

# A Fault Identification Method Based on an Ensemble Deep Neural Network and a Correlation Coefficient

Yanli Yang (✉ [yyl070805@163.com](mailto:yyl070805@163.com))

Tiangong University <https://orcid.org/0000-0003-1919-8857>

Yichuan He

tiangong university

---

## Research Article

**Keywords:** Rotating machinery, ensemble deep neural networks, fault diagnosis, deep learning, correlation coefficient.

**Posted Date:** January 31st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1307228/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# A fault identification method based on an ensemble deep neural network and a correlation coefficient

Yanli Yang\* and Yichuan He

Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems,  
Tiangong University, Binshuixi Road 399, Xiqing District, Tianjin 300387, China

**Abstract**—An intelligent fault identification algorithm based on ensemble deep neural networks and correlation coefficients is proposed for rotating machinery fault detection. In this algorithm, three deep neural networks (DNNs) are arranged to initially identify faults in the frequency domain, wavelet domain and envelope spectrum angle, respectively. Then, correlation coefficients are adopted to evaluate the recognition results of the DNNs. The reliabilities of the DNNs' recognition results are divided into 6 levels according to the outputs of the DNNs and the correlation coefficients. Finally, the evaluation results of the three DNNs are put through fusion processing to generate more reliable recognition results. This algorithm identifies fault signals from multiple angles and combines the correlation coefficients to make a comprehensive judgment, which is beneficial for fault identification. The proposed algorithm is tested with the bearing data provided by the Case Western Reserve University Bearing Data Center. The test results show that the algorithm has a high recognition accuracy for homologous data and a good generalization ability for nonhomologous data.

**Index Terms**—Rotating machinery; ensemble deep neural networks; fault diagnosis; deep learning; correlation coefficient.

## 1 Introduction

With the development of modern industrial technology and the proposal of industry 4.0, rotating machinery has developed rapidly in the direction of large-scale, intelligent, continuous, and complex structures. In view of the long-term operation of rotating machinery, the performance of components will gradually decrease with increasing runtime. The failure of any one component can cause the entire device to work abnormally or even cause serious equipment failure, which will affect the normal production activities of the enterprise. For example, damage to the bearings, which are a precision mechanical support component, can lead to machine failures and even serious accidents [1]. Thus, the reliability of machine operation is receiving substantial attention. The fault identification and diagnosis of rotating machinery have become an important research focus in the field of fault diagnosis.

Currently, fault diagnosis has entered the era of big data with the rapid development of the Industrial Internet of Things. Deep learning [2] has become a useful tool for dealing with big data and has facilitated breakthroughs in the fields of speech recognition and image recognition [3]. Deep learning can approximate complex functions through deep network structures and learn the laws behind the data, which is beneficial to improving the intelligence of fault detection. Moreover, this approach does not require prior knowledge of signal processing techniques [4]. This can bridge the gap between big data and machine intelligence health monitoring [5-7]. Note that a deep learning approach requires a lot of data, computational resources

---

\* Corresponding author. Tel.: +8615522720085.

*E-mail address:* yy1070805@163.com (Y.L. Yang)

and a black-box solution, which restricts its application in the field of fault diagnosis. Generally, the application of deep learning to the diagnosis of mechanical faults has become a trend and has also ushered in a new direction for the fault diagnosis of rotating machinery.

With deep learning, it is possible to learn and memorize fault modes and then use the results of training to classify and identify faults. Applying deep learning to fault detection has attracted the attention of many researchers. Gan et al. [8] proposed a two-layer hierarchical diagnosis network using deep belief networks (DBNs) to identify fault types and fault severities. Shen et al. [9] proposed an improved hierarchical adaptive DBN for bearing fault diagnosis. Using an automatic encoder, Jia et al. [10] proposed an intelligent diagnostic method based on a deep neural network (DNN). Using a DNN, Yang et al. proposed fault signal automatic classification methods in the frequency domain [11] and the wavelet domain [12]. Shao et al. [13] used a three-layer DBN to diagnose rolling-element bearing faults. Zhou et al. [14] proposed a method to identify failure modes with two automatic encoder networks connected in series. Xiang et al. [15] proposed a fault diagnosis method based on the Teager computed order spectrum and the stacked autoencoder (SAE). Lu et al. [4] used a stacked denoising autoencoder (SDAE) to diagnose faults. Guo et al. [16] proposed an automatic denoising and feature extraction method based on SDAE for fault identification, which gives a recognition accuracy that is 7% higher than that of DBN in noisy conditions using the Case Western Reserve University (CWRU) bearing data [17]. In addition, a one-dimensional convolutional neural network (1D CNN) [18] and a deep capsule neural network [19] were also used to diagnose bearing faults. Chen et al. [1], after comparing the results of a DBN, stacked autoencoders (SAEs) and deep Boltzmann models with the bearing signals measured in the laboratory, found that these models can achieve the high-precision fault diagnosis of rolling-element bearings.

A comprehensive analysis of these studies shows that the popular method used to diagnosis faults is the application of a single deep learning model in a specific domain, such as the time-frequency domain, frequency domain or other domains. However, fault diagnosis usually requires comprehensive judgment from multiple angles. In many cases, multiple faults are coupled into one signal due to the complexity and variability of machine working conditions. A single model seldom performs well across a variety of applications because the model has a limited learning ability for complicated physical problems [20]. The varying working conditions, such as the rotating speed oscillation or load variation, pose challenges for fault identification [15, 21, 22]. Any observer can see that analyzing the fault information requires multiple feature modalities [1, 20, 23]. Thus, it is difficult to achieve practical results when using a single network to diagnose faults from a single perspective.

Naturally, combining different deep learning models to diagnose machine faults has also attracted attention. In [1], three deep neural network models—deep Boltzmann machines (DBMs), DBNs, and SAEs—were employed to identify the fault conditions. By combining two layers of sparse self-encoding neural networks with a DBN, Chen et al. [24] proposed a multisensor feature fusion method. An ensemble deep learning diagnosis method, in which the convolutional residual network (CRN), DBN and deep autoencoder (DAE) are weighted and integrated, was proposed to realize the effective diagnosis of rotor and bearing faults [20]. A hybrid intelligent model consisting of a fuzzy min-max (FMM) neural network and a random forest (RF) model was proposed in [25] to classify ball bearing faults. A multimodal deep support vector classification (MDSVC) approach was proposed in [23] to diagnose gearbox faults. In this method, three Gaussian-Bernoulli deep Boltzmann machines (GDBMs) were used to learn the feature patterns of time modality, frequency modality and wavelet modality, which was followed by a support vector classifier to fuse the GDBMs' outputs. The provided experiments show that the best fault classification rate is achieved in these experiments compared to representative deep and shallow learning methods. However, these methods have only been tested based on homologous data. These methods actually used the ensemble neural network (ENN) to achieve fault diagnosis.

