Detection of Pitting in Gears Using a Deep Sparse Autoencoder

Yongzhi Qu [1], Miao He [2], Jason Deutsch [2] and David He [2,3,*]

[1] School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China; quwong@whut.edu.cn
[2] Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA; mhe21@uic.edu (M.H.); jdeuts4@uic.edu (J.D.)
[3] College of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China
[*] Correspondence: davidhe@uic.edu; Tel.: +86-24-8369-6132

Abstract: In this paper; a new method for gear pitting fault detection is presented. The presented method is developed based on a deep sparse autoencoder. The method integrates dictionary learning in sparse coding into a stacked autoencoder network. Sparse coding with dictionary learning is viewed as an adaptive feature extraction method for machinery fault diagnosis. An autoencoder is an unsupervised machine learning technique. A stacked autoencoder network with multiple hidden layers is considered to be a deep learning network. The presented method uses a stacked autoencoder network to perform the dictionary learning in sparse coding and extract features from raw vibration data automatically. These features are then used to perform gear pitting fault detection. The presented method is validated with vibration data collected from gear tests with pitting faults in a gearbox test rig and compared with an existing deep learning-based approach.

Keywords: gear; pitting detection; deep sparse autoencoder; vibration; deep learning

1. Introduction

Gears are one of the most critical components in many industrial machines. Health monitoring and fault diagnosis of gears are necessary to reduce breakdown time and increase productivity. Pitting is one of the most common gear faults and normally difficult to detect. An undetected gear pitting fault during the operation of the gears can lead to catastrophic failures of the machines.

In recent years, many gear pitting fault detection methods have been developed. Following the same way to classify machine fault diagnostic and prognostic methods by [1,2], gear pitting fault detection methods can be classified into two main categories, namely model-based methods and data-driven methods. The model-based techniques rely on accurate dynamic models of the systems, while the data-driven approaches use data to train fault detection models. Model-based approaches obtain the residuals between actual system and output. These residuals are then used as the indicator of the actual faults [3,4]. However, the model-based approaches require not only expertise in dynamic modeling, but also accurate condition parameters of the studied system. On the other hand, data-driven approaches do not require the knowledge of the target system and dynamic modeling expertise. In comparison with model-based techniques, data-driven approaches can design a fault detection system that can be easily applied when massive data is available. Data-driven techniques are appropriate when a comprehensive understanding of system operation is absent, or when it is sufficiently difficult to model the complicated system [5].

Data-driven-based gear pitting fault detection methods in general relies on feature extraction by human experts and complicated signal processing techniques. For example, Reference [6] used

a zoomed phase map of continuous wavelet transform to detect minor damage such as gear fitting. References [7,8] used the mean frequency of a scalogram to get features for gear pitting fault detection. Reference [9] extracted condition indicators from time-averaged vibration data for gear pitting damage detection. Reference [10] used empirical mode decomposition (EMD) to extract features from vibration signals for gear pitting detection. Reference [11] combined EMD and fast independent component analysis to extract features from stator current signals for gear pitting fault detection. Reference [12,13] applied spectral kurtosis to extract features for gear pitting fault detection. Reference [14] extracted statistical parameters of vibration signals in the frequency domains as an input to artificial neural network for gear pitting fault classification. One challenge facing the abovementioned data-driven gear fault detection methods in the era of big data is that features extracted from vibration signals depend greatly on prior knowledge of complicated signal processing and diagnosis expertise. Besides, features are selected per the specific fault detection problems and may not be appropriate for different fault detection problems. An approach that can automatically and effectively self-learns gear fault features from the big vibration data and effectively detect the gear fault is necessary to address the challenge.

As a data-driven approach, sparse coding is a class of unsupervised methods for learning sets of overcomplete bases to represent data efficiently. Unlike principal component analysis (PCA) that learn a complete set of basis vectors efficiently, sparse coding learns an overcomplete basis. This gives sparse coding the advantage of generating basis vectors that are able to better capture structures and patterns inherent in the input data. Recently, sparse coding-based methods have been developed for machinery fault diagnosis [15–18]. However, these methods used manually constructed overcomplete dictionaries that cannot guarantee to match the structures in the analyzed data. Sparse coding with dictionary learning is viewed as an adaptive feature extraction method for machinery fault diagnosis [19]. Study reported in Reference [19] developed a feature extraction approach for machinery fault diagnosis using sparse coding with dictionary learning. In their approach, the dictionary is learned through solving a joint optimization problem alternatively: one for the dictionary and one for the sparse coefficients. One limitation with this approach is that solving the joint optimization problem alternatively for massive data is NP-complete [20] and therefore is not efficient for automation.

In this paper, a new method is proposed. The proposed approach combines the advantages of sparse coding with dictionary learning in feature extraction and the self-learning power of the deep sparse autoencoder for dictionary learning. Autoencoder is an unsupervised machine learning technique and a deep autoencoder is a stacked autoencoder network with multiple hidden layers. To the knowledge of the authors, no attempt to combine sparse coding with dictionary learning and deep sparse autoencoder for gear pitting fault detection has been reported in the literature.

