws + b), \quad (11)$$

where  $o$ ,  $W$ ,  $b$ , and  $g$  are the latent code, weight matrix, bias vector, and activation function, respectively. Equation (12) presents the decoding process of an autoencoder:

$$r = g'(W's + b') \quad (12)$$

where  $r$ ,  $W'$ ,  $b'$ , and  $g'$  are the reconstructed output, weight matrix, bias vector, and activation function of the decoder, respectively. The loss function is calculated between the original data and the reconstructed data in basic autoencoders by using the following loss function:

$$L(s, r) = \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^K (s_{km} - r)^2 \quad (13)$$

where  $L$  is the loss calculated between the original data  $s$  and the reconstructed data  $r$ . A sparsity constraint can be introduced in an autoencoder by introducing a sparsity regularization term to the loss function. The sparsity constraint enables the autoencoder to learn useful features that can be used for classification [45]. The modified loss function can be represented as follows:

$$L(s, r) = \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^K (s_{km} - r)^2 + \lambda \cdot \Omega_{weights} + \lambda' \cdot \Omega_{sparsity} \quad (14)$$



where  $\lambda$  is the  $L_2$  regularization coefficient and  $\lambda'$  is the sparsity regularization coefficient.  $\Omega_{weights}$  is the  $L_2$  regularization term and  $\Omega_{sparsity}$  is the sparsity regularization term.  $L_2$  regularization and  $\Omega_{sparsity}$  regularization help in avoiding the overfitting problem in sparse autoencoders.

### 3. Dataset

To demonstrate the efficacy of the proposed model, seeded fault data provided by Case Western Reserve University was used. As illustrated in Figure 6, the main components of the seeded fault test rig include a 2 horsepower (hp) electric motor, a dynamometer, and a torque transducer [46].

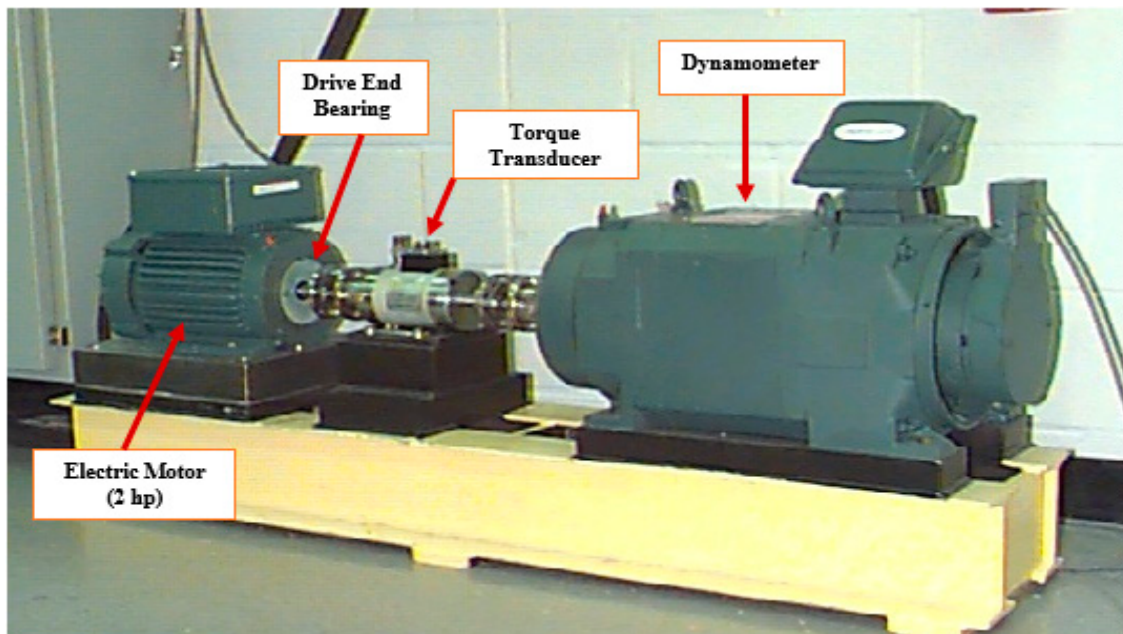


Figure 6. Case Western Reserve University's seeded fault bearing testbed [46].

