*Article*

# Deep Fault Recognizer: An Integrated Model to Denoise and Extract Features for Fault Diagnosis in Rotating Machinery

**Xiaojie Guo [1], Changqing Shen [1,2,\*] and Liang Chen [1]**

[1]  School of Mechanical and Electric Engineering, Soochow University, Suzhou 215131, China;
   gxjcoco@outlook.com (X.G.); chenl@suda.edu.cn (L.C.)

[2]  School of Urban Rail Transportation, Soochow University, Suzhou 215131, China

\*  Correspondence: cqshen@suda.edu.cn; Tel.: +86-512-6716-2703

**Abstract:** Fault diagnosis in rotating machinery is significant to avoid serious accidents; thus, an accurate and timely diagnosis method is necessary. With the breakthrough in deep learning algorithm, some intelligent methods, such as deep belief network (DBN) and deep convolution neural network (DCNN), have been developed with satisfactory performances to conduct machinery fault diagnosis. However, only a few of these methods consider properly dealing with noises that exist in practical situations and the denoising methods are in need of extensive professional experiences. Accordingly, rethinking the fault diagnosis method based on deep architectures is essential. Hence, this study proposes an automatic denoising and feature extraction method that inherently considers spatial and temporal correlations. In this study, an integrated deep fault recognizer model based on the stacked denoising autoencoder (SDAE) is applied to both denoise random noises in the raw signals and represent fault features in fault pattern diagnosis for both bearing rolling fault and gearbox fault, and trained in a greedy layer-wise fashion. Finally, the experimental validation demonstrates that the proposed method has better diagnosis accuracy than DBN, particularly in the existing situation of noises with superiority of approximately 7% in fault diagnosis accuracy.

**Keywords:** fault diagnosis; deep learning; stacked denoising autoencoder; feature extraction; integrated deep fault recognizer

## 1. Introduction

Gears and bearings are important components that are commonly used in the rotating machinery parts of trains, ships, and automobile manufacturing, among others. However, gears and bearing are prone to breakdown. Therefore, an unexpected failure may lead to serious project accidents, large economic losses, and even human casualties [1].

Detecting and diagnosing fault to enhance the safety and reliability of machinery, as well as reduce operation and maintenance costs, are essential and have practical significant because of the effect of unexpected accidents [2]. Vibration signals can accurately indicate the health conditions of mechanical equipment; hence, these signals are extensively used in fault diagnosis based on artificial methods, such as multinomial logistic regression, wavelet packet transforms (WPT), and support vector machines (SVMs) [3–6]. Yuan et al. [7] selected kurtosis and entropies of the signals as the feature of the input, and put these into the neural network to do fault diagnosis. This work showed that kurtosis and entropies are useful and unique features to classify faults. Jiang et al. [8] improved SVM, which is included in the fault dictionary category, and proposed a novel approach to diagnose actual analog circuits. Ahcène et al. [9] used wavelet-packet method to generalize wavelet decomposition for signal

analysis. The research showed that the wavelet decomposition was a satisfactory method for analyzing motor faults over load torque or non-stationary signals. Lei et al. [10] proposed complete ensemble empirical mode decomposition with adaptive noise into application of fault diagnosis in rolling element bearings, where a unique residue was computed after each IMF extraction and used as feature to do diagnosis. Wang et al. [11] introduced a Bayesian approach based on a linked posterior probability density function of wavelet parameters, and [12] combined a novel approach to the Gauss–Hermite integration on Bayesian theory to estimate the posterior distribution of wavelet parameters. These signal processing methods based on crucial math analysis are beneficial in feature extraction of fault signals. Shen et al. [13] presented a model that used empirical mode decomposition (EMD) to select silent features and put these into multi-class transductive SVM (TSVM), thereby obtaining an accuracy of 91.62% in diagnosing the faults of a gear reducer. Feng et al. [14] proposed a method called Teager energy spectrum to extract the fault induced impulses as features to conduct bearing fault diagnosis, as well as proved the superiority of this method in recognizing transient components in signals and in identifying the characteristic frequency of bearing faults. Cai et al. [15] introduced a high order spectrum to reconstruct the signals' power spectrum, and used it to extract fault feature information. This novel method proved to extract more useful feature information than others. The above methods have obtained satisfactory performance. However, some of the methods select features according to the signals statistical properties, such as kurtosis and variance. Sometimes, the researchers also rely on the knowledge about faults, such as the fault characteristic frequency, and extract the related frequency band energy as features, and then establish the relationship between the feature vectors and their labels (fault type or healthy condition). Moreover, the performance of the traditional methods relies heavily on the representability of the features which are usually manually selected.

Nevertheless, the effective diagnosis of the machinery health status based on vibration signals remains a challenge when machinery systems are considerably complex, and sometimes the decision process requires some expertise and signal processing techniques [16]. Hence, automatically extracting failure features from machine signals without human interference is significant [17]. Up to date, intelligent methods (e.g., BPNN) have been extensively investigated and used to diagnose faults in rotating machinery [18,19]. Liang et al. [17] combined BPNN with wavelet packet decomposition, where the wavelet packet decomposition coefficients were used to extract eight energy features, and the BPNN was used to do recognition with validation accuracy at 92.5%. Recently, deep-learning methods, such as the DBN–WPT combination, have been determined to overcome the obstacles in fault diagnosis in complex machines [20–22]. Deep architectures have been proven to be more effective fault recognition than shallow architecture. However, in these methods, features still need to be selected manually at first and deep models just function as classifiers.

The early vibration signals of gear or bearing fault are often characterized as non-stationary and are well-affected by vibrations from other components in the equipment and transmission path [23]. This way, the beneficial information of signals is often restricted by other noises or even completely overwhelmed. Obtaining useful information from a signal polluted by noise is essential for effective fault diagnosis methods. However, inadequate denoising or over denoising can distort the original signals, thereby resulting in useless machinery fault diagnosis or reduced recognition rate. Therefore, an efficient denoising technique is required before analyzing the signals for the characteristic fault frequency retrieval. If the components of the signal are known, then optimal filters can be employed for denoising [24]. Moreover, numerous methods have been developed and applied to the denoising step, such as wavelet transform (WT), ensemble empirical mode decomposition (EEMD), and undecimated discrete wavelet transform (UDWT), among others. Tan et al. [25] denoised signals with digital wavelet frame (DWF) and performed fault diagnosis based on a stacked autoencoder (SAE). It combined low-level features to form more abstract high-level features to represent data distributed characteristics and obtained accuracy around 99%. Santhana et al. [26] also proposed a method that combined UDWT and EMD to complete the denoising and diagnosis progress. The UDWT was used to denoise signals and EMD was used to decompose the signals into a number of Intrinsic Mode Functions (IMFs).

Even though EMD is an adaptive signal processing method, it suffers from several shortcomings, such as mode-mixing problem that makes analyses of IMFs difficult and empirical. Some fault related signatures may reside in several IMFs, which make the selection/combination of useful IMFs for machine fault feature extraction difficult. Besides, some combinations of the denoising step with the feature extraction and classification step bring more human intervention, such as the base function or other parameters selection affecting the final performance.