2. The Methodology

The general procedure of the presented method is shown in Figure 1 below. As shown in Figure 1, the presented method is composed of three main steps. Firstly, the dictionary and the corresponding representation of raw data will be obtained through unsupervised learning by the deep sparse autoencoder. Then, a simple backpropagation neural network constructed as the last hidden layer and the output layer is trained to classify the healthy and pitting gear condition using the learnt representation. With the learnt dictionary and trained classifier, the testing raw data are then imported into the network for pitting fault detection. It should be noted that the dictionary learning process is an unsupervised learning process. Thus, the representations regarded as features extracted from raw signals are learnt completely unsupervised without fine-tuning. Sections 2.1 and 2.2 give a brief introduction on dictionary learning and autoencoder, respectively. Dictionary learning using deep sparse autoencoder for gear pitting detection is explained in Section 2.3.
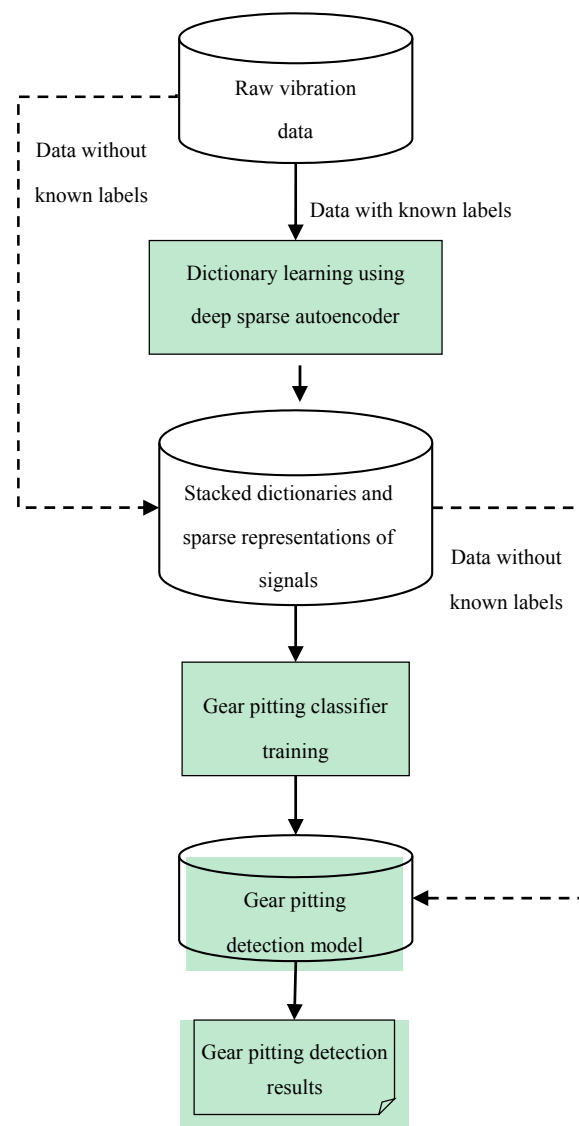
Figure 1. General procedure of the deep sparse autoencoder-based gear pitting detection.

2.1. Dictionary Learning

In recent years, the application of dictionary learning has been popularized in various fields, including image and speech recognition [21–25]. The study of dictionary learning application in vision can be traced back to the end of the last century [26]. The goal of dictionary learning is to learn a basis for representation of the original input data. The expansion of dictionary learning based applications is benefited from the introduction of K-SVD [27,28]. K-SVD is an algorithm that decomposes the training data in matrix form into a dense basis and sparse coefficients. Given an input signal $x = [x_1, x_2, \ldots, x_n]^T$, the basic dictionary learning formula can be expressed as:

$$\min_{D,S} \|x - DS\|_2^2 \tag{1}$$

where $D \in \mathbb{R}^{n \times K}$ represents the dictionary matrix to be learnt with dimension n as number of data points in the input signal x and K as the number of atoms in the dictionary D, each column of D the basic function d_k also known as atoms in dictionary learning, $S = [s_1, s_2, \ldots, s_k]^T$ the representation coefficients of the input signal x, and $\|\cdot\|_2$ the approximation accuracy accessed by the $l^2 norm$.

The goal of dictionary learning is to learn a basis which can represent the samples in a sparse presentation, meaning that S is required to be sparse. Thus, fundamental principle of dictionary learning with sparse representations is expressed as:

$$\min_{S} \|S\|_0 \tag{2}$$

Subject to:

$$\|x - DS\|_2^2 \leq \gamma \tag{3}$$

where function $\|\cdot\|_0$ is referred to as l_0 *norm* that counts the nonzero entries of a vector, as a sparsity measurement, and γ the approximation error tolerance.

As shown in Equation (2), solution of the $l_0 norm$ minimization is a NP hard problem [20]. Thus, the orthogonal matching pursuit (OMP) [29] is commonly used to solve approximation of $l_0 norm$ minimization. As mentioned previously, the popular used dictionary learning algorithm K-SVD was developed with employment of OMP as well. The K-SVD is constituted by two main procedures. In the first procedure, the dictionary matrix is firstly learnt and then it is used in the second procedure to represent the data sparsely. In the procedure of dictionary learning, K-SVD estimate the atoms one at a time according to the ranking update with efficient technique. Such strategy leads to the disadvantage of K-SVD as relatively low computing efficiency since the singular value decomposition (SVD) is required in each iteration.

The basic functions of dictionary matrix D can be either manually extracted or automatically learned from the input data. The manually extracted basic functions are simple and will lead to fast algorithms, however with poor performance on matching the structure in the analyzed data. An adaptive dictionary should be learned from input data through machine learning based methods, such that the basic functions can capture a maximal amount of structures of the data.

2.2. Autoencoder

The structure of autoencoder is shown in Figure 2. A typical autoencoder contains two parts, namely the encoding and decoding part. As shown in Figure 2, the encoding part maps the input data to the latent expression in the hidden layer, and then the decoding part reconstructed the latent expression to the original data as output. In an autoencoder, all the neurons in the input layer are connected to all the neurons in the hidden layer, and vice versa. With a given input data (bias term included) vector x, the latent expression in the hidden layer h can be written as:

$$h = f_e(wx) \tag{4}$$

where w represents the weights matrix between each neuron in the input layer and the one in the hidden layer, f_e the non-linear activation function used to smooth the output of the hidden layer. Commonly, the activation function is selected as sigmoid or tanh function.

The decoding portion reverse the maps the latent expression to the data space as:

$$\hat{x} = f_d\big[w'f_e(wx)\big] \tag{5}$$

where $\hat{x}$ represents the reconstructed data mapped from the latent expression in the hidden layer, $w' = w^T$ the weight matrix between hidden layer and the output layer, and f_d the activation function to smooth the output layer results. Likewise, f_d is usually selected as sigmoid or tanh function.

The objective in the autoencoder training procedure is to obtain the set of encoding weights w and decoding weights w' such that the error between the original input data and the reconstructed data is minimized. The learning objective can be written as:

$$\underset{w,w'}{\mathrm{argmin}} \|x - \hat{x}\|_2^2 \tag{6}$$

The smooth and continuously differentiable activation function in the Equation (5) guarantees that even as a non-convex problem in Equation (6), the smooth results leads it can be solved by gradient descent techniques.

Figure 2. Scheme of single autoencoder.

Furthermore, multiple autoencoders can be stacked to construct a deep structure. The deep autoencoder structure is illustrated in Figure 3.

Figure 3. Scheme of deep autoencoder structure.