Using an electro-discharge machine, faults with diameters of 0.007, 0.014, and 0.021 inches were seeded on the inner raceway (IR), outer raceway (OR), and rolling elements (RE) at the drive end bearings. Variable length vibration acceleration signals were collected via an accelerometer attached to the housing of the drive end bearing at 12 o'clock with a sampling data rate of 12,000 Hz. The motor was subject to four loads ranging from 0 to 3 horsepower (hp), which resulted in four motor speeds, approximately from 1722 to 1797 revolutions per minute (r/min).

In this study, the dataset comprises vibration acceleration signals for normal bearings and bearings with three types of faults, i.e., faults on the inner raceway, outer raceway, and rolling element. For each fault condition, the dataset consists of signals recorded for bearings with three levels of fault severities (i.e., 0.007, 0.014, and 0.021 inches) at four different shaft loads. For normal bearings, there are four signals in the dataset—one for each shaft load. The signals are subjected to a segmentation process using a fixed sized window of 12,000 data points. The segmentation process splits all the fault signals into 10 samples each, but three of the four normal signals yield 20 samples each, while the fourth normal signal yields only 10 samples. The length of each sample for both normal and faulty bearings is 12,000 data points. Thus, the seeded fault dataset used for the experiments contains a total of 610 samples (70 normal samples + 3 fault types  $\times$  3 fault severities  $\times$  60 samples). After segmentation, a hybrid feature vector is constructed for each sample in the dataset. These feature vectors are then divided into training and test sets. The training set contains feature vectors for 310 samples (40 normal samples + 3 fault types  $\times$  3 fault severities  $\times$  30 samples), while the test set consists of feature vectors for 300 samples (30 normal samples + 3 fault types  $\times$  3 fault severities  $\times$  30 samples). The details of the bearing dataset with seeded faults are given in Table 3.

Table 3. Bearings and dataset specification.

Fault Type	Fault Location	Fault Size (Inches)	Training Samples	Test Samples	Sample Length	Accelerometer Position	Shaft Load (hp)
Normal	None	0	40	30	12,000	Drive End Bearings	0, 1, 2, 3
Inner raceway	IR	0.007	30	30			
	IR	0.014	30	30			
	IR	0.021	30	30			
Outer raceway	OR	0.007	30	30			
	OR	0.014	30	30			
	OR	0.021	30	30			
Roller	RE	0.007	30	30			
	RE	0.014	30	30			
	RE	0.021	30	30			

#### 4. Results and Analysis

A bearing dataset with seeded faults was provided by Case Western Reserve University [46] and used to validate the proposed fault diagnosis model. The dataset is composed of four health conditions and three different fault severities. For training and evaluation of the first layer, all the samples from training set were used to train the first sparse stacked autoencoder (SSAE)-based deep neural network (DNN). On the other hand, while training the rest of the three SSAE-DNNs in the crack size identification layer, only samples from the respective fault classes were considered. To produce stable results, the experiment was repeated 20 times with random selection of samples to form the train and test sets each time. To evaluate the effectiveness of the proposed scheme, the results were compared with those of the state-of-the-art algorithms, including the radial basis function (RBF) kernel-based one-against-all support vector machines (OAASVMs) and backpropagation neural networks (BPNNs). All the SSAE-DNNs in the proposed scheme were replaced with RBF-OAASVMs and BPNNs to create a similar hierarchical structure. The same set of features were provided as input to the RBF-OAASVMs and two layered BPNNs with 10 hidden nodes. The Levenberg–Marquardt backpropagation optimization function was used in the BPNNs to update the weights. Figure 7 presents the results of the fault pattern identification that is proposed to identify the bearing health conditions (i.e., normal health or having a fault on the inner raceway, outer raceway, or roller element) for the proposed method and the state-of-the-art algorithms. The overall average accuracy of the proposed model for the fault pattern identification layer is 99.5%.

