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

Yanli Yang1 • Yichuan He1

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. Tests on some public bearing data sets show that the proposed algorithm can improve the reliability of fault identification.

Keywords 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 (Chen et al. 2017). 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 (Hinton and Salakhutdinov 2006) has become a useful tool for dealing with big data and has facilitated breakthroughs in the fields of speech recognition and image recognition (LeCun et al. 2015). 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 (Lu et al. 2017). This can bridge the gap between big data and machine intelligence health monitoring (Zhao et al. 2019; Samir and Yairi 2018; Lei et al. 2020). Note that a deep learning approach requires a lot of data and computational resources and provides black-box solutions, 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. (2016) proposed a two-layer hierarchical diagnosis network using deep belief networks (DBNs) and wavelet domain features to identify fault types and fault severities. Shen et al. (2019) proposed an improved hierarchical adaptive DBN for bearing fault diagnosis using frequency domain features. Using an automatic encoder and frequency domain features, Jia et al. (2016) proposed an intelligent diagnostic method based on a deep neural network (DNN). Using a DNN, Yang et al. (2018) proposed fault signal automatic classification methods in the frequency domain and the wavelet domain (Yang and Fu 2018). Shao et al. (2015) used a three-layer DBN to diagnose rolling-element bearing faults in the time domain. Zhou et al. (2017) proposed a method to identify failure modes with two automatic encoder networks connected in series using frequency domain features. Xiang et al. (2019) proposed a fault diagnosis method based on the Teager computed order spectrum and the stacked autoencoder (SAE) using frequency domain features. Lu et al. (2017) used a stacked denoising autoencoder (SDAE) to diagnose faults in the time domain. Guo et al. (2017) 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 (Case Western Reserve University Bearing Data Center 2022). In addition, a one-dimensional convolutional neural network (1D CNN) (Peng et al. 2019) and a deep capsule neural network (Chen et al. 2019) were also used to diagnose bearing faults in the time domain. Chen et al. (2017), 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 wavelet 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 (Ma and Chu 2019). The varying working conditions, such as the rotating speed oscillation or load variation, pose challenges for fault identification (Xiang et al. 2019; Qian et al. 2019; An et al. 2019). Any observer can see that analyzing the fault information requires multiple feature modalities (Chen et al. 2017; Ma and Chu 2019; Li et al. 2015). 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. Three deep neural network models—deep Boltzmann machines (DBMs), DBNs and SAEs—were employed to identify the fault conditions (Chen et al. 2017). By combining two layers of sparse self-encoding neural networks with a DBN, Chen and Li (2017) 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 (Ma and Chu 2019). A hybrid intelligent model consisting of a fuzzy min–max (FMM) neural network and a random forest model was proposed in (Seeraa et al. 2017) to classify ball bearing faults. A multimodal deep support vector classification approach was proposed in Li et al. (2015) 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, the combination of more than one deep learning model still cannot overcome the black-box problem of neural networks.

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. In practice, fault identification requires the help of more than one tool. For example, frequency spectrum and envelope spectrum are commonly used tools for bearing fault identification. Our motivation is to find a fault identification method from multiple angles using neural networks, so as to be closer to the practical applications. It was reported in Ma and Chu (2019) that ensemble learning can improve the generalization ability of the model ensemble. Theoretical and empirical work has shown that an ensemble neural network (ENN) often exhibits a better generalization performance compared to any individual neural network by itself (Yang et al. 2013). The ENN still gives a black-box solution, although it has an excellent performance in fault diagnosis. It is known that the correlation coefficient is an important tool for determining the correlation degree between samples (Shi 2016). Thus, we try to use the correlation coefficient to assist the ENN in fault diagnosis.
The main contribution of this paper is to propose a fault identification method based on ensemble deep neural networks (EDNNs) and correlation coefficients to improve the diagnostic reliability, which is inspired by the work in Ma and Chu (2019), Yang et al. (2013) and Shi (2016). This 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 (Yang et al. 2013). 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 Sect. 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 some public bearing data sets. Finally, the conclusions are drawn in Sect. 5.

2 Deep neural network and correlation coefficient

2.1 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 (Guo et al. 2016).

The neurons summarize the inputs and give the output through an activation function. Commonly used activation functions are sigmoid functions, rectified linear functions, hyperbolic tangent functions and so on. 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 Hinton and Salakhutdinov (2006) and LeCun et al. (2015).