For the deep autoencoders shown in Figure 3, the overall cost function can be expressed based on Equation (6) as: $\min\limits_{w_1,\ldots,w_{m-1},w_1',\ldots,w_m'} \|x - \hat{x}\|_2^2$ where $\hat{x} = f_{d1}w_1'\{w_2' \ldots w_m'[E(x)]\}$, and

$$E(x) = f_{e(m-1)}\left\{ w_{m-1} f_{e(m-2)}[w_{m-2} \ldots f_{e1}(w_1 x)] \right\} \tag{7}$$

where w_i and w_i' ($i = 1, 2, \ldots, m$) represent the encoding and decoding weight matrix of the ith autoencoder in the network respectively, f_{ei} and f_{di} the encoding and decoding activation function of

the ith autoencoder. The computational complexity of massive amount of parameters (weight matrix) in Equation (7) results in computation challenge and over fitting phenomena. Thus, searching for the appropriate solution is commonly accomplished through the layer-wise learning behavior.

2.3. Dictionary Learning Using Deep Sparse Autoencoder

Based on the previously reviewed dictionary learning and stacked autoencoder models, a deep sparse autoencoder based dictionary learning is presented in this section. Like the structure of a deep autoencoder, the deep sparse autoencoder based dictionary learning can be illustrated as Figure 4 below.
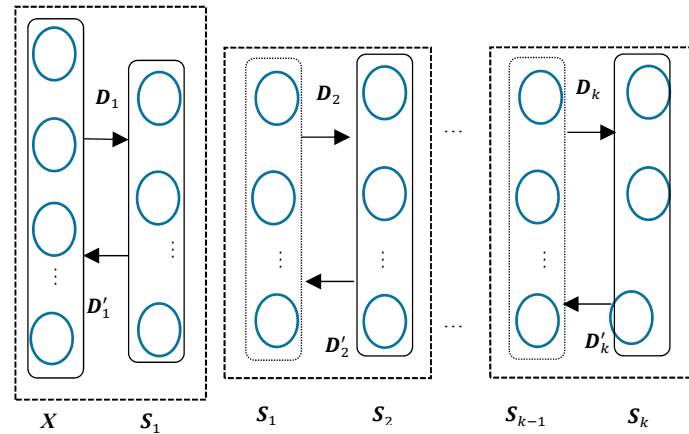
Figure 4. Scheme of stacked dictionary learning.

As shown in Figure 4, each dash block represents a shallow/single dictionary learning process. The first dictionary learning process can be written as:

$$X = D_1 S_1 \tag{8}$$

where $X = \left\{ x_i \in \mathbb{R}^d \right\}_{i=1}^{N}$ stands for the set of N input signals, x_i the signal vector with a length of d, $D_1 = \left\{ d_1^1, d_1^2, \ldots, d_1^j, \ldots, d_1^K \right\}$ for $d_1^j \in \mathbb{R}^d$ the first learnt dictionary, and $S_1 = \left\{ s_1^1, s_1^2, \ldots, s_1^g, \ldots, s_1^N \right\}$ for $s_1^g \in \mathbb{R}^K$ the first latent expression of X in D_1. Treating the deep autoencoder as a dictionary learning network, one can define $D_1' = \left\{ d_1'^1, d_1'^2, \ldots, d_1'^p, \ldots, d_1'^K \right\}$ for $d_1'^p \in \mathbb{R}^d$ as the reconstruction weight from latent expression S_1 to original input X. Like the expression in the autoencoder, the reconstructed input data $\hat{X}$ can be written as:

$$\hat{X} = f_d \left[D_1' f_e(S_1) \right] \tag{9}$$

where f_e and f_d represent the encoding and decoding activation functions, respectively.

Substitute S_1 in Equation (9) with Equation (8), $\hat{X}$ can be written as:

$$\hat{X} = f_d \left\{ \left[D_1' f_e \left(D_1^{-1} X \right) \right] \right\} \tag{10}$$

Here in this study, the activation function for both encoding and decoding processes are selected as sigmoid function. The cost function of dictionary learning using deep sparse autoencoder can be expressed as:

$$\min_{D, D'} \frac{1}{2N} \| X - \hat{X} \|_2^2 + \beta \sum_{j=1}^{J} KL(\rho || \hat{\rho}_j) \tag{11}$$

$$KL(\rho || \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \tag{12}$$

$$\hat{\rho}_j = \frac{1}{N}\left[1 + \exp\left(-\boldsymbol{D}^{-1}\boldsymbol{X}\right)\right]^{-1} \tag{13}$$

where N represents the number of input vectors, β the parameter controlling the weight of the sparsity penalty term, ρ the sparsity parameter, $\hat{\rho}_j$ the average activation of the hidden unit j over the all N training samples. The sparsity penalty term is defined as Kullback-Leibler (KL) divergence, which is used to measure the difference between two distributions. It is defined as $KL(\rho||\hat{\rho}_j) = 0$ when $\rho = \rho_j$, otherwise the KL divergence increases as $|\rho - \rho_j|$ increases. In comparison with the similar k-sparse autoencoder proposed in [30], the advantages of the deep sparse autoencoder include: (1) The introduction of the sparsity penalty leads to the automatic determination of the sparsity rather than pre-defined as k-sparsity. It enables the deep sparse autoencoder to extract the sparse features more accurately based on the characteristics of the data. (2) The dictionary is learnt in the encoding procedure. The encoding dictionary is different from the encoding weight matrix. (3) The deep sparse autoencoders does not require the fine-tuning process while the performance of k-sparse autoencoders relies on the supervised fine-tuning process.

In the deep autoencoder, the output of a hidden layer in the previous autoencoder can be taken as the input to the next autoencoder. Let the first layer of the kth autoencoder in the deep autoencoder be the kth layer and the second layer as the $(k + 1)^{th}$ layer. Also, let $\boldsymbol{D}_k$ and $\boldsymbol{D}'_k$ be the dictionary and reconstruction weight for the kth layer in the deep autoencoder, the encoding procedure in the kth autoencoder can be expressed as:

$$\boldsymbol{a}^k = f_e\left(\boldsymbol{z}^k\right) \tag{14}$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{D}_k^{-1}\boldsymbol{a}^k \tag{15}$$

where $\boldsymbol{a}^k$ stands for the output of the kth layer, $\boldsymbol{z}^k$ and $\boldsymbol{z}^{k+1}$ the input for the kth and $(k + 1)^{th}$ layer, respectively.