Although some significant advances have been made in detecting faults using deep learning, many technical problems remain to be solved. The development of fault detection algorithms with good generalization capabilities is one of the technical challenges. Ensemble learning can improve the generalization ability of the model ensemble [20]. Theoretical and empirical work has shown that an ENN often exhibits a better generalization performance compared to any individual neural network by itself [26]. The ENN still gives a black-box solution, although it has an excellent performance in fault diagnosis. Thus, we try to use the correlation coefficient to assist the ENN in fault diagnosis.

It was reported that the correlation coefficient is an important tool for determining the correlation degree between samples [27]. Inspired by the work in [20, 26, 27], we propose a fault diagnosis method based on ensemble deep neural networks (EDNNs) and correlation coefficients. The method first uses an EDNN to identify faults from the frequency domain, wavelet domain and envelope spectrum angle and then combines the correlation coefficients to evaluate the DNNs recognition results. The correlation coefficients are used to help the DNN accurately identify specific fault samples. Then, the fault diagnosis result is given by a comprehensive judgment. In fact, the ENN can be classified into a homogeneous ensemble with the same learning model and as a heterogeneous ensemble characterized by different machine-learning models [26]. Here, a homogeneous EDNN, which combines three DNNs, is used to recognize faults from the frequency domain, wavelet domain and envelope spectrum analysis perspectives. The proposed method identifies fault signals from multiple angles and combines the correlation coefficients to improve the diagnostic reliability. The details of the proposed algorithm are introduced in the following steps. In Section 2, a brief review of the theory of DNNs and the correlation coefficients is given. Section 3 presents a detailed description of the proposed approach. Section 4 shows some experiments and testing on the CWRU bearing data. Finally, the conclusions are drawn in Section 5.

## 2 Deep neural network and correlation coefficient

### A. Basic theory of DNN

The DNN has a similar hierarchical structure to that of the traditional neural network. It is characterized by a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. A DNN usually has more than one hidden layer. Each successive layer uses the output from the previous layer as the input. Higher-level features are derived from lower-level features to form a hierarchical representation. A DNN can adaptively capture the representation information from raw data through multiple nonlinear transformations and approximate complex nonlinear functions with little error [28].

The neurons summarize the inputs and give the output through an activation function. Commonly used activation functions are sigmoid functions, rectified linear functions, and hyperbolic tangent functions. The activation function used in our algorithm is a sigmoid function. The sigmoid function can map a real number to the interval of  $(0, 1)$ . Note that the sigmoid function has two saturation regions, which results in low discrimination between the inputs. This is also one reason why we evaluate the output of DNNs.

Like traditional neural networks, weights and thresholds can be updated by gradient descent methods during training. However, DNN training has its own unique features. A DNN can be trained one layer at a time by supervised, semisupervised, or unsupervised learning methods. One of the outstanding advantages is that DNNs can use unsupervised training on pattern analysis. The depth of the network can be very deep as a result of the layer-by-layer training mechanism. For more information about deep learning, refer to [2, 3].

A neural network with a proper architecture is necessary for pattern analysis. As discussed in the previous section, deep learning has many specific models, such as DNN, DBN, SDAE and SAEs. The structure of a DNN is similar to that of traditional neural networks, but it has a deep hidden layer. To utilize the ensemble

method and facilitate comparison with traditional neural network methods, we use DNNs to perform faults diagnosis.

### B. Basic theory of the correlation coefficient

The correlation coefficient is a statistical indicator used to reflect the strength of the correlation between variables. It is a tool commonly used for measuring the relationship between two random variables or two signals.

As one of the most commonly used correlation coefficients, the Pearson's correlation coefficient of two signals,  $r(Y, X)$ , can be calculated by [29]

$$r(Y, X) = \frac{Cov(X, Y)}{\sqrt{Var[X]Var[Y]}} \quad (1)$$

where  $|r(Y, X)| \leq 1$ ,  $Cov(X, Y)$  represents the covariance of  $X$  and  $Y$  and  $Var[\cdot]$  denotes the variance. Thus,  $X$  and  $Y$  are positively correlated when  $r(Y, X) > 0$ , and they are negatively correlated when  $r(Y, X) < 0$ .

The greater value of  $|r(Y, X)|$ , the greater is the correlation between  $X$  and  $Y$  [30]. The closer  $|r(Y, X)|$  is to 1, the closer is the linear relationship between  $X$  and  $Y$ . Otherwise, the closer  $|r(Y, X)|$  is to 0, the weaker is the linear relationship between  $X$  and  $Y$ .  $X$  and  $Y$  are completely linearly related, especially when  $|r(Y, X)| = 1$  and  $X$  is not related to  $Y$  when  $|r(Y, X)| = 0$ .

The absolute value of the correlation coefficient takes a value between 0 and 1 to reflect a certain degree of linear correlation between two variables or signals [30]. Generally, the degree of correlation can be divided into strong correlation, moderate correlation, and weak correlation according to the correlation coefficient. A more specific division is shown in Table 1, which is commonly used in application.

Table 1. The correlation coefficient and degree of linear correlation.

Absolute value of correlation coefficient	Degree of linear correlation
0.80~1.0	extremely strong
0.60~0.80	strong
0.40~0.60	moderate
0.20~0.40	weak
0~0.20	extremely weak

### 3 Faults identification algorithm

As discussed in the previous section, a comprehensive analysis from different angles using multiple spectral analysis methods is helpful for diagnosing faults. It was reported previously [1, 8, 10, 13-16, 20, 24, 25] that the DNN is a powerful tool for fault diagnosis. Alongside this thought, we propose an algorithm with which to diagnose faults using an EDNN and correlation coefficients, as shown in Fig. 1. The method combines the EDNN and correlation coefficients to recognize faults from multiple angles.

It was pointed out in [1] that the feature extraction of vibration signals is still a necessary step for DNN-based classifiers. Here, a time series signal is decomposed by the fast Fourier transform (FFT), the wavelet packet transform (WPT), and envelope analysis. Then, the features are extracted from the frequency domain, the wavelet domain, and the envelope spectrum. These features are put into the EDNN to further learn the inherent laws of the data and identify faults.

It is well known that neural networks need to be trained before use. There is nothing noteworthy about training a neural network, so we will not detail that process here. A well-trained neural network plays an important role in the subsequent steps of our method, so we address this step first. For neural networks, three DNN models are used in our method. The three models are a DNN for handling frequency-domain

features, a DNN for handling wavelet-domain features, and a DNN for handling envelope-spectrum features, which are noted as DNN-FDF, DNN-WDF and DNN-ESF, respectively.