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.

2.2 Brief review of the correlation coefficients

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 Rodgers and Nicewander (1988)

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

where $|r(Y, X)| \leq 1$, $\text{Cov}(X, Y)$ represents the covariance of $X$ and $Y$ and $\text{Var}[]$ 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 $Y$ (Taylor 1990). 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 relationship.
correlation between two variables or signals (Taylor 1990). 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.

| 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 (Chen et al. 2017; Gan et al. 2016; Jia et al. 2016; Shao et al. 2015; Zhou et al. 2017; Xiang et al. 2019; Guo et al. 2017; Ma and Chu 2019; Chen and Li 2017; Seeraa et al. 2017) 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 Chen et al. (2017) 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.

It is a common thought that the power of a neural network depends to a large extent on the data set. 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 Smith and Randall (2015). 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 in spectral line distribution and amplitude 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 outputs of DNNs.

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.

3.1 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 \((H^f_i, H^e_i)\), where \(H^f_i\) denotes the category label of the neuron with the largest output value in the output layer of the DNN-FDF and \(H^e_i\) represents the output of the corresponding neuron. Since the activation function of the output layer neurons is the sigmoid function, \(H^f_i\) takes a value between 0 and 1.

In the frequency domain, we calculate the correlation coefficients between a test signal and the training samples. Let \(C^f_i\) denote the maximum correlation coefficient, and let \(C^e_i\) represent the category label of the corresponding training sample. Suppose that \((S^f_i, S^e_i)\) denotes the assessment result of the DNN-FDF’s recognition results. The values of \(S^f_i\) and \(S^e_i\) are determined by \((H^f_i, H^e_i)\) and \((C^f_i, C^e_i)\). If \(H^f_i = C^f_i\), we then let \(S^f_i = H^f_i\) and have
A fault identification method based on an ensemble deep neural network and a correlation coefficient

Fig. 1 Flowchart of the proposed algorithm.

Fig. 2 Illustration of the difference between training and test samples. The training samples of types 1, 2 and 3 are the normal data, IR fault data and ball fault data, respectively.

\[ S_{r}^{f} = \begin{cases} 
5, & \text{if } |C_{r}^{f}| > 0.8 \text{ and } H_{r}^{f} > 0.8, \\
4, & \text{if } 0.6 < |C_{r}^{f}| \leq 0.8 \text{ and } H_{r}^{f} > 0.8, \\
3, & \text{if } (|C_{r}^{f}| > 0.4 \text{ and } 0.6 < H_{r}^{f} \leq 0.8) \text{ or } (0.4 < |C_{r}^{f}| \leq 0.6 \text{ and } H_{r}^{f} > 0.8), \\
2, & \text{if } (|C_{r}^{f}| > 0.2 \text{ and } 0.5 < H_{r}^{f} \leq 0.6) \text{ or } (0.2 < |C_{r}^{f}| \leq 0.4 \text{ and } H_{r}^{f} > 0.6), \\
1, & \text{others}. 
\]
Otherwise, if $H^r_w \neq C^r_e$, we still let $S^r_e = H^r_w$, but we let $S^w_e = 0$. 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 $S^r_e$ takes the value of the DNN-FDF’s output and $S^w_e$ represents the reliability of the DNN-FDF’s recognition results.

Similarly, we obtain the recognition result of the DNN-WDF, which is denoted as $(H^r_w, H^w_w)$, 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 $H^w_w \in [0, 1]$ 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 $C^w_e$, and determine the category label of the corresponding training sample, which is denoted as $C^w_e$. Then, we evaluate the output of DNN-WDF by considering the values of $(H^r_w, H^w_w)$ and $(C^w_e, C^w_e)$. The reassessment result is denoted as $(S^w_e, S^w_w)$. We have $S^w_w = H^w_w$, regardless of whether $H^r_w$ and $C^w_e$ are the same, but $S^w_e$ has a different value. If $H^r_w = C^w_e$, $S^w_w$ can be calculated by (2) with the superscript $f$ in (2) replaced by $w$. Otherwise, $S^w_w = 0$. 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 $(H^r_e, H^w_e)$, and the correlation coefficients $(C^r_e, C^w_e)$. Among the correlation coefficients calculated by the envelope features, we select the maximum correlation coefficient, which is denoted as $C^r_e$, and determine the category label of the corresponding training sample, which is denoted as $C^r_e$. Let $(S^r_e, S^w_e)$ 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 $S^w_e$ by (2) with the superscript $f$ in (2) replaced by $e$. Otherwise, $S^r_e = H^r_e$, but $S^w_e = 0$. The DNN-ESF’s results are also divided into six levels to indicate the reliability of the recognition results.