Similarly, the decoding procedure in the kth autoencoder can be expressed as:

$$\boldsymbol{a}^k = f_d\left(\boldsymbol{z}^k\right) \tag{16}$$

$$\boldsymbol{z}^{k-1} = \boldsymbol{D}_k'\boldsymbol{a}^k \tag{17}$$

Thus, the original input $\boldsymbol{X}$ can be expressed by the latent expression S_k in the $(k + 1)^{th}$ layer as:

$$\boldsymbol{X} = (\boldsymbol{D}_1\boldsymbol{D}_2\ldots\boldsymbol{D}_k)S_k \tag{18}$$

where $\boldsymbol{D}_k$ represents the learnt dictionary in the kth dictionary learning process, S_k the latent expression of $\boldsymbol{X}$ in $\boldsymbol{D}_k$.

The stacked dictionaries $(\boldsymbol{D}_1\boldsymbol{D}_2\ldots\boldsymbol{D}_k)$ will be learnt in a greedy layer by layer way. The greedy layer by layer learning guarantees the convergence at each layer.

3. Gear Test Experimental Setup and Data Collection

The gear pitting tests were performed on a single stage gearbox installed as an electronically closed transmission test rig. The gearbox test rig includes two 45 kW Siemens servo motors. One of the motors can act as the driving motor while the other can be configured as the load motor. The configuration of the driving mode is flexible. Compared with traditional open loop test rig, the electrically closed test rig is economically more efficient, and can virtually be configured with arbitrary load and speed specifications within the rated power. The overall gearbox test rig, excluding the control system, is showed in Figure 5.
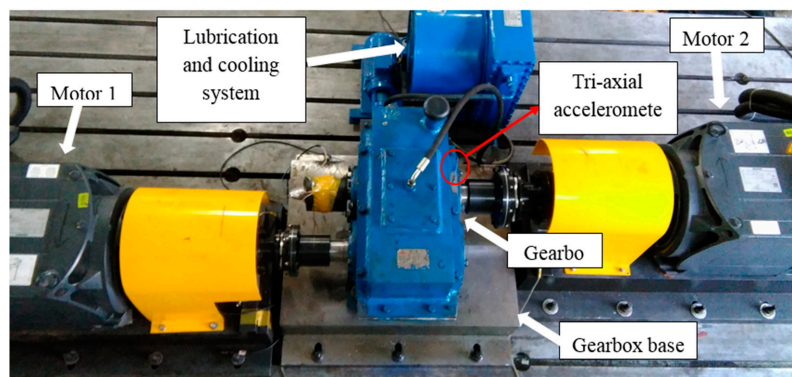
Figure 5. The gearbox test rig.

The testing gearbox is a single stage gearbox with spur gears. The gearbox has a speed reduction of 1.8:1. The input driving gear has 40 teeth and the driven gear has 72 teeth. The 3-D geometric model of the gearbox is shown in Figure 6.
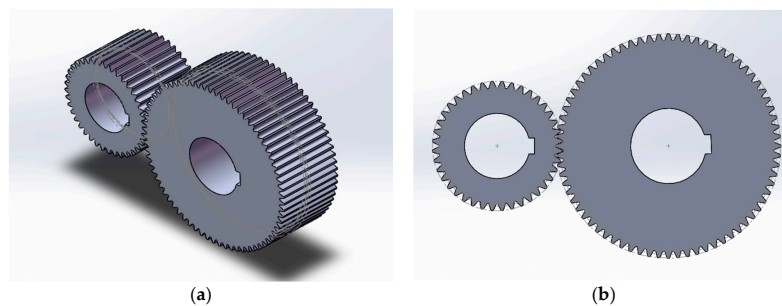

(a)   (b)

Figure 6. 3-D model of the gears under testing. (**a**) gear models in 3-D dimension; (**b**) gear models in 2-D dimension.

Gear parameters are provided in Table 1.

Table 1. List of gear parameters for the tested gearbox.

Gear Parameter	Driving Gear	Driven Gear
Tooth number	40	72
Module	3 mm	3 mm
Base circle diameter	112.763 mm	202.974 mm
Pitch diameter	120 mm	216 mm
Pressure angle	20°	20°
Addendum coefficient	1	1
Coefficient of top clearance	0.25	0.25
Diametric pitch	8.4667	8.4667
Engaged angle	19.7828°	19.7828°
Circular pitch	9.42478 mm	9.42478 mm
Addendum	4.5 mm	3.588 mm
Dedendum	2.25 mm	3.162 mm
Addendum modification coefficient	0.5	0.196
Addendum modification	1.5 mm	0.588 mm
Fillet radius	0.9 mm	0.9 mm
Tooth thickness	5.8043 mm	5.1404 mm
Tooth width	85 mm	85 mm
Theoretical center distance	168 mm	168 mm
Actual center distance	170.002 mm	170.002 mm

The pitting fault was simulated by using electrical discharge machine to erode gear tooth face. The pitting location is on one of the teeth on the output driven gear with 72 teeth. Approximately, the gear tooth face was eroded with a depth of 0.5 mm. One row of pitting faults was created along the tooth width. The simulated pitting fault is shown in Figure 7.
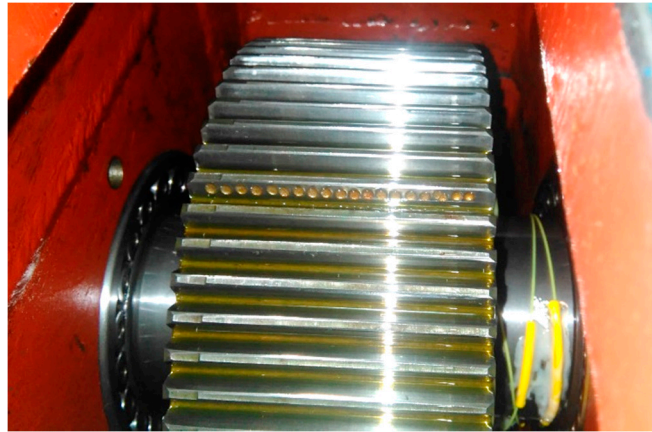

Figure 7. Simulated gear pitting fault.