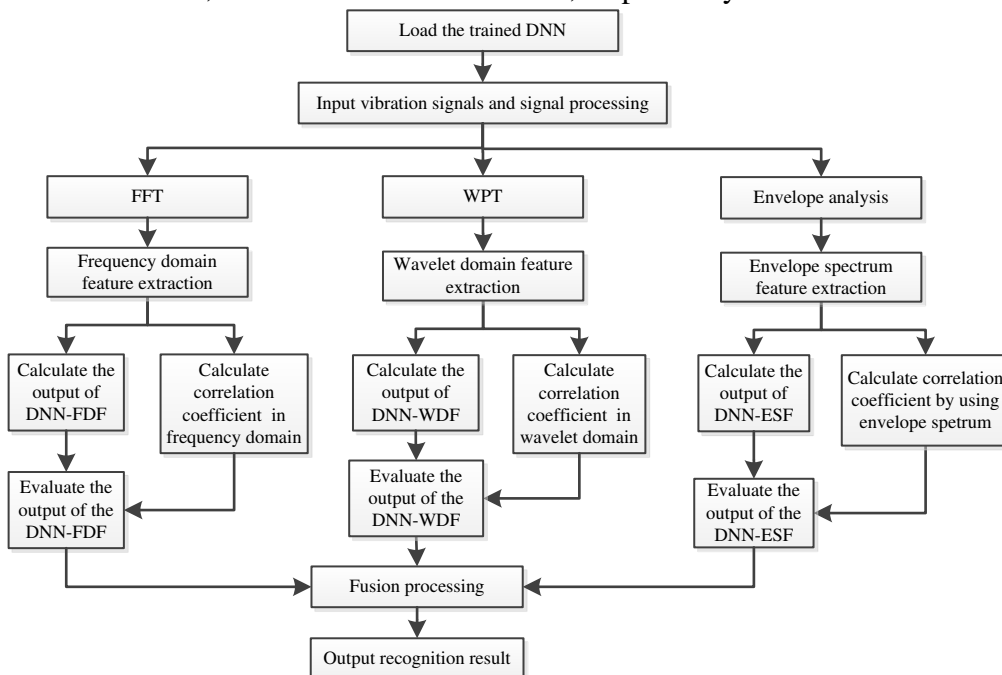


Fig. 1. Flowchart of the proposed algorithm.

It is a common thought that the power of a neural network depends to a large extent on the dataset. The incompleteness of the training samples will affect the generalization ability of neural networks. Take the DNN with layers 1024-260-130-80-50-3 as an example—the training samples are 4 normal data, 4 inner race (IR) fault data, and 4 rolling element (ball) fault data, and the test sample is one outer race (OR) fault data. These data are selected from the CWRU bearing data, of which the characteristics can be found in [31]. After 100 epochs of training, the DNN has completely remembered these training samples. The test sample with the OR fault is identified as a type of ball fault, in which the output of the corresponding neuron exceeds 0.95. However, it is quite clear that there is a big difference between the test sample and the training samples, as shown in Fig. 2. In addition, the correlation coefficients, among which the maxima are 0.390, 0.293 and 0.217, show that there is little correlation between the test sample and the training samples. Thus, this requires an evaluation of the output of the DNN.

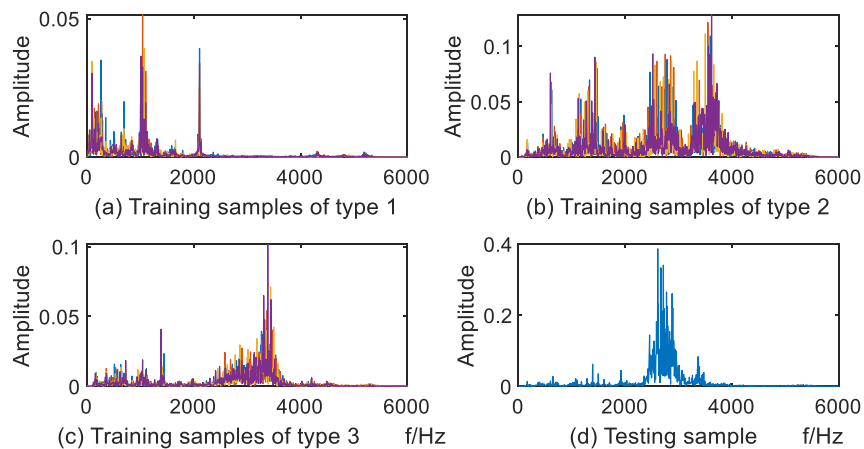


Fig. 2. Illustration of the difference between training and test samples.

The correlation coefficient is used to help evaluate the outputs of DNNs. The correlation coefficients are calculated using the features put into DNNs to identify faults. By incorporating the correlation coefficients, we evaluate the outputs of the DNNs. Finally, the evaluation results of the DNNs are analyzed by fusion processing to obtain the diagnosis results. Detailed descriptions of the evaluation of the DNN outputs and the fusion analysis of the evaluation results are shown in the following text.

#### A. Evaluation of the DNNs output

We can get a result after putting the extracted frequency domain features from test signals into the DNN-FDF. The result is noted as  $(H_r^f, H_c^f)$ , where  $H_r^f$  denotes the category label of the neuron with the largest output value in the output layer of the DNN-FDF and  $H_c^f$  represents the output of the corresponding neuron. Since the activation function of the output layer neurons is the sigmoid function,  $H_c^f$  takes a value between 0 and 1.

In the frequency domain, we calculate the correlation coefficients between the test signal and the training samples. Let  $C_c^f$  denote the maximum correlation coefficient, and let  $C_r^f$  represent the category label of the corresponding training sample. Suppose that  $(S_r^f, S_c^f)$  denotes the assessment result of the DNN-FDF's recognition results. The values of  $S_r^f$  and  $S_c^f$  are determined by  $(H_r^f, H_c^f)$  and  $(C_r^f, C_c^f)$ . If  $H_r^f = C_r^f$ , we then let  $S_r^f = H_r^f$  and have

$$S_c^f = \begin{cases} 5, & \text{if } |C_c^f| > 0.8 \text{ and } H_c^f > 0.8, \\ 4, & \text{if } 0.6 < |C_c^f| \leq 0.8 \text{ and } H_c^f > 0.8, \\ 3, & \text{if } (|C_c^f| > 0.4 \text{ and } 0.6 < H_c^f \leq 0.8) \text{ or } (0.4 < |C_c^f| \leq 0.6 \text{ and } H_c^f > 0.8), \\ 2, & \text{if } (|C_c^f| > 0.2 \text{ and } 0.5 < H_c^f \leq 0.6) \text{ or } (0.2 < |C_c^f| \leq 0.4 \text{ and } H_c^f > 0.6), \\ 1, & \text{others.} \end{cases} \quad (2)$$

Otherwise, if  $H_r^f \neq C_r^f$ , we still let  $S_r^f = H_r^f$ , but we also let  $S_c^f = 0$ . In fact, we divide the DNN-FDF's recognition results into six levels based on the correlation between the test sample and the training samples to indicate the reliability of the DNN-FDF's recognition results. Thus, we have that  $S_r^f$  takes the value of the DNN-FDF's output and  $S_c^f$  represents the reliability of the DNN-FDF's recognition results.