To determine a concise and efficient method that can simultaneously denoise the signals while extracting features, a stacked denoising autoencoder (SDAE) based on a deep network is proposed to construct a fault diagnosis system. SDAE, which was first proposed by Vincent et al., is a stacked neural network comprising several classical denoising autoencoders (DAs) that are trained not to construct their input but rather denoise an artificially corrupted version of their input [27]. DAs have been previously shown to be competitive to restricted Boltzmann machines, constructive unit of a DBN, and for the unsupervised pre-training of each layer [28]. Vincent et al. [29,30] extended DA with a greedy layer-wise procedure of deep learning algorithm. Pierre et al. [31] introduced a general mathematical framework for the study of both linear and non-linear autoencoders, thereby enabling the autoencoder to solve additional types of tasks. Thereafter, SDAE has been extensively used in different types of application [32–36]. In the field of fault diagnosis, Lin et al. [37] placed a stacked autoencoder (SAE) into fault diagnosis. However, with an additional independent component analysis (ICA), the diagnosis system became unintelligent and complicated. Natarajan et al. [38] used DWT to select features as well as denoise the signals in the first step, and put the features into artificial neural network (ANN) to do classification. Wavelet Daubechies-6 was selected for processing 49 signal samples per condition in their work. It is important to select a proper wavelet mother function which should be highly similar to impulses generated by a bearing or gear defect, and to determine a wavelet decomposition level/depth for retaining the resonant frequency band excited by the defect. In general, the performance of DWT processing relies on the selection of wavelet mother function and the depth of wavelet decomposition to obtain the representative features. On the other hand, a proper classifier construction is also needed for the extracted DWT features. The current study proposes a deep-learning-based fault diagnosis method. Accordingly, a deep fault recognizer based on SDAE is used to deal with the random noises of the original signals and extract features to simultaneously perform fault pattern recognition. The SDAE is an integral model, in which the weights of the feature selection part are updated based on the final output every iteration. The proposed method has been empirically shown to avoid obtaining poor convergence, which is typically reached with random initialization [39]. In addition, the experiment manifests that the SDAE method obtains superior diagnosis performance compared to the DBN methods, particularly in the existing situation of noises.

The remainder of this paper is organized as follows. Section 2 details the methodology of the traditional SDAE. Section 3 provides details of the structure and the algorithm of the proposed SDAE based diagnosis system. Section 4 validates the effectiveness of the SDAE method on rolling bearing datasets and gearbox dataset. Moreover, this section further tested the advantage of the proposed method through a comparative study between DBN and SDAE on both original datasets, as well as the dataset mixed with artificially white noises. Finally, Section 5 presents the conclusion.

## 2. Brief Introductions to SDAE

DA (denoised autoencoder) is a simple modification of the classical autoencoder neural networks that are trained to denoise an artificially corrupted version of their input. Meanwhile, SDAE often comprises several DAs that are trained in a bottom-up and layer-wise manner. In SDAE, DAs use corrupted versions when several signal elements are missing. The hidden layers of a traditional SDAE simultaneously aid in dealing with blank noises and extracting the features.

### 2.1. Denoising Autoencoder

Given a training input pattern $X$, a corrupted version $\widetilde{X}$ is first generated. The corrupted $\widetilde{X}$ is generated in a way of random mapping: $\widetilde{X} \sim p(\widetilde{X}|X)$, where refers to a conditional distribution. For the

input matrix $X$, a determinate number $d$ of points are selected randomly, the value of which is forced to 0, whereas the others remained unchanged.

Thereafter, the corrupted input $\widetilde{X}$ is projected to the succeeding hidden layer:

$$Y = f_\theta(\widetilde{X}) = \text{sigmoid}(W * \widetilde{X} + b), \tag{1}$$

$$\theta = \{W, b\}, \tag{2}$$

where $\theta$ is the weight and $b$ is the bias parameters from the input to the hidden layers. Thereafter, the input is reconstructed as:

$$Z = g_{\theta'}(y) = \text{sigmoid}(W' * Y + b'), \tag{3}$$

$$\theta' = \{W', b'\}, \tag{4}$$

where $\theta'$ is the weight and $b$ is the bias parameters from the hidden layer to the output layer. To minimize the number of parameters, tied weights are used as follows:

$$W' = W^T. \tag{5}$$

The joint function is defined as follows:

$$p^0(X, \widetilde{X}, Y) = p^0(X) p_d(\widetilde{X} \mid X) \sigma_{f_\theta(\widetilde{X})}(Y), \tag{6}$$

The computation $\sigma_{f_\theta(\widetilde{X})}(Y)$ indicates placing mass 0 if $Y \neq f_\theta(\widetilde{X})$, and the parameters of $p^0(X, \widetilde{X}, Y)$ is determined by $\theta$. Thereafter, the loss function to be minimized can be defined as follows:

$$\underset{\theta, \theta'}{\text{argmin}} E_{p^0(X, \widetilde{X})} [L_H(X, Z)]. \tag{7}$$

The cross-entropy loss is computed as follows:

$$L_H(X, Z) = X \log Z + (1 - X) \log(1 - Z). \tag{8}$$

The gradient of the loss is computed to update the parameters as follows:

$$[\theta, \theta']^{k+1} = [\theta, \theta']^k + \alpha \frac{\partial L_H(X, Z)}{\partial [\theta, \theta']^k}, \tag{9}$$

where $\alpha$ denotes the learning rate of the training process. The process of encoding and decoding is illustrated in Figure 1.
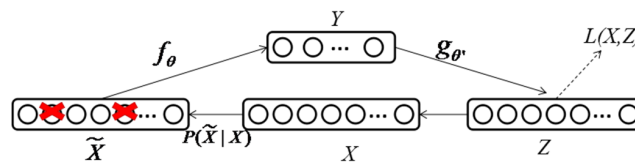


**Figure 1.** Encoding and decoding process of denoising autoencoder.

After pre-training, the autoencoder is robust to the particular corrupted input and other signal deficiencies; thus, the denoising autoencoder is able to "fill in" these "blanks" of input. When the dimension of $Y$ is constrained shorter than that of input $X$, the intermediate representation may be served as various coordinate system transformations for points. Denoising training can only be possible because of dependencies between dimensions in high dimensional distributions, thereby

possibly determining these dependencies. This process also explains why this approach is appropriate in fault diagnosis, which has high dimension input signals.

## 2.2. SDAE

For SDAE, a deep network is initialized by stacking several denoising autoencoders. The corrupted input is only used in the process of initial denoising training for each layer to obtain beneficial features from the signals mixed with noises. Once the initial denoising layer has been trained, this layer will be used on the original inputs thereafter. Accordingly, no corrupted input is applied to generate the presentation, thereby serving as clean input for the training of the next layer. Thus, output *Y* is inputted into the next layer without considering output *Z*. The complete procedure of stacking several layers of denoising autoencoder is depicted in Figure 2.
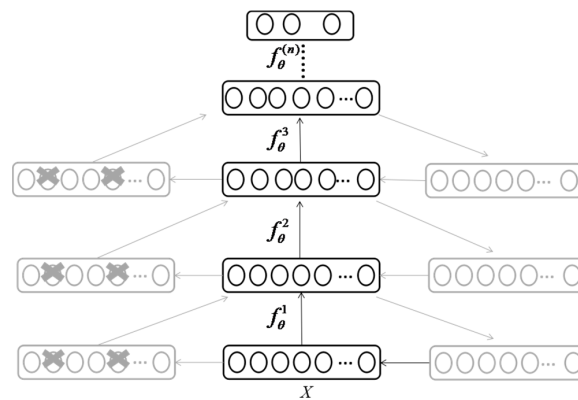


**Figure 2.** Process of SDAE training.

## 3. Proposed Integrated Deep Fault Recognizer

### 3.1. Denoising Part of Deep Fault Recognizer

There is a problem that the ambient noise mixed within the dynamic vibration signals cannot be ignored. The SDAE method of Vincent et al. [27] focused on masking noise, which is a common technique for handling missing values. However, the main type of noise in the practical situation of fault diagnosis of a machine is random noise, which is different from masking noises. Thus, a new diagnosis recognizer is applied based on the random corrupted mechanism.

