

## **Research Article**

# Fault Diagnosis of Rolling Bearing Based on Modified Deep Metric Learning Method

Zengbing Xu<sup>(b)</sup>,<sup>1,2,3</sup> Xiaojuan Li,<sup>1,2</sup> Hui Lin<sup>(b)</sup>,<sup>4</sup> Zhigang Wang,<sup>1,2</sup> and Tao Peng<sup>5</sup>

<sup>1</sup>Key Laboratory of Metallurgical Equipment and Control Technology (Wuhan University of Science and Technology), Ministry of Education, Wuhan 430081, China

<sup>2</sup>Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering,

Wuhan University of Science and Technology, Wuhan 430081, China

<sup>3</sup>The State Key Lab. of Digital Manufacturing Equipment & Technology, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>4</sup>The 708 Institute of China State Shipbuilding Corporation Limited, Shanghai 200011, China

<sup>5</sup>Datang Interconnection Technology (Wuhan) Co., Ltd., Wuhan 430056, China

Correspondence should be addressed to Hui Lin; linhui@maric.com.cn

Received 16 October 2020; Revised 16 December 2020; Accepted 7 January 2021; Published 27 February 2021

Academic Editor: Riccardo Rubini

Copyright © 2021 Zengbing Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel fault diagnosis method of rolling bearing based on deep metric learning and Yu norm is proposed in this paper, which is called a deep metric learning method based on Yu norm (DMN-Yu). In order to solve the misclassification caused by the traditional deep metric learning based on distance metric function, a similarity criterion based on Yu norm is introduced into the traditional deep metric learning. Firstly, the deep metric learning neural network (DMN) is used to adaptively extract the fault feature parameters. Secondly, considering that the data samples at the boundary between different fault categories can be misclassified, the marginal Fisher analysis method based on Yu norm is used to optimize the features. And then, BPNN classifier of DMN-Yu method is used to fine tune the network parameters and diagnose the fault category. Finally, the effectiveness and feasibility of the proposed DMN-Yu method is verified with the rolling bearing fault diagnosis test. And the superiority of the proposed diagnosis method based on Euclidean distance (DMN-Euc), traditional deep belief network (DBN), and support vector machine (SVM) combined with the common time-domain statistical features.

#### 1. Introduction

The rolling bearing is one of the key components in rotating machinery, and its running state has an important influence on the health of rotating machinery. In order to prevent production losses and casualties, it is very necessary to monitor the running condition of rolling bearings and identify their fault categories [1]. The condition monitoring system is an effective tool to ensure the normal operation of the bearing. Owing to the long data acquisition time from the beginning of service to the end of life, the relatively large number of measuring points and the high sampling frequency, the mechanical big data can be obtained to reflect the health condition of bearings [2–4]. Therefore, it is of great significance to study how to effectively use the big data to diagnosis the bearing fault.

In recent years, many traditional intelligent diagnosis methods based on machine learning have been applied to the field of fault diagnosis. KNN, Bayes network, and others shallow-layer neural networks are all utilized to diagnose the different fault categories of mechanical equipment [5–9], but these ANN methods need the feature parameters which are manually extracted by diagnosis experience and signal processing method. Afterwards, deep learning [10, 11] is developed and applied to the field of fault diagnosis because of the ability of automatically extracting feature parameters and directly identifying faults from big data [2]. Some deep learning models are also designed to diagnose bearing faults; for example, deep belief network (DBN) is proposed to diagnose faults of aircraft engine and power transformer and rolling bearing [12, 13], and deep convolution neural network (DCNN) is also used to identify outer ring raceway faults and different severity degrees of lubricity faults [14]. Although these deep learning models have strong feature extraction capability to diagnose the fault categories with high accuracy, their diagnosis mechanism is unexplainable in the process of diagnosis. In addition, they have a poor ability to eliminate the overlapping region of different fault categories and can misclassify the faulty data sample at the boundary region between different fault categories [15].