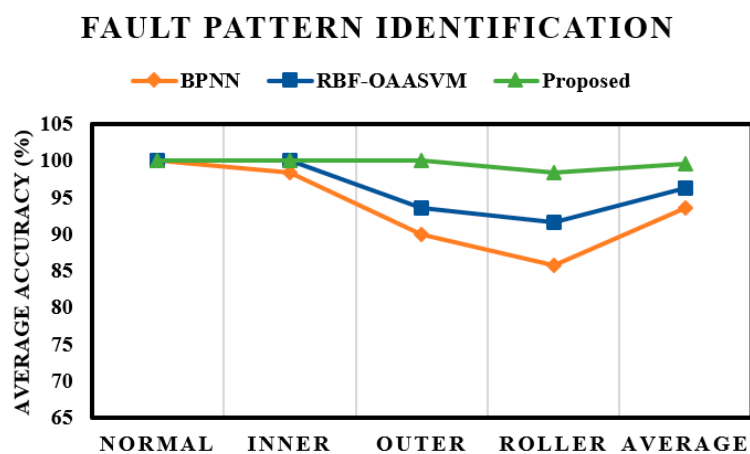


Figure 7. Bearing fault classification results for the first layer (i.e., the fault pattern recognition layer), BPNN and RBF-OAASVM stand for back-propagation neural network and radial basis function-one against all support vector machine, respectively.

The SSAE-based DNN, because of its hierarchical structure and by using nonlinear transformation in the hidden layers, could extract discriminating information from the hybrid feature pool, enhancing the overall performance of the proposed model. This observation is validated by Figure 8, which contains the distribution of the first two feature vectors extracted by using SSAEs. It is worth noticing that the proposed method correctly classified all the samples for inner and outer faults; however, it misclassified a few of the roller fault samples. The hybrid feature pool fails to provide enough intrinsic information, and, thus, SSAEs fail to extract more discriminant high-level features in this case. In the case of the normal condition and inner and outer faults, high-level feature extraction seems relatively easy for SSAEs. In the case of roller fault, the extracted high-level features overlap with some of the samples from the normal and inner fault, which leads to the misclassification of roller fault samples. The roller fault signals possess the properties of inner as well as outer faults. This observation is validated by Figure 2, where the envelope power spectrum of the roller fault is given. The presence of inner and outer fault defect frequencies can be clearly seen in the envelope power spectrum. Therefore, features extracted from the time domain, envelope spectrum, and wavelet energy, in this case, were either confused with inner or roller fault. Moreover, from the comparison results, it is evident that the proposed model provided 3.32% and 6.12% better average accuracy for the fault pattern identification layer than the RBF-OAASVMs and BPNNs, respectively.

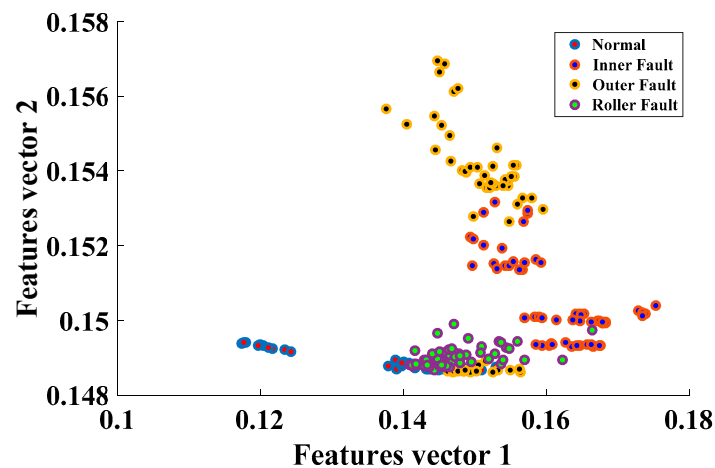


Figure 8. The extracted high-level features for the fault pattern recognition layer.