In the process of generating a corrupted version $\widetilde{X}$, $\widetilde{X}$ is now computed as follows:

$$\widetilde{X} = X + R, \tag{10}$$

where *R* is a random signal added to the input *X*. The corruption part can be represented by capturing the statistical dependencies among the input. The degree of denoising function is controlled by corruption level *p*, which is a variance of *R*:

$$p^2 = \mathrm{Var}(R) \tag{11}$$

Thereafter, after $\widetilde{X}$ is obtained, the following steps of training and stacking several autoencoders together are similar to the previous description in Section 2. It should be noted that the signal corruption process is conducted in all layers of the SDAE model instead of the input layer only.

### 3.2. Structure and Algorithm of Deep Fault Recognizer

This study uses the SDAE based model as basis to propose a self-governed intelligent fault diagnosis method that can automatically mine fault features from original signals, classify different

fault conditions, and deal with the noises existing in raw signals. The original signals in this study refer to the measured data in the frequency domain, which is an advantage in presenting the distribution of constitutive components with discrete frequencies, as well as containing legible information about the health conditions of rotating machinery [40]. The successive flow chat of deep fault recognizer is presented in Figure 3.
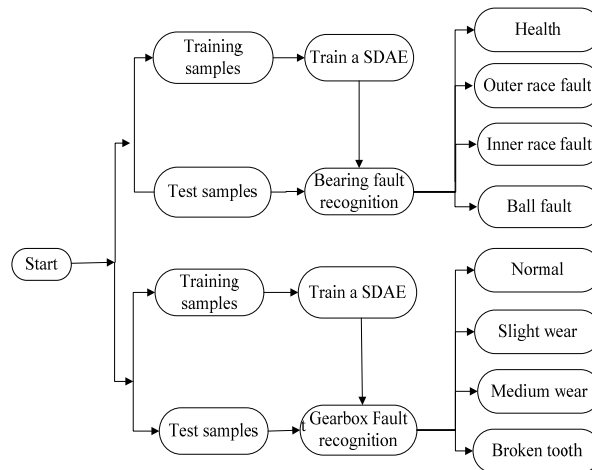


**Figure 3.** Flow chart of deep fault recognizer.

With the training set of the frequency spectra [$X$, $Y$], several DAs can be trained and stacked. After a stack of denoising DAs has been built, the highest level representation output is inputted to a stand-alone supervised training algorithm, such as a logistic regression layer (see Figure 4). This process yields a deep neural network to supervised learning. Figure 4 illustrated the structure of an SDAE based model, the length of each layer and number of layers of which will be different because of different tasks. To date, no validated method to immediately determine these parameters has been developed. Therefore, the grid search method was adopted to determine a suitable model structure based on the machine learning empirical analysis.
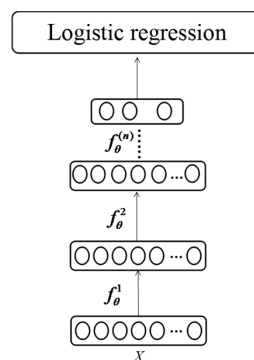


**Figure 4.** Architecture of SDAE.

For a neural network, training the model with gradient-based optimization technique as a back propagation method in a supervised manner is straightforward. However, to resolve the gradient missing problem occurring in a deep network, the greedy layer-wise learning algorithm is adopted to train this deep network. First, the network is pre-trained per layer in a bottom-up direction. The parameters are updated by a set of batches, and each batch comprises several samples. Thereafter, the parameters of the entire model can be fine-tuned with the back propagation method in a top-down direction.

In the pre-training process, SDAE can learn multiple nonlinear deformations and extract the main beneficial inherent features of the frequency spectra. Thereafter, the process of fine-tuning enables SDAE to obtain different information from the input to further promote the classification ability [41]. The whole training process of SDAE is described as Algorithm 1.

---

**Algorithm 1. Training deep fault recognizer**

---

Given training samples {*X*, *Y*}, the number of data batch m, the number of hidden layers *l*,
Step 1: Pre-train the SDAE
For *j* in *l*:
For *i* in *m*:

　　　　　　–Random Initialization for weight matrices and bias vectors {*W*, *b*}.
　　　　　　–Compute the output of the *j*th hidden layer and make it as the input of the (*j* + 1)th hidden layer.
　　　　　　–Update encoding parameters {*W*, *b*} with the gradients computed from the loss function for the (*j* + 1)th hidden layer to minimize the loss function.

---

Step 2: Fine-tune the whole network
–Initial {*W*, *b*} with the pre-training result.
For *i* in *m*:

　　　　　　–Use the BP method to compute the overall gradient of each batch and the optimization with the gradient-based optimization technique.
　　　　　　–Fine-tune the {*W*, *b*} with the mean optimization of a batch by updating all the parameters of the network in a top-down direction.

---

## 4. Experiments

### 4.1. Fault Diagnosis of Rolling Bearings

#### 4.1.1. Dataset Description

The analyzed experimental dataset was gathered from the Case Western Reserve University [42]. Figure 5 shows a test rig that generates the experimental data. The experimental platform comprises a torque transducer, a 2 hp motor (left), and a dynamometer (right).The type of bearings that support the motor shaft was 6205-2RS JEM SKF. Electro-discharge machining is added to cause a single point fault on the inner race of the bearing, ball elements, and outer race. Accelerometers were installed at the dead end to sample vibration signals at 12 kHz for the faults localized on the inner race, rolling elements, and outer race. The data samples obtained from the different health conditions, including four conditions, are employed for analysis. Table 1 shows the two sets that have been analyzed. The defect size of the first dataset was 0.007 with no loading, and the second dataset was obtained with a defect size of 0.014 under 1 hp loading.

**Table 1.** Description of the bearing fault dataset.

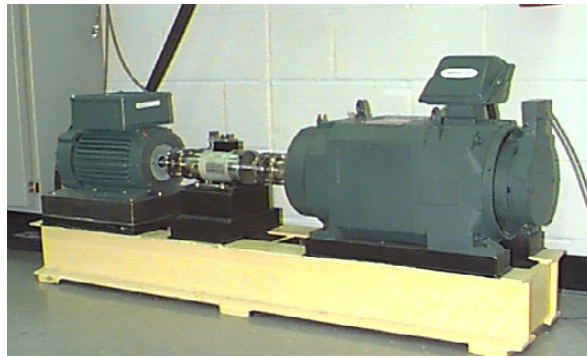| Dataset | Health Condition | Loading (hp) | Defect Size (Inches) | Training Samples | Testing Samples | Target Value |
|---|---|---|---|---|---|---|
| Bearing Dataset I | Health | 0 | - | 50 | 50 | 1 |
| | Inner race fault | 0 | 0.007 | 50 | 50 | 2 |
| | Ball fault | 0 | 0.007 | 50 | 50 | 3 |
| | Outer race fault | 0 | 0.007 | 50 | 50 | 4 |
| Bearing Dataset II | Health | 1 | - | 50 | 50 | 1 |
| | Inner race fault | 1 | 0.014 | 50 | 50 | 2 |
| | Ball fault | 1 | 0.014 | 50 | 50 | 3 |
| | Outer race fault | 1 | 0.014 | 50 | 50 | 4 |

**Figure 5.** Description of the experimental platform.

4.1.2. Results

The rolling bearing signals were processed in accordance with the structure of the SDAE fault diagnosis method shown in Figure 4. For each dataset, 200 samples were randomly selected and inputted into the system for training the model with output of four target values. After optimization of the parameters in the model, another 200 randomly selected samples were tested for the fault pattern recognition.