Recently, metric learning has been proved to have the capacity to reduce intraclass scatter and interclass similarity within fault categories by measuring the distance or similarity between different data samples [16]. In view of the advantages of the deep learning and metric learning, deep metric learning (DML) methods are developed to improve the classification ability, which combines deep learning with metric learning to map original feature parameters to discriminative feature space by maximizing interclass variation and minimizing intraclass variation [17, 18]. Many typical DML models have been also proposed and applied to the field of pattern recognition. A discriminative deep metric learning (DDML) method is proposed for face recognition, which employed a fully connected deep neural network to learn multiple nonlinear transformations to map face samples into a discriminative distance space in which the similarity of each positive pair increased and the similarity of each negative pair decreased [19]. Hu et al. [20] suggested a deep transfer metric learning (DTML) method for face recognition to learn the deep metric network by maximizing interclass variation and minimizing intraclass variation. Song et al. [21] proposed a DML method for vision recognition and employed the lifted structure feature embedding to learn semantic feature embeddings, and the similar samples are mapped close to each other and dissimilar samples are far from each other. An adaptive interval DML method is proposed for video classification, which can adaptively allocate the distance according to semantic distance between sample pairs [16]. Obviously, these DML methods can use the distance criterion to classify these face images with high accuracy, but they have not strong ability to discriminate between fuzzy face images.

In the field of fault diagnosis, some mechanical signals are very difficult to be classified because of the complexity of signal transmission path and insensitiveness to fault classes of parameter features, and especially, some data samples in the boundary region between different fault categories can be misclassified [22]. Fuzzy operators have the ability to deal with various fuzzy sets. A novel clustering method that combines the adaptive resonance theory (ART) with the similarity measure based on Yu's norm is proposed to diagnose the faults of rolling element bearings, which can recognize faulty data samples in the boundary region through the fuzzy formalism [23]. But it needs to manually extract feature parameters from original data samples.

As mentioned above, a deep metric learning method based on Yu norm- (DMN-Yu-) based similarity measure is

proposed to diagnose mechanical faults, which use the Yu norm-based similarity measure to calculate the similarity between different data samples and use the deep network architecture to extract the feature parameter and built the nonlinear relation between data sample and fault category. The rest of this paper is organized as follows. Section 2 reviews the deep neural network and deep metric learning. A deep metric learning method based on Yu norm is depicted in Section 3. Section 4 describes the details of DMN-Yu algorithm, and the diagnosis analysis of bearing fault is conducted in Section 5. Finally, the conclusions are drawn in Section 6.

#### 2. Theory of Deep Metric Learning

2.1. Deep Neural Network. Deep neural network is a deep structure with multiple hidden layers, which can learn the hierarchical feature representation by constructing the network architecture from the low-level layer to high-level layer. Thus, some feature parameters can be extracted directly from the original data automatically. Generally, a deep neural network consists of an input layer, multiple hidden layers, and an output layer, and the network architecture is shown in Figure 1. Assume that there is a deep neural network with N + 1 layers and  $P^{(n)}$  units in the *n*th layer, where the parameters  $n \in [1, 2, ..., N]$ ,  $(W, b) = (W^1, b^{(1)}, W^{(2)}, b^{(2)}, ..., W^{(N)}, b^{(N)})$ . Given a data sample set  $X \in \mathbb{R}^d$ , the input of the first layer is X, and its corresponding output  $h^{(1)}$  can be expressed as follows:

$$Z^{(1)} = W^{(1)}X + b^{(1)}, (1)$$

$$h^{(1)} = s(Z^{(1)}) \in R^{P^{(1)}},$$
(2)

where  $Z^{(1)}$  is the weighted sums of the input of the first layer,  $W^{(1)} \in \mathbb{R}^{P^{(1)} \times d}$  is the projection matrix learned in the first layer,  $b^{(1)} \in \mathbb{R}^{P^{(1)}}$  is the bias vector in the first layer, and *s* is the nonlinear activation function of each layer, which can be a sigmoid function or a tanh function. The tanh function was used as the nonlinear activation function in this paper.