### 3.2 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 $(S^r_e, S^w_e)$, where $S^r_e$ denotes the type of fault identified and $S^w_e$ represents the credibility of the recognition result.

The value of $S^w_e$ is chosen among $S^r_e$, $S^w_w$, and $S^w_e$. The previous analysis shows that the values of $S^r_e$, $S^w_w$, and $S^w_e$ are determined by the outputs of the DNN-FDF, DNN-WDF and DNN-ESF, respectively. Hence, $S^w_e$ is one of the values output by three networks. For a test sample, if $S^r_e$, $S^w_w$ and $S^w_e$ 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 $S^w_e = 10$.

In this way, the credibility of the recognition result is divided into ten levels. The value of $(S^w_e, S^w_w)$ and the corresponding conditions are listed in Table 2. These rules listed in Table 2 are visually shown in “Appendix” Fig. 8. One point to note is that the values of vector $(S^r_e, S^w_e)$ 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.

After fusion processing, the outputs of the DNNs are re-evaluated. The effectiveness of the proposed algorithm is tested using some public bearing data sets, which is provided in the next section.

### 4 Testing and analysis

The proposed method is implemented using the DeepLearn Toolbox (2022). We first test the proposed method with the CWRU bearing data set and then validate it with more data sets. The CWRU bearing data was collected on normal bearings and bearings with single-point drive end (DE) and fan-end (FE) defects (Case Western Reserve University Bearing Data Center 2022). 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 were only collected at 12 kHz. To test with the same structure of the network, 48 k 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 a 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 wavelet packet 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 2 Fusion processing rules (the smaller the No. is, the higher the priority)

| No | Vector \((S_f, S_o)\) | Condition |
|----|----------------|-----------|
| 1  | \((S_o, 10)\)  | \(S_f = S_o \) & \(S_f > 1 \) & \(S_o = 0\) & \(S_o = 0\) |
| 2  | \((S_o, 9)\)    | \(S_f = S_o \) & \(S_f > 4\) & \(S_o = 0\) & \(S_o = 0\) |
| 3  | \((S_o, 9)\)    | \(S_o = S_o \) & \(S_o > 4\) & \(S_o = 0\) & \(S_o = 0\) |
| 4  | \((S_o, 8)\)    | \(S_o = 5\) |
| 5  | \((S_o, 8)\)    | \(S_o = 5\) |
| 6  | \((S_o, 8)\)    | \(S_o = 5\) |
| 7  | \((S_o, 7)\)    | \(S_f = S_o \) & \(2 < S_f < 4\) & \(2 < S_f < 4\) |
| 8  | \((S_o, 7)\)    | \(S_f = S_o \) & \(2 < S_f < 4\) & \(2 < S_f < 4\) |
| 9  | \((S_o, 7)\)    | \(S_f = S_o \) & \(3 < S_f < 4\) & \(3 < S_f < 4\) |
| 10 | \((S_o, 6)\)    | \(S_o = 4\) |
| 11 | \((S_o, 6)\)    | \(S_o = 4\) |
| 12 | \((S_o, 6)\)    | \(S_o = 4\) |
| 13 | \((S_o, 5)\)    | \(S_o = S_o\) |
| 14 | The vector corresponding to the maximum of \(S_f, S_o, S_o\) | \(S_f \neq S_o \neq S_o \neq 0\) & \(S_o > 0\) & \(S_o > 0\) & \(S_o > 0\) |
| 15 | \((S_o, 1)\)    | Others |

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 |

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

| Fault width (in.) | Training data set | Labels | Testing samples | Misidentified samples |
|-------------------|-------------------|--------|-----------------|-----------------------|
|                   | Normal + Fault    |        |                 |                       |
| 0.007             | 4 + 60            | 16     | 640             | 0                     |
| 0.014             | 4 + 36            | 10     | 400             | 1                     |
| 0.021             | 4 + 60            | 16     | 640             | 8                     |