Figure 6 shows the example signal waveforms and their spectra for the validated four health conditions. Each signal contains 1024 data points, and each sample has its frequency spectrum with 513 Fourier coefficients.
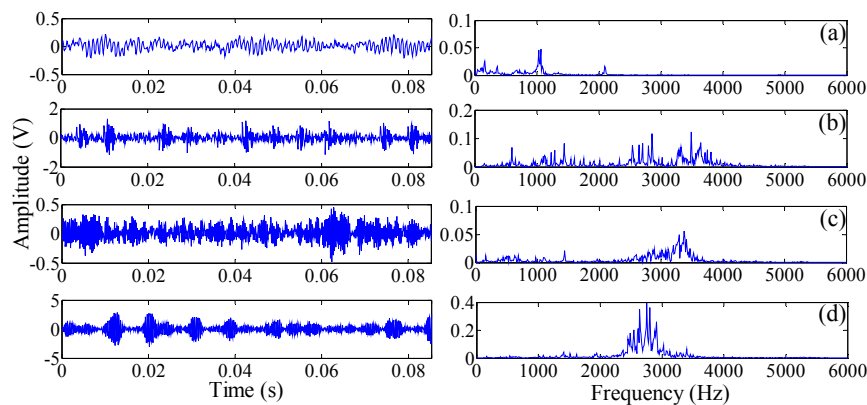


**Figure 6.** Example of the original signals and their spectra for each health condition (**a**) normal-bearing signal; (**b**) inner race fault-bearing signal; (**c**) ball fault-bearing signal; and (**d**) outer race fault-bearing signal.

To determine the structure of a SDAE network, several parameters need to be selected, such as the length of the first visible layer, number of units in each hidden layer, and number of hidden layers. 1024 data unit to are used to form a sample not only to attempt to minimize the computing complexity but also to cover all the information representing vibration signals; thus, the size of the frequency input is 513. The complexity of the model structure is highly related to the recognition task, and it has a great influence on the recognition result. Too complex structure will lead to over-fitting, while too simple structure can result in under-fitting. The hidden layers are selected from 1 to 3 according to the number of samples, length of a sample, and the complexity of the task. The number of units in each layer is selected in the range of (300, 400, 500, 600). To obtain the best architecture for this task, a grid researching is performed. Accordingly, the selected architecture is (500, 500, 500) with three hidden layers. The learning rates for pre-training and fine-tuning are 0.8 and 0.001, respectively. The corruption level for this dataset is designed to be 0.6, which is the optimal value from 0 to 1.

Figure 7 shows the testing results on the health of the bearings, inner race, ball faults, and outer race pattern recognition for dataset I with the best structure selected. The recognition accuracy of the training and testing sets are 100%, and the output value has a perfect clustering result as well. Figure 8 shows the testing and training results of dataset II, which still obtained 100% classification result although with slight floating.
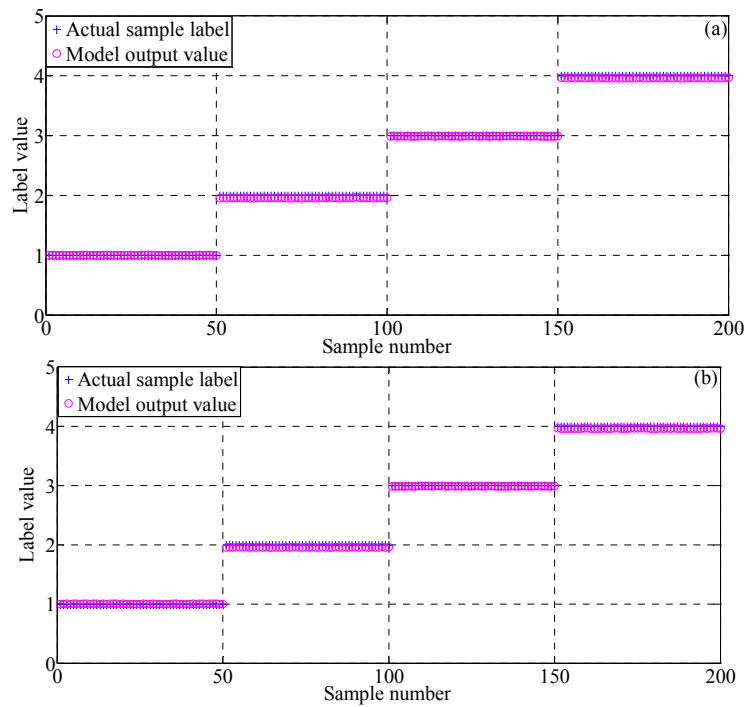


**Figure 7.** (**a**) Training results and (**b**) testing results for dataset I by SDAE.
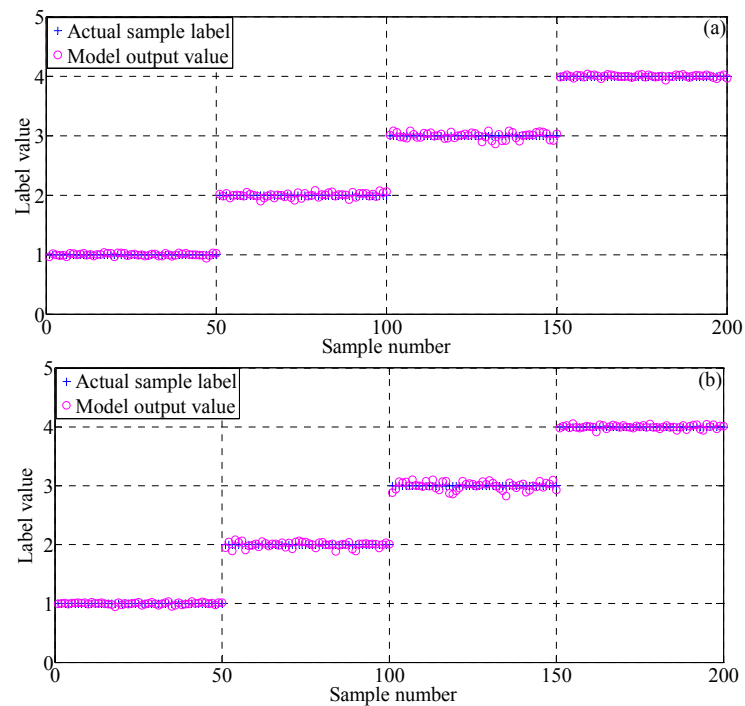


**Figure 8.** (**a**) Training results and (**b**) testing results for dataset II by SDAE.

A DBN model was used for comparison to extrude the superiority of SDAE as a dependent component and its simplicity in the recognition system. DBN is a deep neural network comprising multiple restricted Boltzmann machine (RBM) layers, and has been applied to fault diagnosis since 2014. Each RBM comprises visible and hidden layers, where the units in the same layer are completely unconnected. DBN has been extensively used and obtained satisfactory performance in fault diagnosis since the breakthrough in deep learning.