Similarly, we obtain the recognition result of the DNN-WDF, which is denoted as  $(H_r^w, H_c^w)$ , using the extracted wavelet domain features from the test signals.  $H_r^w$  denotes the category label of the neuron with the largest output value in the output layer of the DNN-WDF, and  $H_c^w \in [0,1]$  represents the output of the corresponding neuron. At the same time, we calculate the correlation coefficients between the test signal and the training samples in the wavelet domain. Among these correlation coefficients, we select the maximum correlation coefficient, which is denoted as  $C_c^w$  and determine the category label of the corresponding training sample, which is denoted as  $C_r^w$ . Then, we evaluate the output of DNN-WDF by considering the values of  $(H_r^w, H_c^w)$  and  $(C_r^w, C_c^w)$ . The reassessment result is denoted as  $(S_r^w, S_c^w)$ . We have  $S_r^w = H_r^w$ , regardless of whether  $H_r^w$  and  $C_r^w$  are the same, but  $S_c^w$  has a different value. If  $H_r^w = C_r^w$ ,  $S_c^w$  can be calculated by (4) with the superscript  $f$  in (4) replaced by  $w$ . Otherwise,  $S_c^w = 0$ . The DNN-WDF's results are divided into six levels to indicate the reliability of the recognition results.

For the features extracted by envelope spectrum, we calculate the recognition result of DNN-ESF, which is denoted as  $(H_r^e, H_c^e)$ , and the correlation coefficients  $(C_r^e, C_c^e)$ . Among the correlation coefficients calculated by the envelope features, we select the maximum correlation coefficient, which is denoted as  $C_c^e$ , and determine the category label of the corresponding training sample, which is denoted as  $C_r^e$ . Let

$(S_r^e, S_c^e)$  denote the evaluated recognition result of DNN-ESF. If  $H_r^e = C_r^e$ , we then let  $S_r^e = H_r^e$  and calculate  $S_c^e$  by (4) with the superscript  $f$  in (4) replaced by  $e$ . Otherwise, we still have  $S_r^e = H_r^e$ , but  $S_c^e = 0$ . The DNN-ESF's results are also divided into six levels to indicate the reliability of the recognition results.

### B. Fusion processing

Fusion processing is performed to obtain the final fault identification result based on the evaluated recognition results of the three networks. The final fault identification result is written as a vector  $(S_r^o, S_c^o)$ , where  $S_r^o$  denotes the type of fault identified and  $S_c^o$  represents the credibility of the recognition result.

The value of  $S_r^o$  is chosen among  $S_r^f$ ,  $S_r^w$  and  $S_r^e$ . The previous analysis shows that the values of  $S_r^f$ ,  $S_r^w$  and  $S_r^e$  are determined by the outputs of the DNN-FDF, DNN-WDF, and DNN-ESF, respectively. Hence,  $S_r^o$  is one of the values output by three networks. For a test sample, if  $S_r^f$ ,  $S_r^w$  and  $S_r^e$  are equal, this means that the DNN-FDF, DNN-WDF, and DNN-ESF identified the same fault result. We believe that this case has the highest credibility level and set  $S_c^o = 10$ . In this way, the credibility of the recognition result is divided into ten levels. The value of  $(S_r^o, S_c^o)$  and the corresponding conditions are listed in Table 2. One point to note is that the values of vector  $(S_r^o, S_c^o)$  are calculated according to the sequence numbers in Table 2 from small to large. That is, the smaller the number is, the higher the considered priority.

Table 2. Fusion processing rules (the smaller the No. is, the higher the priority).

No.	Vector $(S_r^o, S_c^o)$	Condition
1	$(S_r^f, 10)$	$S_r^f = S_r^e = S_r^w$ & $S_c^f > 1$ & $S_c^e > 1$ & $S_c^w > 1$
2	$(S_r^f, 9)$	$(S_r^f = S_r^e$ & $S_c^f > 4$ & $S_c^e > 4)$ or $(S_r^f = S_r^w$ & $S_c^f > 4$ & $S_c^w > 4)$
3	$(S_r^w, 9)$	$S_r^w = S_r^e$ & $S_c^w > 4$ & $S_c^e > 4$
4	$(S_r^e, 8)$	$S_c^e = 5$
5	$(S_r^f, 8)$	$S_c^f = 5$
6	$(S_r^f, 7)$	$S_r^f = S_r^e$ & $2 < S_c^f \leq 4$ & $2 < S_c^e \leq 4$
7	$(S_r^e, 7)$	$S_r^e = S_r^w$ & $2 < S_c^e \leq 4$ & $2 < S_c^w \leq 4$
8	$(S_r^f, 7)$	$S_r^f = S_r^w$ & $3 < S_c^f \leq 4$ & $3 < S_c^w \leq 4$
9	$(S_r^f, 6)$	$S_c^f = 4$
10	$(S_r^e, 6)$	$S_c^e = 4$
11	$(S_r^f, 5)$	$S_r^f = S_r^w$
12	The vector corresponding to the maximum of $S_c^f, S_c^e, S_c^w$	$S_r^f \neq S_r^e \neq S_r^w$ & $S_c^f > 0$ & $S_c^e > 0$ & $S_c^w > 0$
13	$(S_r^e, 1)$	Others

We expect to use ensemble networks to improve the accuracy of fault detection. In fact, we use DNNs to detect faults from multiple perspectives and combine the results with correlation coefficients to improve the diagnostic reliability. The effectiveness of the proposed algorithm is tested using the CWRU bearing data, which is provided in the next section.

### 4 Testing and analysis

The proposed method is implemented using the DeepLearn Toolbox [32], and it is tested using the CWRU bearing data. The CWRU bearing data was collected on normal bearings and bearings with single-point drive end (DE) and fan-end (FE) defects [17]. The faulty data include fault types of IR, OR, and ball. For the drive-end bearing experiments, the data were sampled at 12 kHz and 48 kHz. The fan-end bearing



data was only collected at 12 kHz. To test with the same structure of the network, 48k data are resampled, and the sampling rate is converted to 12 kHz. In each training set, a small number of normal samples are added.

For the bearing data, we extract three types of fault characteristics. We use FFT to transform a time-series signal into the frequency domain and then extract 1024 frequency domain features. At the same time, we extract 1024 frequency domain features by performing FFT on the envelope of the data. By using the WPT, we get 256 WP energy characteristics.

For the different types of features, the DNN structure is shown in Table 3. As shown in Table 3, the input layer is different for different DNNs. Nevertheless, the number of output layer neurons is determined by the types of sample labels. The weights of the DNN are initialized randomly. The learning rate is set to 1. The activation function of the neurons is the sigmoid function. The samples are selected randomly as the input for training the DNN. The designed DNN has six layers. The minimum training epoch is 150.