The fault pattern recognition layer is followed by the crack size identification layer. The results of the subsequent layer mainly depend on the results of the first layer; if the performance of the first layer is poor, the results of the subsequent layer will also be poor. From the results of the pattern recognition layer, it is evident that the proposed method could classify most of the input instances, which ultimately boosted the performance of the proposed scheme. This observation is validated by the results of the crack size identification layer. Figure 9 shows the fault severity classification of an inner fault; once again, the performance of the proposed method is better than those of the RBF-OAASVM and BPNN methods. The proposed method provides an average accuracy of 100%, whereas RBF-OAASVMs and BPNNs provide average accuracies of 94.4% and 90.44%, respectively. It can be interpreted from the results that the proposed scheme successfully classifies all the samples into their respective classes. However, RBF-OAASVMs and BPNNs fail to classify all the samples properly. Figure 10 shows the results for crack size prediction within an outer fault. The average fault severity accuracies for the proposed method, RBF-OAASVMs, and BPNNs are 100%, 93.56%, and 85.03%, respectively. In Figure 11, the results of crack size identification in terms of the average accuracy in a roller fault are given. It is evident that the proposed method outperforms the SVM and BPNN methods. In this case, there is slight deterioration in the performance of the proposed method, but, still, it delivers better performance compared with the other state-of-the-art algorithms.

The deterioration is due to the misclassification of outer fault samples in the fault pattern identification layer, consequently affecting the results of the crack size identification layer in the case of roller faults. Overall, our proposed method has an average accuracy of 96.66%, while the average crack size prediction accuracies of SVMs and BNNs are 92.33% and 83.44%, respectively.

### INNER FAULT

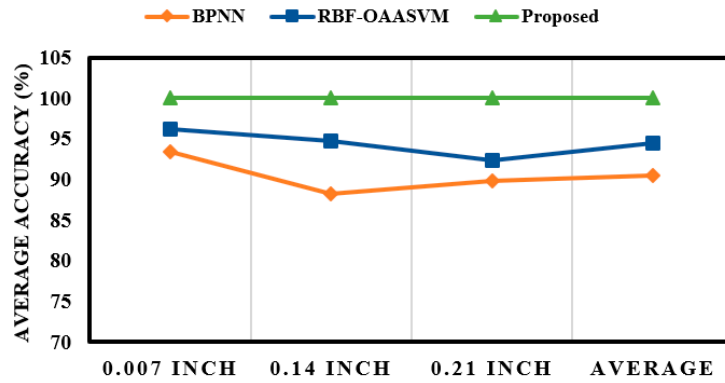


Figure 9. Fault severity prediction in an inner fault.

### OUTER FAULT

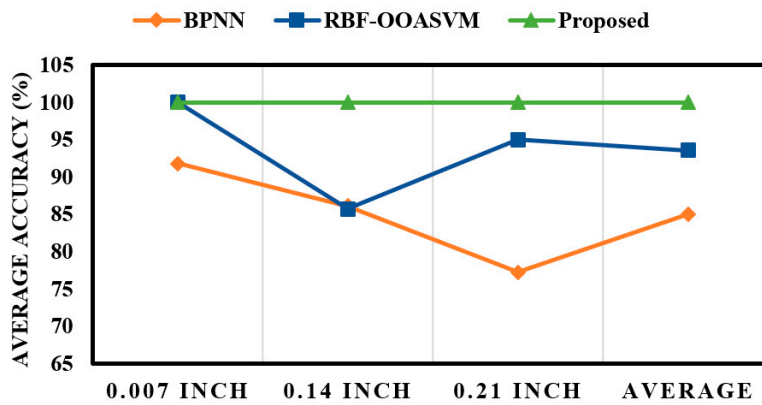


Figure 10. Fault severity prediction in an outer fault.