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

| Fault width (in.) | Training data set | Labels | Testing samples | Misidentified samples |
|-------------------|-------------------|--------|-----------------|-----------------------|
|                   | Normal + Fault    |        |                 |                       |
| 0.007             | 4 + 40            | 11     | 220             | 0                     |
| 0.014             | 4 + 22            | 7      | 130             | 3                     |
| 0.021             | 4 + 40            | 11     | 220             | 0                     |

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 6 The test results for the 12 k fan-end bearing data

| Fault width (in.) | Training data set | Labels | Testing samples | Misidentified samples |
|-------------------|-------------------|--------|-----------------|-----------------------|
|                   | Normal + Fault    |        |                 |                       |
| 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](image_url) The recognition accuracies for the 12 k 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.

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

| Fault width (in.) | Measured position | Training data set | Labels | Testing samples | Misidentified samples |
|-------------------|-------------------|-------------------|--------|-----------------|-----------------------|
|                   | Normal + Fault    |                   |        |                 |                       |
| 0.007             | DE                | 4 + 20            | 6      | 120             | 42 52 14 8            |
|                   | 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 43           |
|                   | FE                | 4 + 20            | 6      | 120             | 13 10 9 6            |

© Springer
Table 8 The test results for the 48 k drive-end bearing data

| Sample | \( (H^r_i, H^c_i) \) | \( (C^r_i, C^c_i) \) | \( (S^r_i, S^c_i) \) | \( (H^r_o, H^c_o) \) | \( (C^r_o, C^c_o) \) | \( (S^r_o, S^c_o) \) | Real label |
|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------|
| 100FE  | (1, 0.995)      | (1, 0.799)      | (1, 4)          | (1, 0.996)      | (1, 0.92)       | (1, 5)          | (1, 0.989)  |
| 97FE   | (1, 0.995)      | (1, 0.81)       | (1, 5)          | (1, 0.996)      | (1, 0.893)      | (1, 5)          | (1, 0.996)  |
| 98FE   | (1, 0.996)      | (1, 0.894)      | (1, 5)          | (1, 0.996)      | (1, 0.976)      | (1, 5)          | (1, 0.819)  |
| 99FE   | (1, 0.995)      | (1, 1)          | (1, 5)          | (1, 0.996)      | (1, 1)          | (1, 5)          | (1, 0.996)  |
| 189DE  | (3, 0.995)      | (3, 0.641)      | (3, 4)          | (3, 0.994)      | (3, 0.789)      | (3, 4)          | (3, 0.992)  |
| 190DE  | (3, 0.993)      | (3, 0.567)      | (3, 3)          | (3, 0.994)      | (3, 0.813)      | (3, 5)          | (3, 0.993)  |
| 191DE  | (3, 0.796)      | (3, 0.556)      | (3, 3)          | (3, 0.994)      | (3, 0.844)      | (3, 5)          | (3, 0.992)  |
| 192DE  | (3, 0.99)       | (3, 0.603)      | (3, 4)          | (3, 0.996)      | (3, 0.883)      | (3, 5)          | (3, 0.989)  |
| 174DE  | (3, 0.805)      | (2, 0.484)      | (3, 0)          | (3, 0.787)      | (2, 0.765)      | (3, 0)          | (2, 0.97)   |
| 175DE  | (3, 0.75)       | (2, 0.463)      | (3, 0)          | (3, 0.679)      | (2, 0.721)      | (3, 0)          | (2, 0.992)  |
| 176DE  | (3, 0.647)      | (2, 0.489)      | (3, 0)          | (3, 0.876)      | (2, 0.76)       | (3, 0)          | (2, 0.993)  |
| 177DE  | (3, 0.845)      | (2, 0.408)      | (3, 0)          | (3, 0.843)      | (2, 0.632)      | (3, 0)          | (2, 0.992)  |
| 201DE  | (2, 0.774)      | (2, 0.568)      | (2, 3)          | (1, 0.986)      | (1, 0.441)      | (1, 3)          | (4, 0.973)  |
| 202DE  | (2, 0.809)      | (2, 0.496)      | (2, 3)          | (3, 0.991)      | (3, 0.805)      | (3, 5)          | (4, 0.937)  |
| 203DE  | (3, 0.861)      | (4, 0.523)      | (3, 0)          | (3, 0.991)      | (3, 0.753)      | (3, 4)          | (4, 0.588)  |
| 204DE  | (3, 0.717)      | (2, 0.515)      | (3, 0)          | (3, 0.994)      | (3, 0.777)      | (3, 4)          | (4, 0.799)  |

4.1 Test on CWRU data collected under same conditions

We consider that the training data set and the testing data set are collected under the same conditions. 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 12 k drive end, 12 k fan end, and 48 k drive end data, respectively. We pick 5 test samples for the 48 k 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 12 k-sampled data, and 16.67% of the samples as the training set and 83.33% of the samples as the test set for the 48 k-sampled data.