Table 3. The DNN structures for different fault characteristics.

Types	Input	Hidden layer
DNN-FDF	1024	260-130-80-50
DNN-WDF	256	200-130-80-50
DNN-ESF	1024	260-130-80-50

#### A. Training and testing on highly similar samples

We consider that the training data set and the testing data set are from the same raw data. Although the data are measured under the same test conditions, the signal also changes with time due to the influence of many factors. Thus, we can know that the training set and the test set are not exactly the same. Nevertheless, the similarity between two samples is high. Taking the record 108DE as an example, the smallest correlation coefficient between two samples is 0.7123.

For each data record, we choose a segment with a specific length as a training sample and select several segments as test samples. For example, we select 10, 10, and 5 segments as the test samples for the 12k drive end, 12k fan end, and 48k drive end data, respectively. We pick 5 test samples for the 48k drive end data because some records are not long enough. In fact, many people choose 70% of the samples as the training set and 30% of the samples as the test set. In comparison, we select 9.09% of the samples as the training set and 90.91% of the samples as the test set for the 12k-sampled data and 16.67% of the samples as the training set and 83.33% of the samples as the test set for the 48k-sampled data.

The test results for the 12k drive end bearing data are shown in Table 4. As seen from this table, for each type of network, the test results are not zero, except for testing on the DNN-FDF using bearing data with a 0.007-inch fault width. The maximum number of misidentified samples is 21, and the corresponding accuracy rate is 94.75%. This shows that the three types of networks exhibit different capabilities and can learn the characteristics of signals from different perspectives. This is one of the reasons why we use ensemble networks for fault diagnosis. No matter which kind of network is used, the DNNs cannot accurately identify all fault samples for the three kinds of data. As a comparison, the misidentification number of our method is 0, and the recognition accuracy is 100%. This shows that our algorithm improves the recognition accuracy.

The testing result on the 48k drive-end bearing data is shown in Table 5. We can see that the number of misidentified samples for the three networks is 0, a small number and partly zero for bearing data with 0.007-, 0.014- and 0.021-inch faults, respectively. In contrast, the numbers of misidentified samples for our method are all zero. Especially in the case in which all three networks misidentified samples for the dataset with a 0.014-inch fault width, our algorithm achieved a recognition rate of 100%. This shows that our algorithm has a good performance.

The testing results for the 12k fan-end bearing data are shown in Table 6, and the recognition accuracies are shown in Fig. 3. Compared with Tables 4 and 5, this table shows that the number of misidentified samples is relatively large. Among the results of the three types of networks, the maximum number of misidentified samples is 92, and the corresponding accuracy rate is only 78.10%. Moreover, the minimum number of misidentified samples is 1. By fusing the results of the three network types and combining the results with the correlation coefficients, our method gives results in which there are only 8 misidentified samples for the testing on the bearing data with a 0.007-inch fault width. The recognition rate has been improved for all three fault widths. Although the recognition rate does not reach 100%, it is increased from the 95.71% of DNN-FDF, 96.90% of the DNN-WDF, and 78.10% of the DNN-ESF to 98.10% for the 0.007-inch fault width. Thus, we can conclude that our method significantly improves the recognition accuracy.

According to the above test, we find that a DNN has good generalization capabilities for homologous data. In this case, our method can reach a high recognition rate.

Table 4. The testing results for the 12k drive-end bearing data.

Fault width (in.)	Training dataset Normal+Fault	Labels	Testing samples	Misidentified samples			
				DNN-FDF	DNN-WDF	DNN-ESF	Our method
0.007	4+60	16	640	0	1	20	0
0.014	4+36	10	400	1	1	21	0
0.021	4+60	16	640	8	10	7	0

Table 5. The testing results for the 48k drive-end bearing data.

Fault width (in.)	Training dataset Normal+Fault	Labels	Testing samples	Misidentified samples			
				DNN-FDF	DNN-WDF	DNN-ESF	Our method
0.007	4+40	11	220	0	0	0	0
0.014	4+22	7	130	3	2	18	0
0.021	4+40	11	220	0	0	2	0

Table 6. The testing results for the 12k fan-end bearing data.

Fault width (in.)	Training dataset Normal+Fault	Labels	Testing samples	Misidentified samples			
				DNN-FDF	DNN-WDF	DNN-ESF	Our method
0.007	4+38	11	420	18	13	92	8
0.014	4+25	9	290	11	1	50	0
0.021	4+24	9	280	2	3	23	0

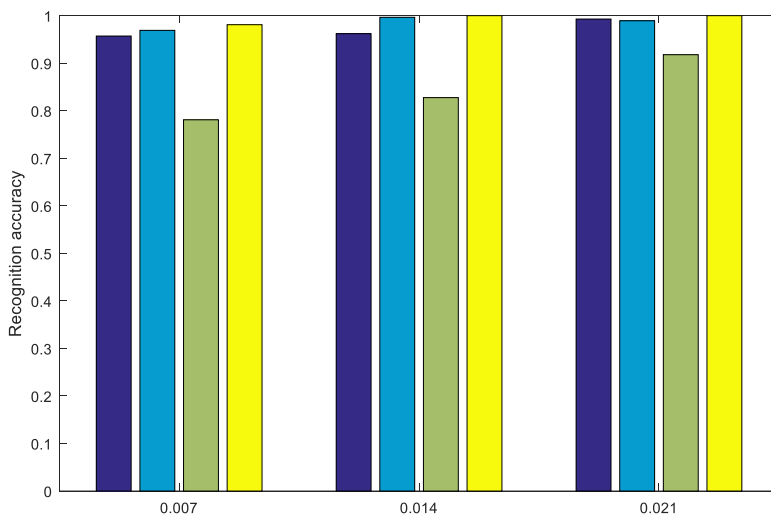


Fig. 3. The recognition accuracies for the 12k fan-end bearing data. For each group, the four column bars from left to right are the accuracies of the DNN-FDF, DNN-WDF, DNN-ESF, and our method.

## B. Training and testing on different datasets

We train the three types of DNNs using the 12k drive-end bearing data and test the trained DNNs with the 48k drive-end bearing data. For the convenience of testing, the 48k drive-end bearing data are resampled to a sampling frequency of 12 kHz. The 12k and 48k drive-end bearing data are measured in two independent testing processes. For the same severity of faults, the signal also shows a certain difference. In this case, the degree of similarity between the training sample and the test sample is less than that for the homologous data. Taking the records 108DE and 112DE as examples, the smallest correlation coefficient between samples of the two records is 0.598.

The test results for the 48k drive-end bearing data are shown in Table 7. In this table, the bearing data of DE and FE are separately trained and tested. Table 7 shows that the number of misidentified samples by our method has decreased, except for the DE of 0.014-inch and 0.021-inch fault widths. For the testing on the DE of 0.014-inch and 0.021-inch fault widths, the number of misidentified samples by our method is only more than that of DNN-ESF, and it is much less than those of DNN-FDF and DNN-WDF.