### ROLLER FAULT

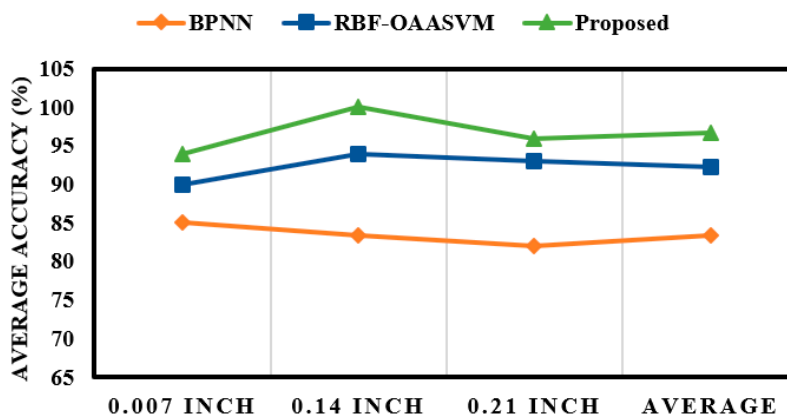


Figure 11. Fault severity prediction in a roller fault.

To further validate the reliability of the proposed method, a comparison is made with an existing bearing fault diagnosis scheme [47], in which the authors used vibration spectrum imaging (VSI) and artificial neural networks (ANN) for bearing fault diagnosis. The bearing dataset used for validation of the scheme was acquired from Case Western Reserve University (shaft load of 2 hp with 1748 r/min). The vibration signals were segmented into fixed sized windows of 1024 data points each, and then a 513 point fast Fourier transform (FFT) was applied to the segmented signals. The resultant spectral information was stacked on top of each other to generate a  $513 \times 8$  pixel grayscale vibration spectrum image. A smoothing filter of size  $8 \times 4$  was applied to the grayscale image, and then the filtered image was converted into a binary image by using an optimum threshold value of 0.7. The optimum threshold value plays a key role in the VSI-based fault diagnosis scheme because it defines the quality of the input vectors to the underlying classifier, and can affect the overall accuracy of the scheme. Then, the binary images, each with 4104 binary spectral components, were provided as an input to an artificial neural network having one hidden layer with three nodes. The comparison results of the proposed method and the VSI-based fault diagnosis scheme are presented in Table 4. The proposed method provides better diagnostic performance as compared with VSI when tested on the dataset containing instances from the seeded fault bearings with multiple fault severities. The proposed method can overcome the nonstationary and nonlinear behavior of the vibration signal in a much better way compared with the VSI-based approach, where the spectral information is more susceptible to variation in working conditions and fault severities.

**Table 4.** The diagnostic performance of the proposed model and vibration spectrum imaging (VSI).

Method	Layer 1 Average Accuracy (%)	Layer 2 Average Accuracy (%)			Total (%)
		0.007 Inches	0.014 Inches	0.021 Inches	
VSI [47]	60.15	55	55	84.6	63.68
Proposed	99.75	100	100	96.66	99.10

## 5. Conclusions

In this paper, a two-layered bearing fault diagnosis scheme was proposed. The first layer is for fault pattern identification in rotary machine bearings, while the subsequent layer is used for crack size identification of a given fault. A hybrid features pool comprising time domain statistical features, envelope power spectrum features, and wavelet energy features is used in combination with sparse stacked autoencoder (SSAE)-based deep neural networks (DNNs) for the diagnosis of different bearing defects with various levels of severity. The hybrid feature pool was formed to overcome the nonstationary and nonlinear behavior of the vibration acceleration signals. A bearing dataset containing four health conditions and three fault severities was used to validate the proposed model. It is observed that the SSAE-based DNN is able to extract effective representative features from the hybrid feature pool, resulting in a superior diagnostic performance of the proposed model for both fault pattern as well as for crack size identification. Moreover, the proposed model was compared with three state-of-the-art fault diagnosis algorithms (i.e., RBF-OAASVMs, BPNNs, and VSI). The results demonstrated that the proposed scheme is more effective compared with the other methods regardless of the nonlinearity contained in the vibration signals due to multiple fault severities. However, in the case of roller fault identification, the performance of the proposed method slightly deteriorated, which underscores the need for more sophisticated signal processing algorithms as future work that could eventually result in superior diagnostic performance. It can be concluded that the proposed method provides satisfactory bearing fault diagnosis results and can be used for fault diagnosis of bearings containing various fault severities.