The test results for the 12 k 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 test result on the 48 k 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 data set 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 test results for the 12 k 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 test 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.

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

4.2 Test on CWRU data collected under different conditions

We train the three types of DNNs using the 12 k drive-end bearing data and test the trained DNNs with the 48 k drive-end bearing data. For the convenience of testing, the 48 k drive-end bearing data are resampled to a sampling frequency of 12 kHz. The 12 k and 48 k 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 training samples 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 48 k 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 by DNN-ESF, and it is much less than those by DNN-FDF and DNN-WDF.

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.

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 belong 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.

Table 9 Test results on some public data sets

| Data set | Training samples | Labels | Testing samples | Misidentified samples |
|----------|------------------|--------|-----------------|----------------------|
|          |                  |        |                 | DNN-FDF | DNN-WDF | DNN-ESF | Our method |
| PU       | 80               | 4      | 240             | 37      | 34      | 47       | 31         |
| FTP      | 20               | 5      | 575             | 199     | 169     | 235      | 116        |
| MFTP     | 7                | 5      | 16              | 6       | 10      | 5        | 0          |
| XJTU     | 15               | 5      | 451             | 15      | 32      | 43       | 5          |
| OU       | 9                | 3      | 27              | 1       | 1       | 4        | 1          |
| MFD      | 12               | 3      | 126             | 31      | 15      | 37       | 19         |
Another case is when all three DNNs’ recognition results are not 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 174 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 Figs. 5 and 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 exhibit 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. Note that it is necessary to establish a complete training set to obtain good recognition results in practical applications.

4.3 Test on other data sets

We test the proposed method using some public bearing data sets. These data sets include Paderborn University (PU) bearing data set (Paderborn University Bearing Data Center 2022), FTP bearing data set (Politecnico and di Torino 2022), Mechanical Failures Prevention Group (MFPT) fault data set (Mechanical Failures Prevention Group (MFPT) Society 2022), Ottawa University (OU) bearing data set (Huang and Baddour 2018), Xi’an Jiaotong University (XJTU) bearing data set (Wang et al. 2020) and Machinery Fault Database (MFD) (2022). The test results are listed in Table 9. We can see that the number of misidentified samples by our method is not greater than that by DNN-ESF, DNN-FDF and DNN-WDF, except for the DNN-WDF result test on the MFD data set. For the test on MFD data set, the number of misidentified samples by our method is slightly bigger than that by DNN-WDF, but it is much smaller than that by DNN-FDF and DNN-ESF. Therefore, we can say that our method has achieved good recognition results on these public data sets.

Although Table 9 shows that the proposed method can improve the recognition rate, there are two points need to be mentioned. The first point is that our method performs very well on some data sets such as the MFPT data set, but it shows mediocre performance on other data sets such as the OU data set. We hold the opinion there are two reasons, one is related to the data, and the other is that the model needs to be optimized. We look forward to improving it. Another point is that our method sometimes does not have a higher recognition rate than all single DNNs such as the test on the MFD data set. Nevertheless, overall, the proposed method uses the correlation coefficients to evaluate the outputs of DNNs and identifies fault signals from multiple angles, which is helpful for fault identification.
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 some public bearing data sets. 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 and more comparisons with the related works. These are the directions of our future research.

Appendix

See Figs. 7 and 8.
Acknowledgements  This work was supported in part by the Natural Science Foundation of Tianjin, China, under Grant 19JCYBJC16400.

Data availability  Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest  The authors declare that they have no conflict of interest.