The comparison is conducted based on the same two datasets of rolling bearing fault. The structure and parameters of the training process are similar with SDAE. The training and testing set recognition accuracy is 100%. However, the label value for the testing set is floating around the actual label compared with the proposed method. One sample is misjudged in both training and testing sets with an overall accuracy of 99.50%, and the output label value has a considerably evident floating. Table 2 generalizes the recognition accuracy of the SDAE method and DBN comparison method. The proposed method achieves satisfactory performance regarding the classification accuracy and the label value clustering performance.

**Table 2.** Identification results of bearing fault obtained by SDAE and DBN.

| Dataset | Method | IR Fault Accuracy (%) | | B Fault Accuracy (%) | | OR Fault Accuracy (%) | | Health Accuracy (%) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Training | Testing | Training | Testing | Training | Testing | Training | Testing |
| Bearing Dataset I | SDAE | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | DBN | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Bearing Dataset II | SDAE | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | DBN | 100 | 100 | 100 | 100 | 98 | 98 | 100 | 100 |

With the same CWRU data set in rolling bearing test, some other classical methods such as SVRM, DWF and LDA, proposed in classical literature [43,44], also have good performance. To further highlight the superiority of the proposed method, comparisons with these methods are shown in Table 3. In [43], WPT was used to transform the signals at different decomposition depths and DET was employed to reduce the dimension of features. Then SVRM was proposed to classify the fault patterns. In [25], sparse coding was introduced to select feature from the vibration signals and linear discriminate analysis (LDA) was used for fault classification. In [44], the recognition model was based on SAE, and a combination of digital wavelet frame (DWF) and nonlinear soft threshold method was carried out to denoise the signals.

For the fault diagnosis task, the SDAE method could achieve better recognition result compared to these methods with the same benchmark. Compared with SVRM, which needs select features manually, the SDAE can automatically extract typical features. For LDA and DWF methods, the integral model of SDAE can achieve better fitting performance.

**Table 3.** Comparison of different Methods in CWRU data set.

| Fault Size | Method | Recognition Accuracy | |
|---|---|---|---|
| | | Training (%) | Testing (%) |
| 0.007 | SDAE | 100 | 100 |
| | SVRM | 100 | 97.91 |
| | LDA | 100 | 99.57 |
| | DWF + SAE | 100 | 99.25 (1 hp load) |
| 0.014 | SDAE | 100 | 100 |
| | SVRM | 100 | 98.33 |
| | LDA | 99.02 | 99.02 |
| | DWF + SAE | 100 | 99.62 (2 hp load) |

*4.2. Fault Diagnosis of Gearbox*

4.2.1. Data Description

To further validate the proposed deep recognizer model in the fault diagnosis, this study took the dataset obtained from an automobile transmission gearbox as the second case analysis. Figure 9 shows the gearbox being referred to, with one speed backward and five other speeds forward. An accelerometer is installed at the outer box of the gearbox to obtain vibration signals during the forward motion of the third pair of gears. Table 4 details the parameters of the third-speed gears, which was the only one used for the wear process testing. The sampling frequency was 3000 Hz, and the input rotating speed was setup to 1600 rpm. After the gears experienced four running cycles during the running (see Table 5), a broken tooth fault occurred at the driving gear in the beginning of the fifth cycle.

**Table 4.** Specification of the third-speed gears.

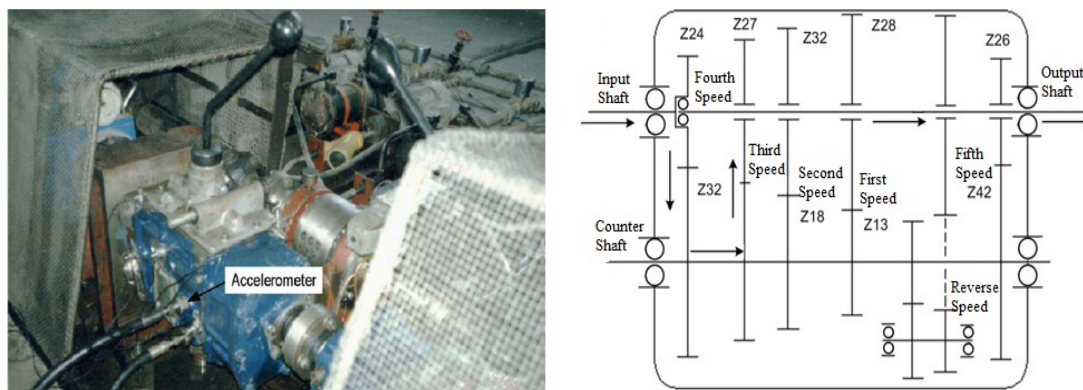| Gear | Number of Teeth | Rotating Frequency (Hz) | Meshing Frequency (Hz) |
|------|-----------------|-------------------------|------------------------|
| Driving gear | 25 | 20 | 500 |
| Driven gear | 27 | 18.5 | 500 |



**Figure 9.** Collection platform of the gearbox vibration signal.

**Table 5.** Gearbox health status and its running period.

| Running Cycle | Health Status | Meshing Times (Thousand) |
|---------------|---------------|--------------------------|
| 1 | Running-in | 0–700 |
| 2 | Normal wear | 700–2800 |
| 3 | Slight wear | 2800–5600 |
| 4 | Medium wear | 5600–6300 |
| 5 | Broken tooth | 6300–7000 |

Table 6 shows that each of the four health statuses has 60 samples selected randomly, in which the half part is used for training and the other half for testing. During the training process of the SDAE model, the slight wear, medium wear, broken tooth, and normal wear labels are set as 1, 2, 3, and 4, respectively.

**Table 6.** Description of the gearbox datasets.

| Health Status | Training Samples | Testing Samples | Target Value |
|---------------|------------------|-----------------|--------------|
| Slight wear | 30 | 30 | 1 |
| Medium wear | 30 | 30 | 2 |
| Broken tooth | 30 | 30 | 3 |
| Normal wear | 30 | 30 | 4 |

4.2.2. Results

The gearbox datasets were still processed as the structure of the SDAE fault diagnosis scheme (see Figure 3). A total of 120 samples are inputted into the system for training the model, with the outputting of four target values. After the optimization of the parameters in the model, another 120 samples are tested for the fault pattern recognition. Figure 10 shows the input signals and their frequency spectra. Each sample contains 1000 data points and the frequency spectra of the sample have 501 Fourier coefficients.