A tri-axial accelerometer was attached on the gearbox case close to the bearing house on the output end as shown in Figure 8.
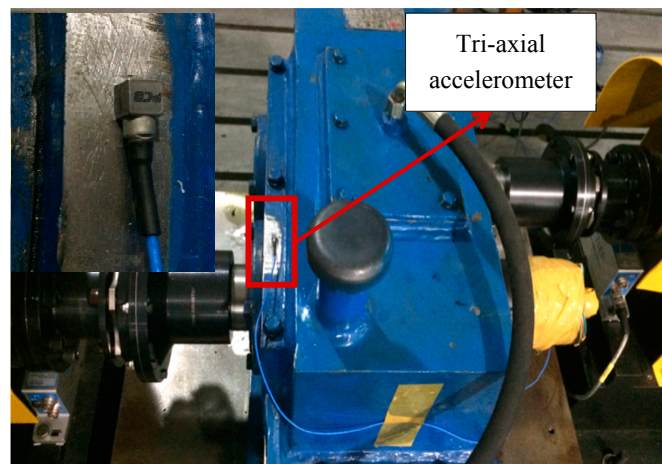
Figure 8. Vibration and torque measurement for the testing gearbox.

Both healthy and pitted gearboxes under various operating condition were run and the vibration signals collected. The tested operation conditions are listed in Table 2. The vibration signals were collected with a sampling rate of 20.48 KHz.

Table 2. Operation condition of the experiments.

Speed (rpm)	100	200	500	1000
Torque (Nm)	50/100/200/300/400/500	50/100/200/300/400/500	50/100/200/300/400/500	50/100/200/300/400/500

Figure 9 shows the raw vibration signals collected for normal gear and pitting gear at loading conditions of 100 Nm and 500 Nm.
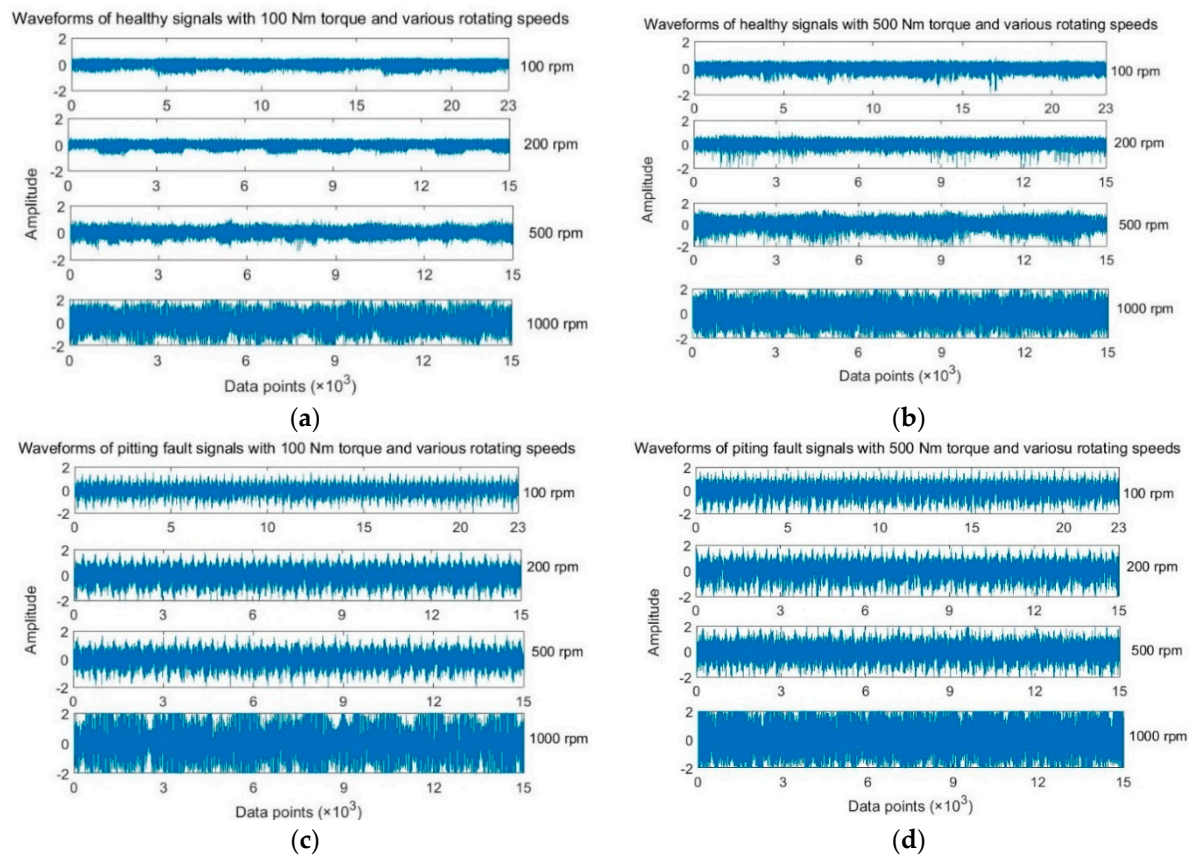
Figure 9. The waveforms of raw vibration data for normal and pitting gear at various rotating speeds under loading conditions of 100 Nm and 500 Nm. (**a**)waveforms of healthy signals with 100 Nm; (**b**) waveforms of healthy signals with 500 Nm; (**c**) waveforms of faulty signals with 100 Nm; (**d**) waveforms of faulty signals with 500 Nm.