Then, the output  $h^{(1)}$  of the first layer is used as the input of the second layer, and the output  $h^{(2)}$  of the second layer can be computed as

$$Z^{(2)} = W^{(2)}h^{(1)} + b^{(2)},$$
(3)

$$h^{(2)} = s(Z^{(2)}) \in R^{P^{(2)}}, \tag{4}$$

where  $Z^{(2)}$  is the weighted sums of the input of the second layer and  $W^{(n)} \in \mathbb{R}^{P^{(2) \times P(1)}}, b^{(2)} \in \mathbb{R}^{P^{(2)}}$ , and *s* are the projection matrix, bias, and activation function of the second layer, respectively.

Similarly, the output of the  $n^{\text{th}}$  layer can be written as follows:

$$Z^{(n)} = W^{(n)} h^{(n-1)} + b^{(n)},$$
(5)

$$h^{(n)} = s(Z^{(n)}) \in R^{P^{(n)}}, \tag{6}$$



FIGURE 1: The network architecture used in the method. *X* is the input of the network,  $h^{(1)} - h^{(N-1)}$  are the output of the hidden layers,  $h^{(N)}$  is the output of the top layer, and  $W^{(n)}$  and  $b^{(n)}$  are the parameters of the network to be learned,  $1 \le n \le N$ .

and the output of the top layer is

$$Z^{(N)} = W^{(N)} h^{(N-1)} + b^{(N)}, (7)$$

$$f^{(x)} = h^{N} = s(Z^{(N)}) \in R^{P^{(N)}},$$
(8)

where mapping function  $f: \mathbb{R}^d \mapsto \mathbb{R}^{P^{(N)}}$  is a parametric nonlinear equation determined by the parameters  $W^{(n)}$  and  $b^{(n)}, n \in [1, 2, ..., N]$ .

2.2. Deep Metric Learning. Deep metric learning can map original data sample into other feature spaces through a set of hierarchical nonlinear transformations, which use the deep neural network structure to integrate feature learning and metric learning into a joint learning framework [24] and to explicitly built the nonlinear mapping function f, and f is initialized by  $W^{(n)}$  and  $b^{(n)}$ . If f(X) is the output of input X, the Euclidean distance of the data points  $x_i$  and  $x_j$  in the deep metric network space is written as follows:

$$d_{f}(x_{i}, x_{j}) = d(f(x_{i}) - f(x_{j})) = \left\| f(x_{i}) - f(x_{j}) \right\|_{2}, \quad (9)$$

where the goal of the deep metric learning is to establish the mapping function f under some certain constraints.

### 3. Deep Metric Learning Model Based on Yu Norm

Metric learning is mainly to measure the distance or similarity between different data samples so that the data sample of the same fault class is as close as possible to each other, and the data sample of different fault classes is as far as possible from each other by metric criterion. Therefore, it is very important to select appropriate metric criteria before conducting classification. However, most existing metric learning methods can only map data samples into feature space through a linear transformation, which makes it difficult to capture the nonlinear relationship between data samples [25]. In order to solve this problem, a kernel-based nonlinear metric method is developed to map data samples to high-dimensional feature space and to perform discrimination distance metric in this high-dimensional feature space [26, 27]. However, the nonlinear mapping function of the nonlinear metric method based on kernel function is not explicit, which is not conducive to understanding and development. In addition, these two metric learning methods need to extract the feature parameters manually. Different from the above two metric learning methods, the Yu-normbased deep metric learning method can extract feature parameters automatically and establish a set of hierarchical nonlinear transformations through deep neural network and map data sample pairs into a feature space for recognition.