The SDAE architecture is required to be determined again due to the change of objects. The number of data units in a gear sample is 1000 and the spectra length is 501; thus, the size of input is 501. After the grid search runs, the selected architecture remains (500, 500, 500). For this dataset, the parameters and structure are similar with the bearing fault dataset, thereby demonstrating the generalization ability of the proposed method to the different datasets.
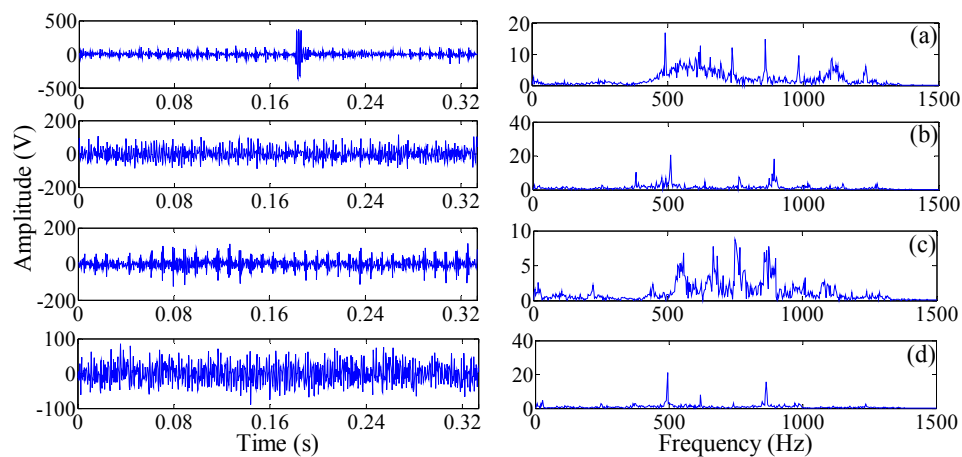


**Figure 10.** Example signals and their spectra of four health status patterns: (**a**) slight wear; (**b**) medium wear; (**c**) broken tooth; and (**d**) normal wear signals.

After selecting the appropriate structure, the training samples are inputted to train an SDAE network. The classification result can be viewed in detail after the testing process. Figure 11 shows the testing results of the four gear pattern recognition for dataset II. The training and testing samples are correctly recognized with 100% accuracy (see Figure 11). However, an evident output value floating exists for the slight wear condition recognition.

Table 7 shows the recognition accuracy of each health status of the gearbox dataset for the proposed and DBN methods. The training set is correctly recognized; however, one sample of the testing set is misjudged.

The contribution of the proposed method is the denoising ability to deal with random noises. To further attest to the superiority of the proposed method, signals of the dataset are combined with different levels of artificially random noises in the time domain and converted to frequency signals thereafter. Figure 12 shows the example of the noise mixed signals of bearing fault in different levels, and Figure 13 shows the example of noise mixed signals of gearbox fault. These artificially added noise dataset led to additional experiments being conducted to validate the proposed method and the comparison method DBN and SAE. Table 8 shows the recognition accuracy on different levels of noise interference.

From Table 8, as the level of noises increases, the superiority of the proposed method becomes evident. Compared with DBN, the proposed method can consistently achieve a high accuracy above 90%, thereby indicating that the proposed method is robust to a range of noise levels.
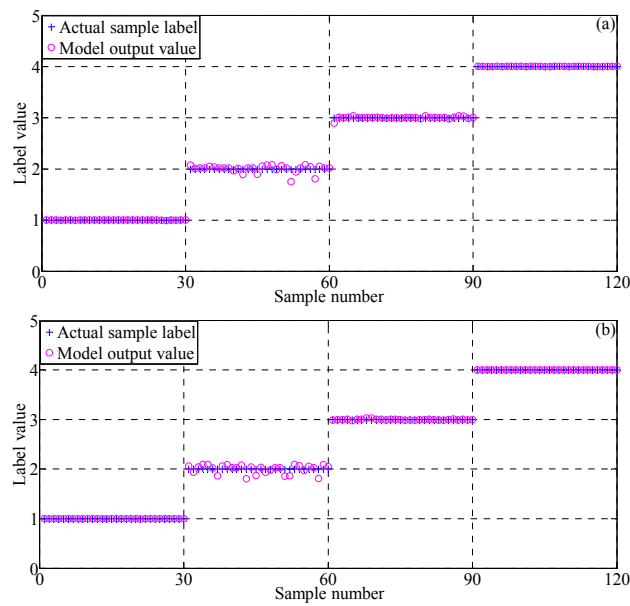
**Figure 11.** (**a**) Training results and (**b**) testing results of gearbox dataset by SDAE.

**Table 7.** Recognition results of gearbox health status obtained by SDAE and DBN.

| Method | Slight Wear Accuracy (%) | | Medium Wear (%) | | Broken Tooth (%) | | Health Wear (%) | |
|---|---|---|---|---|---|---|---|---|
| | Training | Testing | Training | Testing | Training | Testing | Training | Testing |
| SDAE | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| DBN | 100 | 100 | 98 | 100 | 100 | 100 | 100 | 100 |

**Table 8.** Recognition accuracy on different levels of noise interferences.

| Dataset | Noise Level | SDAE Accuracy (%) | | DBN Accuracy (%) | | SAE Accuracy (%) | |
|---|---|---|---|---|---|---|---|
| | | Training | Testing | Training | Testing | Training | Testing |
| Bearing Dataset I | 0.1 | 100 | 100 | 100 | 99.5 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 99.5 | 100 | 100 |
| | 0.3 | 100 | 97 | 100 | 92.5 | 100 | 95 |
| Bearing Dataset II | 0.1 | 100 | 100 | 100 | 98.5 | 100 | 98.5 |
| | 0.2 | 100 | 97 | 100 | 90.5 | 100 | 95 |
| | 0.3 | 100 | 91 | 100 | 84.5 | 98.5 | 82.5 |
| Gearbox | 0.1 | 100 | 100 | 100 | 99.17 | 100 | 99.5 |
| | 0.2 | 100 | 95 | 100 | 93.34 | 100 | 86.7 |
| | 0.3 | 100 | 93.4 | 100 | 87.5 | 100 | 77.5 |
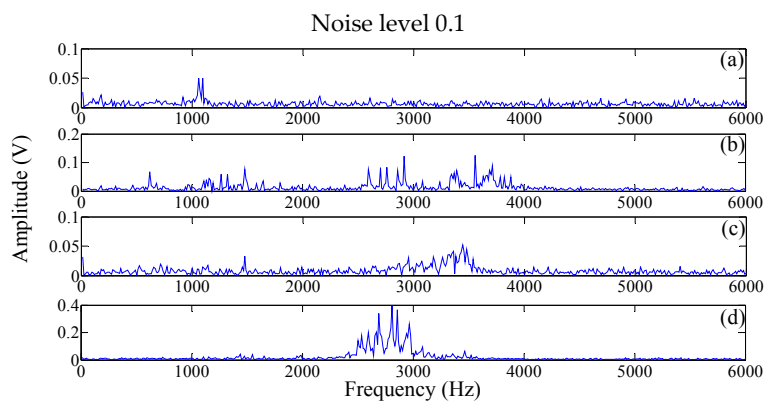