4. The Validation Results

The proposed deep sparse autoencoder structure was implemented to accomplish the greedy deep dictionary learning. The vibration signals along the *z* vertical direction were used in this study since they contain the richest vibration information among the three monitored directions. At first, the gear pitting detection was carried on using signals with light loading as the training data and signals with heavy loadings as the testing data, respectively. Loadings of 100 Nm torque and 500 Nm torque were used as light loading condition and heavy loading condition, respectively. Signals at rotating speeds of 100, 200, 500 and 1000 rpm were used for validation tests. To study the influence of different rotating speeds on the pitting gear fault detection performance of the deep sparse autoencoder, 100 and 1000 rpm were selected as low and high speed for independent validations. The length of the samples was decided to ensure that at least one revolution of the output driven gear was included. Therefore, there were 23,000 data points in each sample for signals at 100 rpm and 15,000 data points for signals at a speed of at least 200 rpm. Thus, 26 samples of signals at 100 rpm were generated for healthy gear and pitting gear, respectively. Hence, there were 52 samples in the training dataset and 52 samples in the testing dataset. Similarly, 40 samples of signals at 1000 rpm were generated for each gear condition, with 80 samples in the training dataset and 80 samples in the testing dataset. The structure of the deep sparse autoencoder was designed separately for signals at 100 rpm and signals at the speed of over 100 rpm as: one input layer (23,000 neurons for signals at 100 rpm and 15,000 neurons for signals at the speed over 100 rpm), four hidden layers (1000-500-200-50 neurons), and one output layer (2 neurons). Particularly, following the suggestions in [31], the sparsity parameter in each sparse autoencoder was set as $\beta = 3$ and $\rho = 0.005$. The sparse representations of the original signals were imported into

classifier for pitting gear fault detection. The last hidden layers of 50 neurons and the output layer of 2 neurons were constructed as a simple backpropagation neural network as a classifier for gear pitting detection. The two neurons in the output layer were setup for classifying the input signals as either gear pitting fault or normal gear. The training parameters of the back propagation neural network classifier were set as: training epoch was 100, learning rate was 0.05 and the momentum was 0.05. For each gear condition, the models were executed 5 times to get average detection accuracy. The detection results are shown in Table 3.

Table 3. Detection results at 100 and 1000 rpm (trained with light loading samples and tested with heavy loading samples).

Gear Conditions	Training Accuracy (100 Nm) (100 rpm/1000 rpm)	Testing Accuracy (500 Nm) (100 rpm/1000 rpm)
Healthy gear	100%/99.50%	99.23%/98.84%
Pitting gear	98.43%/100%	98.43%/98.91%
Overall accuracy	99.22%/99.75%	98.83%/98.88%

The detection results in Table 3 show a good adaptive learning performance of the presented method. The testing accuracy is high as 98.88% overall, which is slightly lower than the training accuracy. It can be explained as that signals at light loading condition contain less fault significant information. Furthermore, the same designed deep sparse autoencoder structure was experimented with heavy loading training data and light loading testing data. The detection results are shown in Table 4.

Table 4. Detection results at 100 and 1000 rpm (trained with heavy loading samples and tested with light loading samples).

Gear Conditions	Training Accuracy (500 Nm) (100 rpm/1000 rpm)	Testing Accuracy (100 Nm) (100 rpm/1000 rpm)
Healthy gear	100%/99.95%	100%/99.90%
Pitting gear	100%/100%	99.23%/100%
Overall accuracy	100%/99.98%	99.62%/99.95%

It can be observed from Table 4 that in comparison with results shown in Table 3, the testing accuracy is slightly higher than the training accuracy. The better adaptive feature extraction and fault detection results are benefited from that the signals with heavy loading condition contain more fault significant information. The results in Tables 3 and 4 show marginal influence of the rotating shaft speeds on the pitting fault detection performance.

Moreover, the signals with stable loading and mixed rotating speed were also tested in the study. The detailed description of each dataset used in the validation is provided in Table 5. The detection results are presented in Tables 6 and 7.

Table 5. Dataset description.

Dataset	Loading Condition of the Training Dataset (Nm)	Loading Condition of the Testing Dataset (Nm)	Rotating Speed (rpm)	Length of Signal Sample
A	100	500	100	23,000
B	500	100	100	23,000
C	100	500	1000	15,000
D	500	100	1000	15,000
E	100	500	100/200/500/1000	15,000
F	500	100	100/200/500/1000	15,000

Table 6. Detection results at mixed rotating speeds (trained with light loading samples and tested with heavy loading samples).

Gear Conditions	Training Accuracy (100 Nm)	Testing Accuracy (500 Nm)
Healthy gear	99.45%	97.21%
Pitting gear	99.65%	97.05%
Overall accuracy	99.55%	97.13%

Table 7. Detection results at mixed rotating speeds (trained with heavy loading samples and tested with light loading samples).

Gear Conditions	Training Accuracy (500 Nm)	Testing Accuracy (100 Nm)
Healthy	99.45%	99.94%
Pitting gear	99.58%	99.84%
Over all	99.52%	99.89%

Still, 100 and 500 Nm torque loadings were selected as the light loading and heavy loading condition. For each loading condition, 52 samples and 80 samples were generated for both healthy and pitting gear condition at 100 rpm and at the speeds over 100 rpm, respectively. In comparison with the results in Tables 3 and 4, even though the detection accuracies in Tables 6 and 7 for both cases (trained with light loading samples and tested with heavy loading samples, and vice versa) are slightly lower, the accuracies obtained by the deep sparse autoencoders are satisfactorily as high as 97.13% and 99.89%. The satisfactory detection results show the good performance without the effects of various rotating speeds of the deep sparse autoencoders. Furthermore, it shows the capability of the deep sparse autoencoders in automatically extracting the adaptive features from the raw vibration signals. The validation results have shown the good robustness of the deep sparse autoencoders on gear pitting detection without much influence of working conditions, including loadings and rotating speeds.

To make a comparison, a typical autoencoder based deep neural network (DNN) presented in [32] was selected to detect the gear pitting fault using the same data. The DNN was designed with a similar structure like the deep sparse autoencoder, namely one input layer (23,000 neurons and 15,000 neurons), four hidden layers (1000-500-200-50 neurons), and one output layer (2 neurons). Like the deep sparse autoencoder, the last hidden layer and the output layer of the DNN was designed as a back propagation neural network classifier. Since the autoencoder based neural network normally requires supervised fine-tuning process for better classification, the designed DNN was tested without and with supervised fine-tuning. The detection results of the DNN are provided in Tables 8 and 9.

Table 8. Detection results of DNN at mixed rotating speeds (trained with light loading samples and tested with heavy loading samples).

Gear Conditions	Training Accuracy (100 Nm)		Testing Accuracy (500 Nm)	
	Without Supervised Fine-Tuning	With Supervised Fine-Tuning	Without Supervised Fine-Tuning	With Supervised Fine-Tuning
Healthy gear	85.25%	90.50%	83.42%	89.85%
Pitting gear	85.85%	89.95%	81.15%	88.24%
Overall accuracy	85.55%	90.23%	82.29%	89.05%

Table 9. Detection results of DNN at mixed rotating speeds (trained with heavy loading samples and tested with light loading samples).