Table 7. The testing results for the 48k drive-end bearing data after training on 12k drive-end bearing data.

Fault width (in.)	Measured position	Training dataset		Testing samples	Misidentified samples			
		Normal+Fault	Labels		DNN-FDF	DNN-WDF	DNN-ESF	Our method
0.007	DE	4+20	6	120	42	52	14	7
	FE	4+20	6	120	0	0	0	0
0.014	DE	4+12	4	80	37	37	15	19
	FE	4+12	4	80	7	7	18	7
0.021	DE	4+20	6	120	62	89	1	34
	FE	4+20	6	120	13	10	9	8

To clearly show the working process of our algorithm, the data with a 0.014-inch fault width used in Table 7 are taken as an example for illustration, as shown in Table 8. The three data rows in italics and boldface are misidentified samples. From Table 8, we can see four situations. The first situation is that the three DNNs' recognition results are the same and are consistent with the results of the correlation coefficients. The data that matches this situation are the first eight samples in this table. At this point, the credibility level has reached the highest value of 10.

Table 8. The testing results for the 48k drive-end bearing data.

sample	$(H_r^f, H_c^f)$	$(C_r^f, C_c^f)$	$(S_r^f, S_c^f)$	$(H_r^w, H_c^w)$	$(C_r^w, C_c^w)$	$(S_r^w, S_c^w)$	$(H_r^e, H_c^e)$	$(C_r^e, C_c^e)$	$(S_r^e, S_c^e)$	$(S_r^o, S_c^o)$	Real label
100FE	(1, 0.995)	(1, 0.799)	(1, 4)	(1, 0.996)	(1, 0.92)	(1, 5)	(1, 0.989)	(1, 0.69)	(1, 4)	(1, 10)	1
97FE	(1, 0.995)	(1, 0.81)	(1, 5)	(1, 0.996)	(1, 0.893)	(1, 5)	(1, 0.996)	(1, 0.734)	(1, 4)	(1, 10)	1
98FE	(1, 0.996)	(1, 0.894)	(1, 5)	(1, 0.996)	(1, 0.976)	(1, 5)	(1, 0.996)	(1, 0.819)	(1, 5)	(1, 10)	1
99FE	(1, 0.995)	(1, 1)	(1, 5)	(1, 0.996)	(1, 1)	(1, 5)	(1, 0.996)	(1, 1)	(1, 5)	(1, 10)	1
189DE	(3, 0.995)	(3, 0.641)	(3, 4)	(3, 0.994)	(3, 0.789)	(3, 4)	(3, 0.992)	(3, 0.726)	(3, 4)	(3, 10)	3
190DE	(3, 0.993)	(3, 0.567)	(3, 3)	(3, 0.994)	(3, 0.813)	(3, 5)	(3, 0.993)	(3, 0.732)	(3, 4)	(3, 10)	3
191DE	(3, 0.796)	(3, 0.556)	(3, 3)	(3, 0.994)	(3, 0.844)	(3, 5)	(3, 0.992)	(3, 0.745)	(3, 4)	(3, 10)	3
192DE	(3, 0.99)	(3, 0.603)	(3, 4)	(3, 0.996)	(3, 0.883)	(3, 5)	(3, 0.989)	(3, 0.712)	(3, 4)	(3, 10)	3
174DE	(3, 0.805)	(2, 0.484)	(3, 0)	(3, 0.787)	(2, 0.765)	(3, 0)	(2, 0.97)	(2, 0.747)	(2, 4)	(2, 6)	2
175DE	(3, 0.75)	(2, 0.463)	(3, 0)	(3, 0.679)	(2, 0.721)	(3, 0)	(2, 0.992)	(2, 0.814)	(2, 5)	(2, 8)	2
176DE	(3, 0.647)	(2, 0.489)	(3, 0)	(3, 0.876)	(2, 0.76)	(3, 0)	(2, 0.993)	(2, 0.858)	(2, 5)	(2, 8)	2
177DE	(3, 0.845)	(2, 0.408)	(3, 0)	(3, 0.843)	(2, 0.632)	(3, 0)	(2, 0.992)	(2, 0.835)	(2, 5)	(2, 8)	2
<b>201DE</b>	<b>(2, 0.774)</b>	<b>(2, 0.568)</b>	<b>(2, 3)</b>	<b>(1, 0.986)</b>	<b>(1, 0.441)</b>	<b>(1, 3)</b>	<b>(4, 0.973)</b>	<b>(4, 0.255)</b>	<b>(4, 2)</b>	<b>(2, 3)</b>	<b>4</b>
202DE	(2, 0.809)	(2, 0.496)	(2, 3)	(3, 0.991)	(3, 0.805)	(3, 5)	(4, 0.937)	(4, 0.645)	(4, 4)	(4, 6)	4
<b>203DE</b>	<b>(3, 0.861)</b>	<b>(4, 0.523)</b>	<b>(3, 0)</b>	<b>(3, 0.991)</b>	<b>(3, 0.753)</b>	<b>(3, 4)</b>	<b>(4, 0.588)</b>	<b>(3, 0.621)</b>	<b>(4, 0)</b>	<b>(3, 5)</b>	<b>4</b>
<b>204DE</b>	<b>(3, 0.717)</b>	<b>(2, 0.515)</b>	<b>(3, 0)</b>	<b>(3, 0.994)</b>	<b>(3, 0.778)</b>	<b>(3, 4)</b>	<b>(3, 0.99)</b>	<b>(3, 0.727)</b>	<b>(3, 4)</b>	<b>(3, 7)</b>	<b>4</b>

The second situation is that the three DNNs' recognition results are the same, but one of them is inconsistent with the results of the correlation coefficients. The 204DE data belongs to this situation. In this case, DNN-FDF's recognition result is inconsistent with the results of the correlation coefficients, so the credibility level is at the minimum. However, the recognition results of the other two DNNs are consistent with the results of the correlation coefficients and have a high level of credibility. Hence, the final recognition results are consistent with the recognition results of the three DNNs and achieve a level of credibility that is not particularly high. Although the recognition result of this sample is inconsistent with the real label, this does not mean that the recognition result is wrong. The spectra of the 204DE data and training samples are shown in Fig. 4. As seen from this figure, the 204DE data have a large difference relative to the samples marked with L4 but have a certain similarity with the samples marked with L3.