3.1. Yu Norm-Based Similarity Measure Criterion. Yu norm is proposed in 1985, which is composed of the *T* norm and the *S* norm in the field of fuzzy mathematics [28], and the *T* norm and the *S* norm are, respectively, defined as follows:

$$T(x, y) = \max[0, (1 + \lambda)(x + y - 1) - \lambda xy],$$
(10)

$$Sn(x, y) = \min[1, x + y + \lambda xy], \qquad (11)$$

where all  $x, y \in [0, 1]$ ,  $\lambda > -1$ . Because of the unique mathematical property of Yu norm, an equivalence expression which is depicted as follows:

$$E(x, y) = T(Sn(\overline{x}, y)), (Sn(x, \overline{y})),$$
(12)

is suggested to compute the similarity between two different data samples [29], and based on the similarity degree, the data samples can be classified.

In the process of classification, the data samples in the overlapping region of the different fault categories can be misclassified easily by the traditional distance metric method because of the nonlinearity of the classification boundary line. However, the basis of classification of the Yu normbased similarity metric method depends on similarity between different data samples rather than the distance. Thus, the Yu norm-based similarity metric method can classify the data samples in the boundary region of fault categories.

3.2. Marginal Fisher Analysis. Marginal Fisher analysis (MFA) is a supervised descendent dimension algorithm that measures the distance between every data sample and its neighbor samples [30], which require the data samples of every category to satisfy approximately the Gauss distribution [31]. Based on the graph embedding framework, MFA can be depicted by two graphs, as shown in Figure 2: an intrinsic graph and a penalty graph. The intrinsic graph characterizes the intraclass compactness, and the penalty graph characterizes the interclass separation. In the intrinsic graph, an adjacency matrix P can be established by the data sample, which is only connected with its k1-nearest neighbors of the same class. In the penalty graph, an adjacency matrix Q can be also established by the sample, which is connected with its k2nearest neighbors of the different class. The matrix P or Q can be expressed as follows:

$$P_{ij} \text{ or } Q_{ij} = \begin{cases} 1, & x \text{ is connected to } x_j, \\ 0, & x_i \text{ is not connected to } x_j. \end{cases}$$
(13)



FIGURE 2: The adjacency relationship of the intrinsic graph and the penalty graph of the marginal Fisher analysis (k1 = k2 = 3): (a) intrinsic graph; (b) penalty graph.

The separation rule of MFA can be described that each sample is drawn close to k1-nearest neighbors of its similar samples and pushed away from k2-nearest neighbors of its dissimilar samples. Based on the graph embedding framework, the intraclass compactness  $S_c$  described by the intrinsic graph and the interclass separability  $S_b$  described by the penalty graph can be obtained, respectively.

Accordingly, after that the metric learning is introduced into the deep neural network, the marginal Fisher analysis is used to optimize the features extracted in the feature space of top layer, which improves the feature representation ability and the classification performance of the deep neural network.

3.3. DMN-Yu Model. For each pair samples  $x_i$  and  $x_j$  in the training sample set X, the output of the *n*th layer of the network can be represented as  $f^{(n)}(x_i)$  and  $f^{(n)}(x_j)$  as mentioned above. The similarity of the two samples can be expressed by Yu norm-based similarity measure criterion as follows:

$$d_{f^{(n)}}(x_i, x_j) = S\langle f^{(n)}(x_i), f^{(n)}(x_j) \rangle, \qquad (14)$$

where  $S\langle x_i, x_j \rangle = \max(0, (1+\lambda)(Sn(\overline{x}_1, x_j) + Sn(x_i, \overline{x}_j) - 1)) - \lambda Sn(\overline{x}_1, x_j)Sn(x_i, \overline{x}_j))$   $Sn(x_i, x_j) = \min[1, x_i + x_j + \lambda x_i x_j], \overline{x}_I = 1 - x_i$ , and  $\lambda$  is a constant.