**Acknowledgments:** This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20162220100050, No. 20161120100350, and No. 20172510102130). It was also funded in part by The Leading

Human Resource Training Program of the Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2016H1D5A1910564) and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927).

**Author Contributions:** All the authors contributed equally to the conception of the idea, the design of experiments, the analysis and interpretation of results, and the writing and improvement of the manuscript.

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

## References

1. Toliyat, H.A.; Nandi, S.; Choi, S.; Meshgin-Kelk, H. *Electric Machines: Modeling, Condition Monitoring and Fault Diagnosis*; CRC Press: Boca Raton, FL, USA, 2012.
2. Frosini, L.; Harlişca, C.; Szabó, L. Induction machine bearing fault detection by means of statistical processing of the stray flux measurement. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1846–1854. [[CrossRef](#)]
3. Islam, M.M.; Kim, J.; Khan, S.A.; Kim, J.-M. Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines. *J. Acoust. Soc. Am.* **2017**, *141*, EL89–EL95. [[CrossRef](#)] [[PubMed](#)]
4. Khan, S.A.; Kim, J.-M. Automated Bearing Fault Diagnosis Using 2D Analysis of Vibration Acceleration Signals under Variable Speed Conditions. *Shock Vib.* **2016**, *2016*. [[CrossRef](#)]
5. Jiang, X.; Wu, L.; Ge, M. A Novel Faults Diagnosis Method for Rolling Element Bearings Based on EWT and Ambiguity Correlation Classifiers. *Entropy* **2017**, *19*, 231. [[CrossRef](#)]
6. Li, S.; Liu, G.; Tang, X.; Lu, J.; Hu, J. An Ensemble Deep Convolutional Neural Network Model with Improved DS Evidence Fusion for Bearing Fault Diagnosis. *Sensors* **2017**, *17*, 1729. [[CrossRef](#)] [[PubMed](#)]
7. Zhang, R.; Peng, Z.; Wu, L.; Yao, B.; Guan, Y. Fault Diagnosis from Raw Sensor Data Using Deep Neural Networks Considering Temporal Coherence. *Sensors* **2017**, *17*, 549. [[CrossRef](#)] [[PubMed](#)]
8. Youssef, T.; Chadli, M.; Karimi, H.R.; Wang, R. Actuator and sensor faults estimation based on proportional integral observer for TS fuzzy model. *J. Frankl. Inst.* **2017**, *354*, 2524–2542. [[CrossRef](#)]
9. Li, L.; Chadli, M.; Ding, S.X.; Qiu, J.; Yang, Y. Diagnostic Observer Design for T-S Fuzzy Systems: Application to Real-Time Weighted Fault Detection Approach. *IEEE Trans. Fuzzy Syst.* **2017**, *PP*, 1. [[CrossRef](#)]
10. Chibani, A.; Chadli, M.; Shi, P.; Braiek, N.B. Fuzzy Fault Detection Filter Design for T-S Fuzzy Systems in the Finite-Frequency Domain. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1051–1061. [[CrossRef](#)]
11. Peng, H.; Wang, J.; Ming, J.; Shi, P.; Perez-Jimenez, M.J.; Yu, W.; Tao, C. Fault Diagnosis of Power Systems Using Intuitionistic Fuzzy Spiking Neural P Systems. *IEEE Trans. Smart Grid* **2017**, *PP*, 1. [[CrossRef](#)]
12. Jardine, A.K.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [[CrossRef](#)]
13. Bediaga, I.; Mendizabal, X.; Arnaiz, A.; Munoa, J. Ball bearing damage detection using traditional signal processing algorithms. *IEEE Instrum. Meas. Mag.* **2013**, *16*, 20–25. [[CrossRef](#)]
14. Tra, V.; Kim, J.; Khan, S.A.; Kim, J.-M. Incipient fault diagnosis in bearings under variable speed conditions using multiresolution analysis and a weighted committee machine. *J. Acoust. Soc. Am.* **2017**, *142*, EL35–EL41. [[CrossRef](#)] [[PubMed](#)]
15. Rauber, T.W.; de Assis Boldt, F.; Varejão, F.M. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 637–646. [[CrossRef](#)]
16. Yin, S.; Ding, S.X.; Zhou, D. Diagnosis and prognosis for complicated industrial systems—Part I. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2501–2505. [[CrossRef](#)]
17. Yin, S.; Ding, S.X.; Zhou, D. Diagnosis and prognosis for complicated industrial systems—Part II. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3201–3204. [[CrossRef](#)]
18. Chang, T.C.; Wysk, R.; Wang, H. *Computer-Aided Manufacturing*; Prentice Hall: Upper Saddle River, NJ, USA, 1991; p. 7632.
19. Qin, X.; Li, Q.; Dong, X.; Lv, S. The Fault Diagnosis of Rolling Bearing Based on Ensemble Empirical Mode Decomposition and Random Forest. *Shock Vib.* **2017**, *2017*. [[CrossRef](#)]
20. Zhao, S.; Liang, L.; Xu, G.; Wang, J.; Zhang, W. Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method. *Mech. Syst. Signal Process.* **2013**, *40*, 154–177. [[CrossRef](#)]