References

An ZH, Li SM, Wang JR, Xin Y, Xu K (2019) Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method. Neurocomputing 352:42–53

Case Western Reserve University Bearing Data Center Website (2022) https://engineering.case.edu/bearingdatacenter

Chen ZY, Li WH (2017) Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. IEEE Trans Instrum Meas 66(7):1693–1702

Chen ZQ, Deng SC, Chen XD, Li C, Sanchez RV, Qin HF (2017) Deep neural networks-based rolling bearing fault diagnosis. Microelectron Reliab 75:327–333

Chen TY, Wang ZH, Yang X, Jiang K (2019) A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals. Measurement 148:1–15

DeepLearnToolbox website (2022). https://github.com/rasmusbergpalm/DeepLearnToolbox

Gan M, Wang C, Zhu CA (2016) 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 72–73:92–104

Guo XJ, Shen CQ (2016) Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 93:490–502

Guo XJ, Shen CQ, Chen L (2017) Deep fault recognizer: an integrated model to denoise and extract features for fault diagnosis in rotating machinery. Appl Sci 7(1):1–17

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507

Huang H, Baddour N (2018) Bearing vibration data collected under time-varying rotational speed conditions. Data Brief 21:1745–1749

Jia F, Lei YG, Lin J, Zhou X, Lu N (2016) Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 72–73:303–315

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444

Lei YG, Yang B, Jiang XW et al (2020) Applications of machine learning to machine fault diagnosis: a review and roadmap. Mech Syst Signal Process 138:1–39

Li C, Sanchez RV, Zurita G, Cerrada M, Cabrera D, Vásquez RE (2015) Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. Neurocomputing 168:119–127
Lu C, Wang ZY, Qin WL, Ma J (2017) Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. Signal Process 130:377–388
Ma S, Chu FL (2019) Ensemble deep learning-based fault diagnosis of rotor bearing systems. Comput Ind 105:143–152
MaFaulDa-Machinery Fault Database (2022). http://www02.smt.ufrj.br/~offshore/mfs/
Mechanical Failures Prevention Group (MFPT) Society Website (2022) https://mfpt.org/fault-data-sets/
Paderborn University Bearing Data Center Website (2022). https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/data-sets-and-download
Peng DD, Liu ZL, Wang H, Qin Y, Jia LM (2019) A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. IEEE Access 7:10278–10293
Qian WW, Li SM, Yi PX, Zhang KC (2019) A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions. Measurement 138:514–525
Rodgers JL, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. Am Stat 42(1):59–66
Samir K, Yairi T (2018) A review on the application of deep learning in system health management. Mech Syst Signal Process 107:241–265
Seeraa M, Wong MLD, Nandi AK (2017) Classification of ball bearing faults using a hybrid intelligent model. Appl Soft Comput 57:427–435
Shao HD, Jiang HK, Zhang X, Niu MG (2015) Rolling bearing fault diagnosis using an optimization deep belief network. Meas Sci Technol 26:1–17
Shen CQ, Xie JQ, Wang D, Jiang XX, Shi JJ, Zhu ZK (2019) Improved hierarchical adaptive deep belief network for bearing fault diagnosis. Appl Sci 9(16):1–18
Shi LL (2016) Correlation coefficient of simplified neutrosophic sets for bearing fault diagnosis. Shock Vib 2016:1–11
Smith WA, Randall RB (2015) Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. Mech Syst Signal Process 64–65:100–131
Taylor R (1990) Interpretation of the correlation coefficient: a basic review. J Diagn Med Sonogr 6(1):35–39
The Politecnico di Torino rolling bearing test rig (2022). ftp://ftp.polito.it/people/DIRG_BearingData/
Wang B, Lei YG, Li NP, Li NB (2020) A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Trans Reliab 69:401–412
Xiang Z, Zhang XN, Zhang WW, Xia XR (2019) Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO spectrum and stacking auto-encoder. Measurement 138:162–174
Yang YL, Fu PY (2018) Rolling-element bearing fault data automatic clustering based on wavelet and deep neural network. Shock Vib 2018:1–11
Yang J, Zeng XQ, Zhong SM, Wu SL (2013) Effective neural network ensemble approach for improving generalization performance. IEEE Trans Neural Netw Learn Syst 24(6):878–887
Yang YL, Fu PY, He YC (2018) Bearing faults automatic classification based on deep learning. IEEE Access 6(1):71540–71554
Zhao R, Yan RQ, Chen ZH, Mao KZ, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 115:213–237
Zhou FN, Gao YL, Wen CL (2017) A novel multimode fault classification method based on deep learning. J Control Sci Eng 2017:1–14

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.