According to the graph embedding framework, the MFA is performed on the output of all the training samples at the top layer of deep neural network, and a strongly supervised deep metric learning model is constructed correspondingly. The parameters of deep metric learning model based on Yu norm (DMN-Yu) can be obtained by the optimization of the following objective function. Thereinto, the parameters  $W^{(n)}$  and  $b^{(n)}$  can be obtained by the gradient descent algorithm:

$$\min_{f^{(N)}} J = S_c^{(N)} - \alpha S_c^{(N)} + \gamma \sum_{n=1}^N \left( \left\| W^{(n)} \right\|_F^2 + \left\| b^{(n)} \right\|_2^2 \right), \quad (15)$$

where  $\alpha$  is the free parameter which balances the importance between intraclass compactness and the interclass separability. The larger the  $\alpha$  is, the greater the interclass scatter is;  $\gamma$ is the adjustable regularization parameter,  $\gamma > 0$ ;  $||Z||_F$  denotes the Frobenius norm of the matrix Z;  $S_c^{(n)}$  and  $S_b^{(n)}$  define the intraclass compactness and interclass separability, respectively, and its formula can be written as follows:

$$S_{c}^{(n)} = \frac{1}{Mk_{1}} P_{ij} d_{f^{(n)}}(x_{i}, x_{j}), \qquad (16)$$

$$S_b^{(n)} = \frac{1}{Mk_2} Q_{ij} d_{f^{(n)}} (x_i, x_j), \qquad (17)$$

where *M* is the number of samples in the training set and *P* and *Q* are adjacency matrixes. If  $X_j$  is one of the *k*1-intraclass nearest neighbors of  $X_i$ , then  $P_{ij}$  is set to 1 and otherwise 0; if  $X_j$  is one of the *k*2-interclass nearest neighbors of  $X_i$ ,  $Q_{ij}$  is set to 1 and otherwise 0.

To solve the problem of parameters optimization in equation (15), the subgradient descent method is used to get the parameters  $\{W^{(n)}, b^{(n)}\}$ , where n=1, 2, ..., N. The gradient of the objective function J with respect to the parameters  $W^{(n)}$  and  $b^{(n)}$  is computed as follows:

$$\frac{\partial J}{\partial W^{(n)}} = \frac{1}{Mk1} P_{ij} \left( L_{ij}^{(n)} h_i^{(n-1)T} + L_{ji}^{(n)} h_j^{(n-1)T} \right) - \frac{\alpha}{Mk2} Q_{ij} \left( L_{ij}^{(n)} h_i^{(n-1)T} + L_{ji}^{(n)} h_j^{(n-1)T} \right) + 2\gamma W^{(n)},$$
(18)

$$\frac{\partial J}{\partial b^{(n)}} = \frac{1}{Mk1} P_{ij} \left( L_{ij}^{(n)} + L_{ji}^{(n)} \right) - \frac{\alpha}{Mk2} \sum_{i=1}^{M} \sum_{j=1}^{M} Q_{ij} \left( L_{ij}^{(n)} + L_{ji}^{(n)} \right) + 2\gamma b^{(n)},$$
(19)

where  $h_i^{(0)} = x_i$  and  $h_j^{(0)} = x_j$  are the original input samples of the network, for all other layers n = 1, 2, ..., N-1, and the updating equations are as follows:

$$L_{ij}^{(N)} = \left(h_i^{(N)} - h_j^{(N)}\right) \odot \varphi'(Z_i^{(N)}),$$
(20)

$$L_{ji}^{(N)} = \left(h_{j}^{(N)} - h_{i}^{(N)}\right) \odot \varphi'\left(Z_{j}^{(N)}\right), \tag{21}$$

$$L_{ij}^{(n)} = \left(W^{(n+1)T} - L_{ij}^{(n+1)}\right) \odot \varphi'\left(Z_i^{(n)}\right),$$
(22)

$$L_{ji}^{(n)} = \left(W^{(n+1)T} - L_{ji}^{(n+1)}\right) \odot \varphi'\left(Z_{j}^{(n)}\right),$$
(23)

where the operation  $\odot$  denotes the element-wise multiplication and  $W^{(n)}$  and  $b^{(n)}$  can be updated by the following gradient descent algorithm until convergence:

$$W^{(n)} = W^{(n)} - \tau \frac{\partial J}{\partial W^{(n)}},\tag{24}$$

$$b^{(n)} = b^{(n)} - \tau \frac{\partial J}{\partial W^{(n)}},\tag{25}$$

where  $\tau$  is the learning rate.