**Figure 12.** *Cont.*
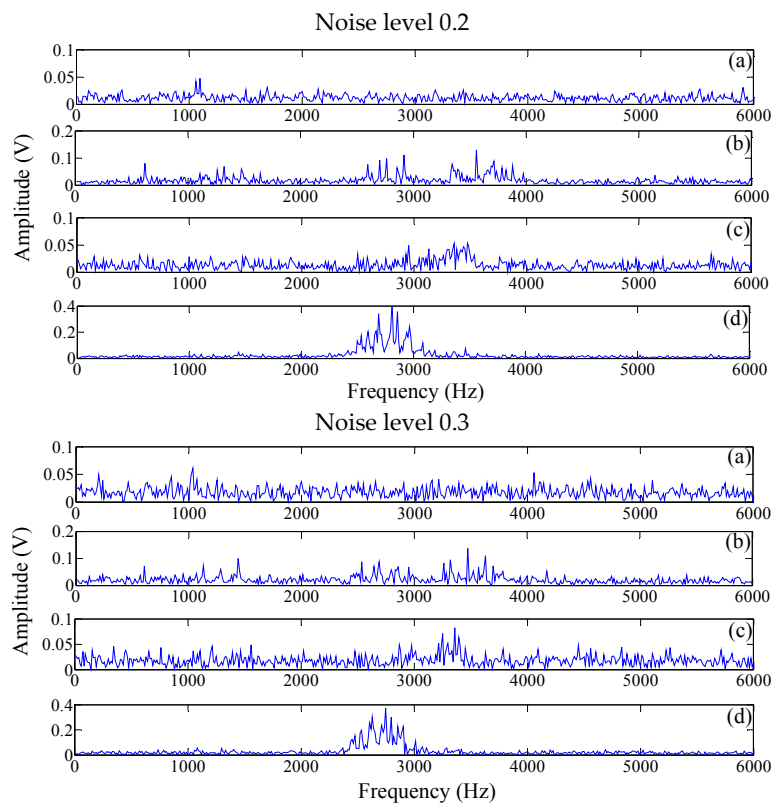
**Figure 12.** Examples of noised signals (from level 0.1 to 0.3) for each health condition (**a**) normal-bearing signal; (**b**) inner race fault-bearing signal; (**c**) ball fault-bearing signal; and (**d**) outer fault-bearing signal.
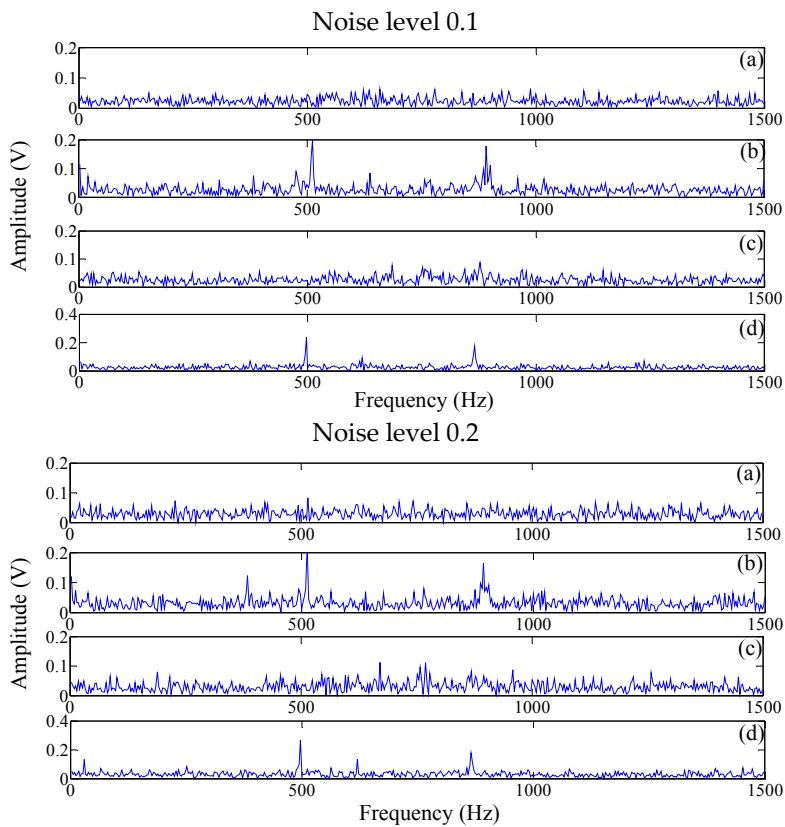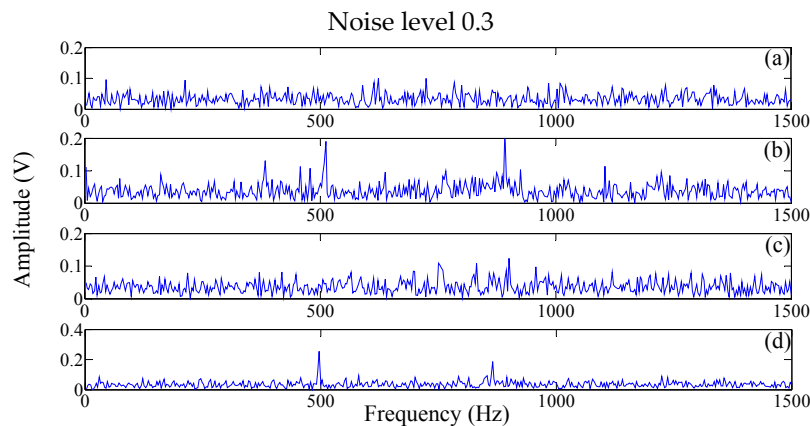


**Figure 13.** *Cont.*

**Figure 13.** Examples of noised signals (from level 0.1 to 0.3) and their spectra of four health status patterns: (**a**) slight wear; (**b**) medium wear; (**c**) broken tooth; and (**d**) normal wear signals.

## 5. Conclusions

In this study, an integrated fault recognizer based on SDAE is presented to denoise and extract features from the raw vibration signals. After the pre-training of each layer with the unsupervised learning method, the BP method is applied to train parameters of the whole model [45].

The proposed method is validated by two datasets, namely, rolling bearing datasets and gearbox datasets. The results of the experiments demonstrate the superiority of the SDAE model to other fault diagnosis methods, such as DBNs, particularly in the mixed noise situations. The proposed model achieves a high degree of accuracy because of its denoising function and offers an automatic feature extraction procedure that is practical and convenient for the application in rotating machine fault diagnosis.

In general, the proposed method has the following advantages.

- The proposed method can adaptively extract useful features from raw signals, which is substantially intelligent and has high recognition accuracy.
- The application of the deep fault recognizer to fault diagnosis has high generalization ability because of its unsupervised pre-training. This method also solves the problem where several labeled vibration signals for training are inaccessible.
- This method can be used independently because of its denoising layer. The proposed method is also more concise than the other traditional fault diagnosis methods, which combines several denoising and feature extraction methods. In addition, the parameters of the trained SDAE can be reused and can reduce quantity cost for training networks, which constantly consumes time and resources.

**Author Contributions:** Xiaojie Guo conceived and designed the experiments; Xiaojie Guo performed the experiments; Xiaojie Guo analyzed the data; Changqing Shen contributed gearbox measurement setup and data; Liang Chen, Xiaojie Guo, Changqing Shen wrote and revised the paper.

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

## References

1. Gao, J.; Wu, L.; Wang, H.; Guan, Y. Development of a Method for Selection of Effective Singular Values in Bearing Fault Signal De-Noising. *Appl. Sci.* **2016**, *6*. [CrossRef]

2.  He, Q.; Wang, J.; Liu, Y.; Dai, D.; Kong, F. Multiscale noise tuning of stochastic resonance for enhanced fault diagnosis in rotating machines. *Mech. Syst. Signal Process.* **2012**, *28*, 443–457. [CrossRef]

3.  Wang, D.; Tse, P.W.; Guo, W.; Miao, Q. Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis. *Meas. Sci. Technol.* **2011**, *22*, 500–502. [CrossRef]

4.  Lei, Y.; Lin, J.; He, Z.; Zi, Y. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2011**, *25*, 1738–1749. [CrossRef]

5.  Yan, R.; Liu, Y.; Gao, R.X. Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. *Mech. Syst. Signal Process.* **2012**, *29*, 474–484. [CrossRef]

6.  Shen, C.; Wang, D.; Liu, Y.; Kong, F.; Tse, PW. Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines. *Smart Struct. Syst.* **2014**, *13*, 453–471. [CrossRef]