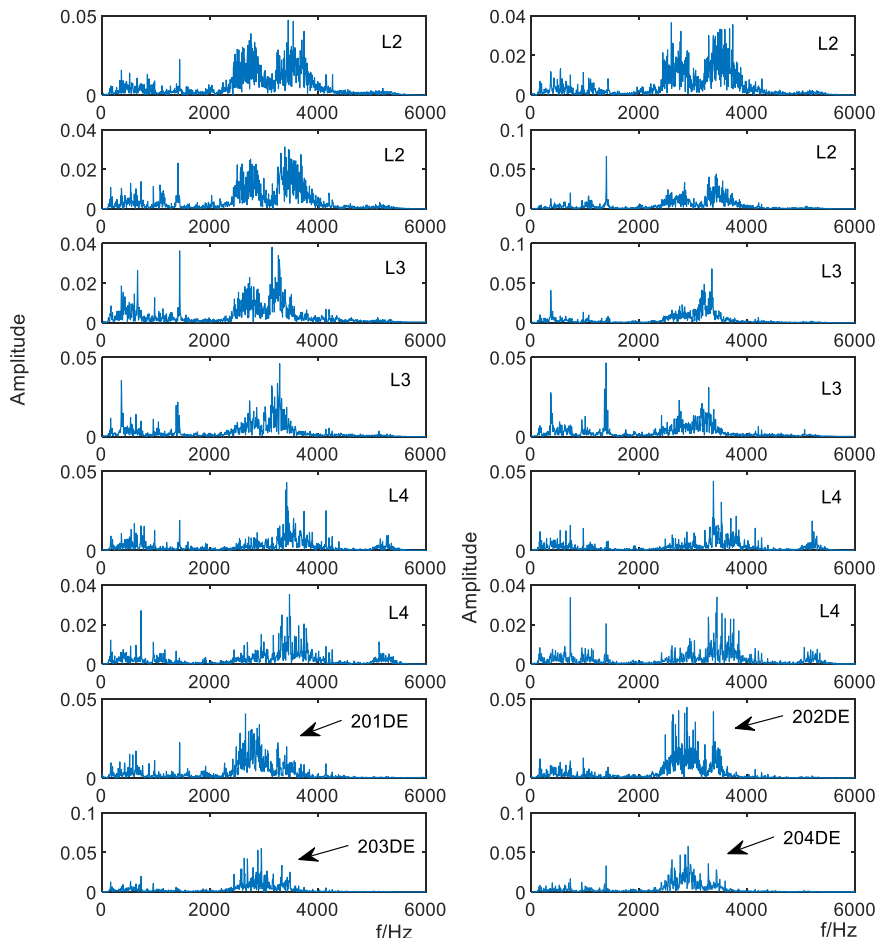


Fig. 4. The spectra of data from 201DE to 204DE and parts of the training samples. Samples labeled with the same tag, such as L2, have the same category label.

Another case is that not all the three DNNs' recognition results are the same and some of them are inconsistent with the results of the correlation coefficients. The data belonging to this situation are the four samples from 174DE to 177DE. The credibility levels of DNN-FDF's and DNN-WDF's recognition results are set to the lowest level because their value is inconsistent with the results of the correlation coefficients. However, the recognition result of the DNN-ESF is consistent with the results of the correlation coefficients, which indicates a high reliability. Although the recognition results of the DNN-FDF and DNN-WDF are the same, the final recognition result is consistent with those of the DNN-ESF, which are also the same as the real labels of these samples.

In addition, the 203DE data also belong to this case. Although the recognition result of the DNN-ESF is consistent with the real label of this sample, the credibility level of the DNN-ESF has the minimum value.

On the one hand, the recognition results of the DNN-FDF and DNN-WDF are the same. On the other hand, the credibility level of the DNN-WDF has a high value. Hence, the final recognition result is the same as the recognition results of the DNN-WDF. However, the recognition result of this sample does not match the actual label. The data are more similar to the samples marked with L3 than to those marked with L4, as shown in Fig. 4.

The last situation is when the three DNNs' recognition results are different from each other, but the DNNs' recognition results are also inconsistent with the corresponding correlation coefficients. There is no doubt that this situation is prone to misidentification. Hence, the results will not achieve a high credibility. The 201DE and 202DE data belong to this situation. It can be seen from the table that the recognition result of the 201DE samples does not match the real label.

Table 7 also shows that the recognition results of the 0.014-inch and 0.021-inch fault widths are less than ideal. To clarify the reasons for this result, we analyzed the spectra of these data. Parts of the data with 0.014-inch and 0.021-inch fault widths measured on FE are shown in Fig. 5 and Fig. 6, respectively. As seen from Fig. 5, 189FE has a significantly different spectrum from those of the other three ball fault data, and the 201FE data also exhibits the same phenomenon. Similarly, Fig. 6 shows that 213FE, 226FE, and 238FE have significantly different spectra from those of other data with the same fault types. We suspect that these data are mislabeled or dominated by other faults.

According to the above test, we determine that the proposed method has good generalization capabilities for nonhomologous data. Overall, our method has a better performance compared to a single DNN for both homologous data and nonhomologous data. However, in practical applications, to obtain good recognition results, it is necessary to establish a complete training set.

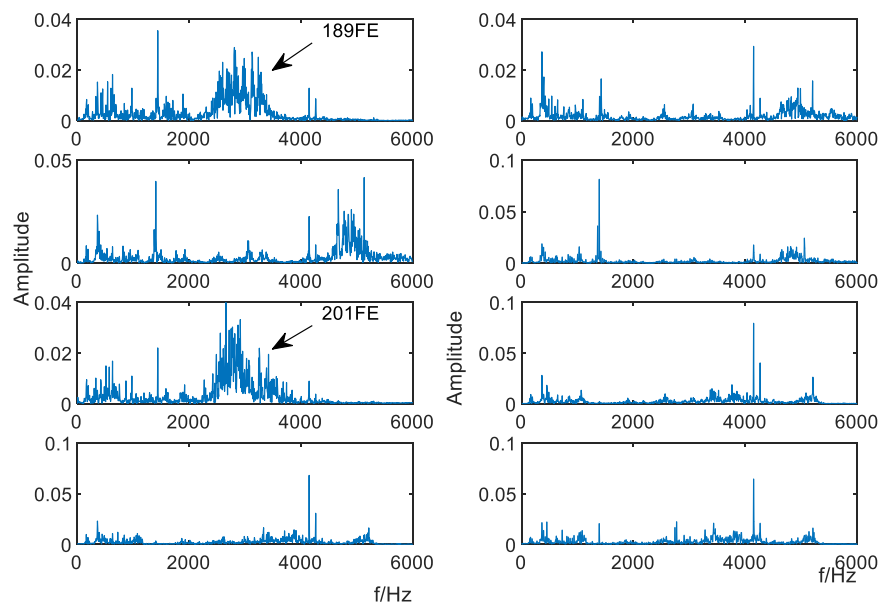


Fig. 5. The spectra of data with a 0.014-inch fault width measured on FE. According to the CWRU data description, the first four data have ball faults, and the last four data have OR faults.