3.4. Classifier Selection. Backpropagation neural network (BPNN) is an effective classifier for multiclassification

problems. Therefore, a BPNN classifier which is introduced into the top feature output layer of the deep metric neural network is utilized to further fine tune the parameters of the network [32, 33] and classify data samples.

#### 4. Fault Diagnosis Method Based on DMN-Yu

The DMN-Yu model is trained by the backpropagation algorithm, and the BPNN classifier is utilized to fine tune the network parameter and to complete the classification of faults. The flowchart of the proposed method is shown in Figure 3. The corresponding algorithm is summarized as follows:

Step 1: normalize a large number of labeled data samples and divide them into the test set and training set according to a certain proportion.

Step 2: construct the DMN-Yu model, set the parameters  $W^{(n)}$  and  $b^{(n)}$  to a number close to 0, and set the iteration number *T*, the intraclass nearest neighbor *k*1, the interclass nearest neighbor *k*2, and the learning rate  $\tau$ ,  $\lambda$ ,  $\gamma$ , and  $\alpha$ .

Step 3: Train the DMN-Yu model, and the data samples are nonlinearly transformed to the top layer and are made by the marginal Fisher analysis to further constrain the extracted features so that the descriptive and discriminative features can be obtained.

Step 4: Add a BPNN classifier to the feature output layer at the top of the network to fine tune the network and optimize the network parameters.

Step 5: use the trained network to classify the test dataset.

Step 6: calculate the classification accuracy.

#### 5. Case Analysis

5.1. Experimental Dataset. In order to verify the effectiveness of the proposed method, the DMN-Yu model was used to diagnose the rolling bearing data [34] from Electrical Engineering Laboratory of Case Western Reserve University. The data are collected on the deep groove ball bearing mounted on the motor drive end type SKF6205-2RS. There are four types bearing faults, namely, normal state, inner race fault, outer race fault, and ball fault, bearings were seeded with faults using electrodischarge machining, each fault has three different depths of damage, and Table 1 shows the description of the bearing dataset. The data are collected by the accelerometer at a sampling frequency of 12 KHz under the operating condition of 0 HP load and 1797 rpm, and the dataset contains 10 fault types. The number of samples for each fault type is 700, a total of 7000 samples, and a single sample has 512 sample points, where the number of samples in the training set is 4900 and the number of samples in the test set is 2100.

5.2. Fault Diagnosis and Analysis. A DMN-Yu model with 3layer (N=2) network is constructed, and the number of neural nodes in each layer was set as 512-100-100,

respectively. BPNN classifier was added to the top-level feature output layer that the classification result output layer nodes were set as 10. It is very important to select proper model parameters for obtaining better diagnosis accuracy. Here,  $\alpha$  was set as 4.0,  $\lambda$  was 0.2, the maximum number of iterations was 10, the regularization parameter  $\gamma$  was 0.5, the initial learning rate  $\tau$  was 0.2, and the learning rate decay was 0.95. The number of the neighbor points k1 and k2 has a greater impact on the diagnosis ability of the model. If the number of the neighbor points is too small, it is difficult for DMN-Yu model to dig out the intrinsic fault information from the high-dimensional data, and if the number of neighbor points is too large, geometric information and nonlinear information of data are easily ignored, so the nearest neighbor values k1 = 5 and k2 = 10 according to the reference [30].

In order to obtain the specific diagnosis information of the model for each fault category, the confusion matrix was used to visualize the diagnosis results, and the diagnosis results were quantitatively described using the precision Pand the recall R [35]. Figure 4 shows the diagnosis results of the training samples. From Figure 4, it can be seen that the diagnosis accuracy of each fault category is 100%. It indicates that the proposed method can extract the fault feature parameter of bearing and diagnose the fault category effectively, and there are no training errors.