Gear Conditions	Training Accuracy (500 Nm)		Testing Accuracy (100 Nm)	
	Without Supervised Fine-Tuning	With Supervised Fine-Tuning	Without Supervised Fine-Tuning	With Supervised Fine-Tuning
Healthy gear	82.18%	90.25%	84.25%	91.50%
Pitting gear	84.15%	88.85%	85.17%	91.50%
Overall accuracy	83.17%	89.55%	84.71%	91.50%

In comparing the results obtained by the DNN in Tables 8 and 9 with those obtained by the deep sparse autoencoder, one can see that the deep sparse autoencoder gives a better performance than the DNN based approach for gear pitting fault detection. In both cases, the detection accuracies obtained by the DNN are much lower than those of the deep sparse autoencoder. In comparison with the DNN, the presented method is more robust in automatically extracting the adaptive features for gear pitting detection. In addition, the presented method does not require the supervised fine-tuning process. Such advantage will increase the computational efficiency and enhance the robustness of the gear pitting fault detection in dealing with massive data.

To verify the ability of the presented method for automatically adaptive features extraction, using a similar approach in [33], the principal component analysis (PCA) was employed to visualize the extracted features. The values of neurons in the last hidden layer were regarded as pitting fault features since they were used for pitting detection in the output layer. Therefore, 50 features were obtained by the deep sparse autoencoders. The first two principle components were used for the visualization since they carried out more than 90% information in the feature domain. Since pitting gear detection results at rotating speed of 1000 rpm is slightly more accurate than that at 100 rpm, only features obtained using signals at 100 rpm were plotted for observation. The scatter plot of principle components of the features automatically extracted from datasets A, B, E and F are presented in Figure 10.
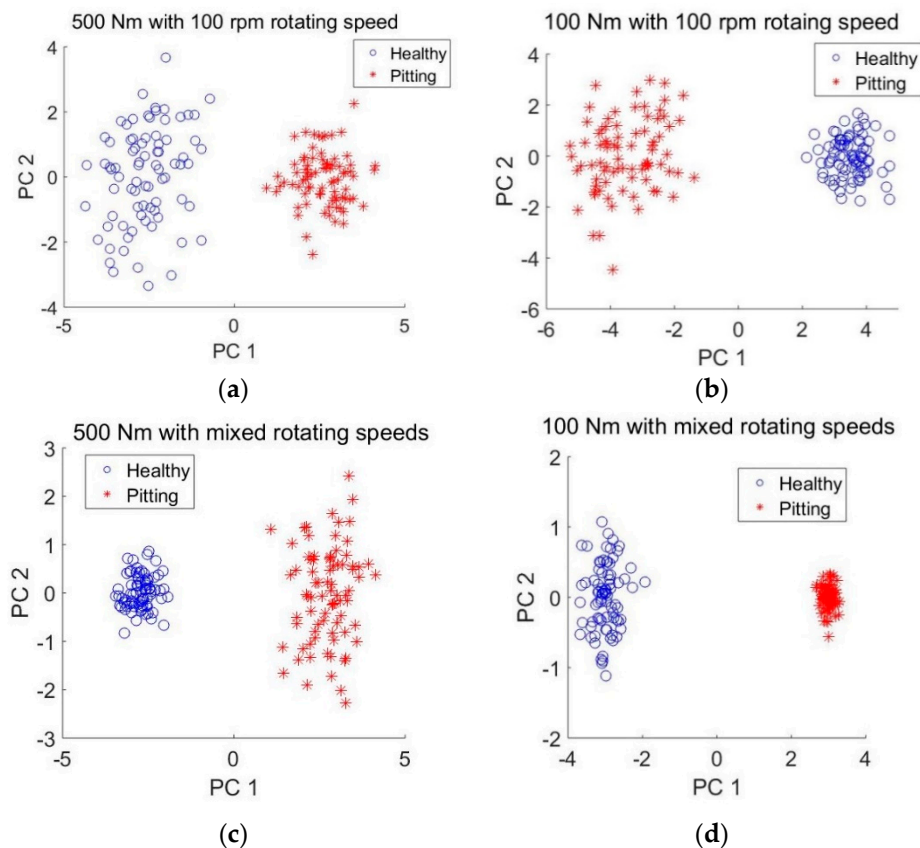

Figure 10. Scatter plot of principle components for the features extracted from: (**a**) dataset A; (**b**) dataset B; (**c**) dataset E, and (**d**) dataset F.

It can be observed from Figure 10 that the features of the same health condition are grouped in the corresponding clusters which are clearly separated from each other. In comparison with Figure 10a,c, Figure 10b,d show a better clustering performance and more clear separation boundary between the healthy and pitting gear conditions. This could be due to the fact that the features of heaving loading conditions in A and E are extracted using the deep sparse autoencoders trained with data of light

loading conditions. The fault features of light conditions are normally less significant than those of heavy loading conditions.

5. Conclusions

Gears are one of the most critical components in many industrial machines and pitting is one of the most common gear faults and normally difficult to detect. An undetected gear pitting fault during the operation of the gears can lead to catastrophic failures of the machines. In this paper, a new method for gear pitting fault detection was presented. The presented method was developed based on a deep sparse autoencoder that integrates dictionary learning in sparse coding into a stacked autoencoder network. The presented method uses a stacked autoencoder network to perform the dictionary learning in sparse coding and automatically extract features from raw vibration data. These features are then used to train a simple backpropagation neural network to perform pitting fault detection. The presented method was validated with vibration data collected from tests with gear pitting faults in a gearbox test rig and compared with a deep neural network based approach. In the validation tests, data obtained from one loading condition was used to train the gear pitting detection model and the model was then tested with data obtained from a different loading condition. The validation results have shown the good robustness of the deep sparse autoencoders on gear pitting detection without much influence of working conditions, including loadings and rotating speeds. The comparison between the deep sparse autoencoder and the deep neural network has shown the outstanding performance of the presented method on automatically extracting the adaptive features than the deep neural network based method.

Author Contributions: Yongzhi Qu conceived, designed, and performed the gear experiments; Miao He and Jason Deutsch analyzed the data; Yongzhi Qu, Miao He, and David He wrote the paper.