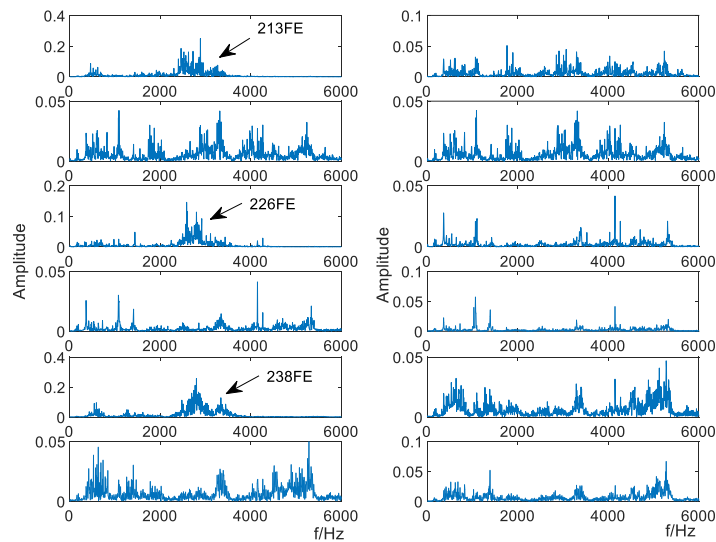


Fig. 6. The spectra of data with a 0.021-inch fault width measured on FE. According to the CWRU data description, the first four data have IR faults, the middle four data have ball faults, and the last four data have OR faults.

## 5 Conclusion

Based on an EDNN and correlation coefficients, an intelligent fault identification algorithm for rotating machinery fault diagnosis is proposed. Two measures are used to ensure that the proposed method has an excellent fault identification capability. One measure is that an EDNN composed of a DNN-FDF, DNN-WDF, and DNN-ESF is adopted to identify faults from the frequency domain, the wavelet domain and the envelope spectrum angle, respectively. The other measure is to evaluate the recognition results of DNNs combined with correlation coefficients and to use a fusion process for the evaluated results. The proposed method identifies fault signals from multiple angles using the EDNN and combines the results with correlation coefficients to make a comprehensive judgment to improve the diagnostic reliability.

The proposed method is tested with CWRU bearing data. The test results show that the proposed algorithm has an excellent generalization ability for homologous data and a good generalization ability for nonhomologous data. However, in engineering practice, the factors that cause equipment to operate abnormally are extremely complicated. The DNNs' outputs can be divided into more levels to facilitate a more detailed assessment. The proposed method also requires further testing on a large amount of real measured data. This is the direction of our future research.

## Compliance with Ethical Standards

The authors declare that they have no conflict of interest.

## References

- [1] Z. Q. Chen, S. C. Deng, X. D. Chen, C. Li, R. V. Sanchez, H. F. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectronics Reliability*, vol. 75, pp. 327–333, 2017.
- [2] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2000.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [4] C. Lu, Z. Y. Wang, W. L. Qin, J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, 2017.
- [5] R. Zhao, R. Q. Yan, Z. H. Chen, K. Z. Mao, P. Wang, R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2019.
- [6] K. Samir and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, 2018.
- [7] Y. G. Lei, B. Yang, X. W. Jiang, et al, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138: Article ID 106587, pp. 1–39, 2020.

- [8] M. Gan, C. Wang, and C. A. Zhu, "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.*, vol. 72-73, pp. 92–104, 2016.
- [9] C. Q. Shen, J. Q. Xie, D. Wang, X. X. Jiang, J. J. Shi, Z. K. Zhu, "Improved hierarchical adaptive deep belief network for bearing fault diagnosis," *Applied Sciences*, vol. 9 no. 16, pp. 1–18, 2019.
- [10] F. Jia, Y. G. Lei, J. Lin, X. Zhou, N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72-73, pp. 303–315, 2016.
- [11] Y. L. Yang, P. Y. Fu, and Y. C. He, "Bearing faults automatic classification based on deep learning," *IEEE Access*, vol. 6, no.1, pp. 71540–71554, 2018.
- [12] Y. L. Yang and P. Y. Fu, "Rolling-element bearing fault data automatic clustering based on wavelet and deep neural network," *Shock and Vibration*, vol. 2018, Article ID 3047830, pp. 1–11, 2018.
- [13] H. D. Shao, H. K. Jiang, X. Zhang, M. G. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, pp. 1–17, 2015.
- [14] F. N. Zhou, Y. L. Gao, and C. L. Wen, "A novel multimode fault classification method based on deep learning," *Journal of Control Science and Engineering*, vol. 2017, Article ID 3583610, pp. 1–14 2017.
- [15] Z. Xiang, X. N. Zhang, W. W. Zhang, X. R. Xia, "Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO spectrum and stacking auto-encoder," *Measurement*, vol. 138, pp. 162–174, 2019.
- [16] X. J. Guo, C. Q. Shen, and L. Chen, "Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery," *Applied Sciences*, vol. 7 no. 1, pp. 1–17, 2017.
- [17] Case Western Reserve University Bearing Data Center Website <<http://csegroups.case.edu/bearingdatacenter/home>>.
- [18] D. D. Peng, Z. L. Liu, H. Wang, Y. Qin, L. M. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019.
- [19] T. Y. Chen, Z. H. Wang, X. Yang, K. Jiang, "A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals," *Measurement*, vol. 148, Article ID 106857, pp. 1–15, 2019.
- [20] S. Ma and F. L. Chu, "Ensemble deep learning-based fault diagnosis of rotor bearing systems," *Computers in Industry*, vol. 105, pp. 143–152, 2019.
- [21] W. W. Qian, S. M. Li, P. X. Yi, K. C. Zhang, "A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions," *Measurement*, vol. 138, pp. 514–525, 2019.
- [22] Z. H. An, S. M. Li, J. R. Wang, Y. Xin, K. Xu, "Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method," *Neurocomputing*, vol. 352, pp. 42–53, 2019.
- [23] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R. E. Vásquez, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, 2015.
- [24] Z. Y. Chen and W. H. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, 2017.
- [25] M. Seeraa, M. L. D. Wong, and A. K. Nandi, "Classification of ball bearing faults using a hybrid intelligent model," *Applied Soft Computing*, vol. 57, pp. 427–435, 2017.
- [26] J. Yang, X. Q. Zeng, S. M. Zhong, S. L. Wu, "Effective neural network ensemble approach for improving generalization performance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 878–887, 2013.
- [27] L. L. Shi, "Correlation coefficient of simplified neutrosophic sets for bearing fault diagnosis," *Shock and Vibration*, vol. 2016, Article ID 5414361, pp. 1–11, 2016.
- [28] X. J. Guo, L. Chen, and C. Q. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [29] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, pp.59–66, 1988.
- [30] R. Taylor, "Interpretation of the correlation coefficient: a basic review," *Journal of Diagnostic Medical Sonography*, vol. 6, no.1, pp. 35–39, 1990.
- [31] W. A. Smith, R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64-65, pp. 100–131, 2015.
- [32] DeepLearnToolbox website <<https://github.com/rasmusbergpalm/DeepLearnToolbox>>.