21. Huang, W.; Sun, H.; Wang, W. Resonance-Based Sparse Signal Decomposition and its Application in Mechanical Fault Diagnosis: A Review. *Sensors* **2017**, *17*, 1279. [[CrossRef](#)] [[PubMed](#)]
22. Abboud, D.; Antoni, J.; Sieg-Zieba, S.; Eltabach, M. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mech. Syst. Signal Process.* **2017**, *84*, 200–226. [[CrossRef](#)]
23. Zhou, S.; Tang, B.; Chen, R. Comparison between non-stationary signals fast Fourier transform and wavelet analysis. In Proceedings of the International Asia Symposium on Intelligent Interaction and Affective Computing, Wuhan, China, 8–9 December 2009; pp. 128–129.
24. Torres, R.; Torres, E. Fractional Fourier Analysis of Random Signals and the Notion of  $1/\alpha$ -Stationarity of the Wigner-ville Distribution. *IEEE Trans. Signal Process.* **2013**, *61*, 1555–1560. [[CrossRef](#)]
25. Burriel-Valencia, J.; Puche-Panadero, R.; Martinez-Roman, J.; Sapena-Bano, A.; Pineda-Sanchez, M. Short-Frequency Fourier Transform for Fault Diagnosis of Induction Machines Working in Transient Regime. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 432–440. [[CrossRef](#)]
26. Zhang, X.; Jiang, D.; Han, T.; Wang, N.; Yang, W.; Yang, Y. Rotating Machinery Fault Diagnosis for Imbalanced Data Based on Fast Clustering Algorithm and Support Vector Machine. *J. Sens.* **2017**, *2017*. [[CrossRef](#)]
27. Islam, M.M.; Kim, J.-M. Time-frequency envelope analysis-based sub-band selection and probabilistic support vector machines for multi-fault diagnosis of low-speed bearings. *J. Ambient Intell. Humaniz. Comput.* **2017**, 1–16. [[CrossRef](#)]
28. 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*, 1–35. [[CrossRef](#)]
29. He, S.; Liu, Y.; Chen, J.; Zi, Y. Wavelet Transform Based on Inner Product for Fault Diagnosis of Rotating Machinery. In *Structural Health Monitoring*; Springer: Cham, Switzerland, 2017; pp. 65–91.
30. Huang, J.; Chen, G.; Shu, L.; Wang, S.; Zhang, Y. An experimental study of clogging fault diagnosis in heat exchangers based on vibration signals. *IEEE Access* **2016**, *4*, 1800–1809. [[CrossRef](#)]
31. Kang, M.; Islam, M.R.; Kim, J.; Kim, J.-M.; Pecht, M. A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3299–3310. [[CrossRef](#)]
32. Tavner, P. Review of condition monitoring of rotating electrical machines. *IET Electr. Power Appl.* **2008**, *2*, 215–247. [[CrossRef](#)]
33. Wang, Y.; Xiang, J.; Markert, R.; Liang, M. Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. *Mech. Syst. Signal Process.* **2016**, *66*, 679–698. [[CrossRef](#)]