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

References

1. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
2. Heng, A.; Zhang, S.; Tan, A.C.C.; Mathew, J. Rotating machinery prognostics: State of the art. Challenges and opportunities. *Mech. Syst. Signal Process.* **2009**, *23*, 724–739. [CrossRef]
3. Rahmounea, C.; Benazzouz, D. Early detection of pitting failure in gears using a spectral kurtosis analysis. *Mech. Ind.* **2012**, *13*, 245–254. [CrossRef]
4. Feki, N.; Cavoret, J.; Ville, F.; Velex, P. Gear tooth pitting modelling and detection based on transmission error measurements. *Eur. J. Comput. Mech.* **2013**, *22*, 106–119.
5. Liu, J.; Wang, G. A multi-step predictor with a variable input pattern for system state forecasting. *Mech. Syst. Signal Process.* **2009**, *23*, 1586–1599. [CrossRef]
6. Lee, S.K.; Shim, J.-S.; Cho, B.-O. Damage detection of a gear with initial pitting using the zoomed phase map of continuous wavelet transform. *Key Eng. Mater.* **2006**, *306–308*, 223–228. [CrossRef]
7. Ozturk, H.; Sabuncu, M.; Yesilyurt, I. Early detection of pitting damage in gears using mean frequency of scalogram. *J. Vib. Control* **2008**, *14*, 469–484. [CrossRef]
8. Ozturk, H.; Yesilyurt, I.; Sabuncu, M. Detection and advancement monitoring of distributed pitting failure in gears. *J. Non-Destruct. Eval.* **2010**, *29*, 63–73. [CrossRef]
9. Lewicki, D.G.; Dempsey, P.J.; Heath, G.F.; Shanthakumaran, P. Gear fault detection effectiveness as applied to tooth surface pitting fatigue damage. In Proceedings of the American Helicopter Society 65th Annual Forum, Grapevine, TX, USA, 27–29 May 2009.
10. Teng, W.; Wang, F.; Zhang, K.; Liu, Y.; Ding, X. Pitting fault detection of a wind turbine gearbox using empirical mode *decomposition*. *Stroj. Vestnik J. Mech. Eng.* **2014**, *60*, 12–20. [CrossRef]
11. He, Q.; Ren, X.; Jiang, G.; Xie, P. A hybrid feature extraction methodology for gear pitting fault detection using motor stator current signal. *Insight Non-Destruct. Test. Cond. Monit.* **2014**, *56*, 326–333. [CrossRef]

12. Peršin, G.; Viintin, J.; Juriic, D. Gear pitting detection based on spectral kurtosis and adaptive denoising filtering. In Proceedings of the 11th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2014/MFPT 2014, Manchester, UK, 10–12 June 2014.

13. Elasha, F.; Ruiz-Carcel, C.; Mba, D.; Kiat, G.; Nze, I.; Yebra, G. Pitting detection in worm gearboxes with vibration analysis. *Eng. Fail. Anal.* **2014**, *23*, 231–241. [CrossRef]

14. Ümütlü, R.; Rafet, C.; Hizarci, B.; Ozturk, H.; Kiral, Z. Pitting detection in a worm gearbox using artificial neural networks. In Proceedings of the INTER-NOISE 2016—45th International Congress and Exposition on Noise Control Engineering: Towards a Quieter Future, Hamburg, Germany, 21–24 August 2016.

15. Liu, B.; Ling, S.F.; Gribonval, R. Bearing failure detection using matching pursuit. *NDT Eval. Int.* **2002**, *35*, 255–262. [CrossRef]

16. Yang, H.; Mathew, J.; Ma, L. Fault diagnosis of rolling element bearings using basis pursuit. *Mech. Syst. Signal Process.* **2005**, *19*, 341–356. [CrossRef]

17. Feng, Z.; Chu, F. Application of atomic decomposition to gear damage detection. *J. Sound Vib.* **2007**, *32*, 138–151. [CrossRef]

18. Zhao, F.; Chen, J.; Dong, G. Application of matching pursuit in fault diagnosis of gears. *J. Shanghai Jiaotong Univ.* **2009**, *43*, 910–913. [CrossRef]

19. Liu, H.; Liu, C.; Huang, Y. Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mech. Syst. Signal Process.* **2011**, *25*, 550–574. [CrossRef]

20. Natarajan, B.K. Sparse approximate solutions to linear systems. *SIAM J. Comput.* **1995**, *24*, 227–234. [CrossRef]

21. Ravishankar, S.; Bresler, Y. MR Image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging* **2010**, *30*, 1028–1041. [CrossRef] [PubMed]

22. Dong, W.; Lin, X.; Zhang, L.; Shi, G. Sparsity-based image denoising via dictionary learning and structural clustering. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011.

23. Yang, M.; Zhang, L.; Feng, X.; Zhang, D. Sparse representation based Fisher discrimination dictionary learning for image classification. *Int. J. Comput. Vis.* **2014**, *109*, 209–232. [CrossRef]

24. Jafari, M.G.; Plumbley, M.D. Fast dictionary learning for sparse representations of speech signals. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 1025–1031. [CrossRef]

25. Sigg, C.D.; Dikk, T.; Buhmann, J.M. Speech enhancement using generative dictionary learning. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 1698–1712. [CrossRef]

26. Olshausen, B.A.; Field, D.J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.* **1997**, *37*, 3311–3325. [CrossRef]

27. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2010**, *54*, 4311–4322. [CrossRef]

28. Rubinstein, R.; Bruckstein, A.M.; Elad, M. Dictionaries for sparse representation modeling. *Proc. IEEE* **2010**, *98*, 1045–1057. [CrossRef]

29. Pati, Y.; Rezaiifar, R.; Krishnaprasad, P. Orthogonal Matching Pursuit: Recursive function approximation with application to wavelet decomposition. In Proceedings of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993.

30. Makhzani, A.; Frey, B. K-Sparse Autoencoders. In Proceedings of the 2nd International Conference on Learning Representations (ICLR2014), Banff, AB, Canada, 14–16 April 2014.

31. Ng, A. *CS 294A Lecture Notes: Sparse Autoencoder*; Stanford University: Palo Alto, CA, USA, 2010.

32. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72–73*, 303–315. [CrossRef]

33. Yunusa-Kaltungo, A.; Sinha, J.K. Sensitivity analysis of higher order coherent spectra in machine faults diagnosis. *Struct. Health Monit.* **2016**, *15*, 555–567. [CrossRef]