7.  Yuan, L.; He, Y.; Huang, J.; Sun, Y. A New Neural-Network-Based Fault Diagnosis Approach for Analog Circuits by Using Kurtosis and Entropy as a Preprocessor. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 586–595. [CrossRef]

8.  Cui, J.; Wang, Y. A novel approach of analog circuit fault diagnosis using support vector machines classifier. *Measurement* **2011**, *44*, 281–289. [CrossRef]

9.  Bouzida, A.; Touhami, O.; Ibtiouen, R.; Belouchrani, A.; Fadel, M.; Rezzoug, A. Fault Diagnosis in Industrial Induction Machines through Discrete Wavelet Transform. *IEEE Trans. Ind. Electron.* **2011**, *58*, 4385–4395. [CrossRef]

10.  Lei, Y.; Liu, Z.; Ouazri, J.; Lin, J. A fault diagnosis method of rolling element bearings based on CEEMDAN. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2015**. [CrossRef]

11.  Wang, D.; Sun, S.; Tse, P.W. A general sequential Monte Carlo method based optimal wavelet filter: A Bayesian approach for extracting bearing fault features. *Mech. Syst. Signal Process.* **2015**, *53*, 293–308. [CrossRef]

12.  Wang, D.; Tsui, K.L.; Zhou, Q. Novel Gauss-Hermite integration based Bayesian inference on optimal wavelet parameters for bearing fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *73*, 80–91. [CrossRef]

13.  Shen, Z.; Chen, X.; Zhang, X.; He, Z. A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM. *Measurement* **2012**, *45*, 30–40. [CrossRef]

14.  Feng, Z.; Wang, T.; Zuo, M.J.; Chu, F.; Yan, S. Teager Energy Spectrum for Fault Diagnosis of Rolling Element Bearings. *J. Phys. Conf. Ser.* **2011**, *305*, 1022–1025. [CrossRef]

15.  Cai, J.H.; Cai, J.H. Fault diagnosis of rolling bearing based on empirical mode decomposition and higher order statistics. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2014**, *229*, 203–210. [CrossRef]

16.  Chen, Z.; Li, C.; Sánchez, R.V. Multi-layer neural network with deep belief network for gearbox fault diagnosis. *J. Vibroeng.* **2015**, *17*, 2379–2392.

17.  Huang, L.; Wu, C.; Wang, J. Fault Pattern Recognition of Rolling Bearing Based on Wavelet Packet Decomposition and BP Network. *Sci. J. Inf. Eng.* **2015**, *67*, 7–13.

18.  Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3137–3147. [CrossRef]

19.  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**. [CrossRef]

20.  Gan, M.; Wang, C.; Zhu, C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mech. Syst. Signal Process.* **2015**, *72–73*, 92–104. [CrossRef]

21.  Tran, V.T.; Althobiani, F.; Ball, A. An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert Syst. Appl.* **2014**, *41*, 4113–4122. [CrossRef]

22.  Tamilselvan, P.; Wang, P. Failure diagnosis using deep belief learning based health state classification. *Reliab. Eng. Syst. Saf.* **2013**, *115*, 124–135. [CrossRef]

23.  Wang, T.; Wu, X.; Liu, T.; Xiao, Z.M. Gearbox Fault Detection and Diagnosis Based on EEMD De-noising and Power Spectrum. In Proceedings of the IEEE International Conference on Information and Automation, Lijiang, China, 8–10 August 2015.

24.  Orfanidis, S.J. *Introduction to Signal Processing*; Prentice Hall International: Upper Saddle River, NJ, USA, 1996; pp. 65–88.

25. Tan, J.; Lu, W.; An, J.; Wan, X. Fault diagnosis method study in roller bearing based on wavelet transform and stacked autoencoder. In Proceedings of the 54th Control and Decision Conference, Osaka, Japan, 15–18 December 2015.

26. Morlet, R.S. Wavelet UDWT De-noising and EMD based Bearing Fault Diagnosis. *Electronics* **2013**, *17*, 1–8.

27. Vincent, P.; Larochelle, H.; Lajoie, I.; Bingio, Y.; Manzagol, P.-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.

28. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the ACM International Conference, Helsinki, Finland, 5–9 June 2008; pp. 1096–1103.

29. Larochelle, H.; Erhan, D.; Vincent, P. Deep Learning using Robust Interdependent Codes. *J. Mach. Learn. Res.* **2009**, *5*, 312–319.

30. Vincent, P. A Connection between Score Matching and Denoising Autoencoders. *Neural Comput.* **2011**, *23*, 1661–1674. [CrossRef] [PubMed]

31. Baldi, P.; Guyon, G.; Dror, V.; Lemaire, V.; Taylor, J.; Silver, D. Autoencoders, Unsupervised Learning, and Deep Architectures Editor. *J. Mach. Learn. Res.* **2012**, *27*, 37–50.

32. Leng, B.; Guo, S.; Zhang, X.; Zhang, X. 3D object retrieval with stacked local convolutional autoencoder. *Signal Process.* **2015**, *112*, 119–128. [CrossRef]

33. Tan, C.C.; Eswaran, C. Using autoencoders for mammogram compression. *J. Med. Syst.* **2011**, *35*, 49–58. [CrossRef] [PubMed]

34. Liu, Y.; Feng, X.; Zhou, Z. Multimodal video classification with stacked contractive autoencoders. *Signal Process.* **2015**, *120*, 761–766. [CrossRef]

35. Li, J.; Struzik, Z.; Zhang, L.; Cichocki, A. Feature learning from incomplete EEG with denoising autoencoder. *Neurocomputing* **2014**, *165*, 23–31. [CrossRef]

36. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 1–9. [CrossRef]

37. Luo, L.; Su, H.; Ban, L. Independent component analysis—Based sparse autoencoder in the application of fault diagnosis. In Proceedings of the 11th World Congress on Intelligent Control and Automation (WCICA), Shenyang, China, 29 June–4 July 2014; pp. 1378–1382.

38. Saravanan, N.; Ramachandran, K.I. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert Syst. Appl.* **2010**, *37*, 4168–4181. [CrossRef]

39. Xu, J.; Li, H.; Zhou, S. An Overview of Deep Generative Models. *IETE Tech. Rev.* **2015**, *32*, 131–139. [CrossRef]

40. Ng, S.S.Y.; Tse, P.W.; Tsui, K.L. A One-Versus-All Class Binarization Strategy for Bearing Diagnostics of Concurrent Defects. *Sensors* **2014**, *14*, 1295–1321. [CrossRef] [PubMed]

41. Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.A.; Vincent, P.; Bengio, S. Why Does Unsupervised Pre-training Help Deep Learning? *J. Mach. Learn. Res.* **2010**, *11*, 625–660.

42. Loparo, K.A. Case Western Reserve University Bearing Data Center. Available online: http://csegroups.case.edu/bearingdatacenter/home (accessed on 25 January 2015).

43. Shen, C.; Wang, D.; Kong, F.; Tse, P.W. Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement* **2013**, *46*, 1551–1564. [CrossRef]

44. Liu, H.; Liu, C.; Huang, Y. Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mech. Syst. Signal Process.* **2011**, *25*, 558–574. [CrossRef]

45. Alejo, R.; Monroy-De-Jesús, J.; Pacheco-Sánchez, J.; López-González, E.; Antonio-Velázquez, J. A Selective Dynamic Sampling Back-Propagation Approach for Handling the Two-Class Imbalance Problem. *Appl. Sci.* **2016**, *6*. [CrossRef]