34. Dong, S.; Chen, L.; Tang, B.; Xu, X.; Gao, Z.; Liu, J. Rotating machine fault diagnosis based on optimal morphological filter and local tangent space alignment. *Shock Vib.* **2015**, *2015*. [[CrossRef](#)]
35. Verstraete, D.; Ferrada, A.; Droguett, E.L.; Meruane, V.; Modarres, M. Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings. *Shock Vib.* **2017**, *2017*. [[CrossRef](#)]
36. Yang, D.; Mu, H.; Xu, Z.; Wang, Z.; Yi, C.; Liu, C. Based on Soft Competition ART Neural Network Ensemble and Its Application to the Fault Diagnosis of Bearing. *Math. Probl. Eng.* **2017**, *2017*. [[CrossRef](#)]
37. Huo, Z.; Zhang, Y.; Francq, P.; Shu, L.; Huang, J. Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures. *IEEE Access* **2017**, *5*, 19442–19456. [[CrossRef](#)]
38. Tahir, M.M.; Khan, A.Q.; Iqbal, N.; Hussain, A.; Badshah, S. Enhancing Fault Classification Accuracy of Ball Bearing Using Central Tendency Based Time Domain Features. *IEEE Access* **2016**, *5*, 72–83. [[CrossRef](#)]
39. Yao, B.; Zhen, P.; Wu, L.; Guan, Y. Rolling Element Bearing Fault Diagnosis Using Improved Manifold Learning. *IEEE Access* **2017**, *5*, 6027–6035. [[CrossRef](#)]
40. Kang, M.; Kim, J.; Wills, L.M.; Kim, J.-M. Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 7749–7761. [[CrossRef](#)]
41. Xia, Z.; Xia, S.; Wan, L.; Cai, S. Spectral regression based fault feature extraction for bearing accelerometer sensor signals. *Sensors* **2012**, *12*, 13694–13719. [[CrossRef](#)] [[PubMed](#)]
42. He, D.; Li, R.; Zhu, J. Plastic bearing fault diagnosis based on a two-step data mining approach. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3429–3440. [[CrossRef](#)]
43. Dubey, R.; Agrawal, D. Bearing fault classification using ANN-based Hilbert footprint analysis. *IET Sci. Meas. Technol.* **2015**, *9*, 1016–1022. [[CrossRef](#)]

44. Nikolaou, N.G.; Antoniadis, I.A. Application of Wavelet Packets in Bearing Fault Diagnosis. In Proceedings of the 5th WSES International Conference on Circuits, Systems, Communications and Computers (CSCC 2001), Rethymno, Greece, 8–15 July 2001; pp. 12–19.
45. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
46. Case Western Reserve University. B.D.C. Seeded Fault Test Data. Available online: <http://csegroups.case.edu/bearingdatacenter/home/> (accessed on 21 January 2017).
47. Amar, M.; Gondal, I.; Wilson, C. Vibration spectrum imaging: A novel bearing fault classification approach. *IEEE Trans. Ind. Electron.* **2015**, *62*, 494–502. [[CrossRef](#)]



© 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/>).