In addition, to describe the diagnosis process of the DMN-Yu model, t-SNE (t-distributed stochastic neighbor embedding) algorithm is used to visualize the high-dimensional data in the 3D (3-dimensional) space by mapping samples from the original feature space to 3D space. The fault features extracted from each layer of DMN-Yu model are optimized and mapped to the 3D space by t-SNE [36]. And the corresponding 3D feature distribution scatter diagrams of the trained DMN-Yu model are obtained. Figures 5-7 show the visualization of the original data, the first hidden layer output feature, and the second hidden layer output feature, respectively. It can be seen that the interclass distance between the features of different fault categories becomes larger and larger and the intraclass distance becomes smaller and smaller with the rising of layer number, and at the top feature output layer, all fault categories can be basically separated. All these indicate that the DMN-Yu model can perform feature extraction through minimizing the intraclass distance and maximizing the interclass distance, and the subsequent classification accuracy can get higher and higher.

The diagnosis result of the bearing fault test set samples is shown in Figure 8. It can be seen that the classification accuracy is 97% and error rate is only 3%. The faults number 1 and fault number 10 are all classified correctly, and their precision P and recall R are all 100%. For other fault numbers, most of them are also correctly classified and recognized, but 15 samples of the fault number 2 are misclassified. Thereinto, 4 samples are misclassified as fault number 3, 2 samples were misclassified as fault number 4, 2 samples were misclassified as fault number 5, 5 samples were misclassified as fault number 6, 2 samples were misclassified as fault number 10, and the recall rate R is 92.9%. 2 samples



FIGURE 3: Flowchart of the proposed method.

TABLE	1:	Description	of	bearings	dataset.
				· · · · · · · · · · · · · · · · · · ·	

Fault number	Fault condition type	Depth of fault (mm)	Training samples	Test samples
1	Normal	0	490	210
2	Mild inner race	0.18	490	210
3	Moderate inner race	0.36	490	210
4	Severe inner race	0.54	490	210
5	Mild ball	0.18	490	210
6	Moderate ball	0.36	490	210
7	Severe ball	0.54	490	210
8	Mild outer race	0.18	490	210
9	Moderate outer race	0.36	490	210
10	Severe outer race	0.54	490	210

of fault number 5 were misidentified as fault number 2, and the corresponding precision P of fault number 2 is 99.0%. For the samples of fault number 7, there was 1 sample

misclassified as fault number 10, 1 sample was misclassified as fault number 6, and the corresponding recall R is 99.0%. For fault number 9, there are 2 samples misidentified as fault

#### Shock and Vibration

1	490	0	0	0	0	0	0	0	0	0	100%
	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0	490	0	0	0	0	0	0	0	0	100%
	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0	0	490	0	0	0	0	0	0	0	100%
	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0	0	0	490	0	0	0	0	0	0	100%
	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0	0	0	0	490	0	0	0	0	0	100%
9	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
dictive sta	0	0	0	0	0	490	0	0	0	0	100%
	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ord	0	0	0	0	0	0	490	0	0	0	100%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0.0%
8	0	0	0	0	0	0	0	490	0	0	100%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%
9	0	0	0	0	0	0	0	0	490	0	100%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%
10	0	0	0	0	0	0	0	0	0	490	100%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1	2	3	4	5	6 Actual state	7	8	9	10	

FIGURE 4: Confusion matrix of fault recognition for rolling bearing training samples based on the DMN-Yu model.



FIGURE 5: 3D feature scatter diagram of the raw data.



FIGURE 6: 3D feature scatter diagram of the first hidden layer.



FIGURE 7: 3D feature scatter diagram of the second hidden layer.

number 7, and its precision P is 99.0%. Correspondingly, the classification accuracies of other fault categories can also be seen in Figure 8. All these can demonstrate that the DMN-Yu model can diagnose the different fault categories of bearings accurately.

5.3. Diagnosis Performance Comparison with Other Methods. In order to evaluate the superiority of the DMN-Yu model, the deep metric learning method based on Euclidean distance (DMN-Euc), the traditional DBN, and the traditional support vector machine (SVM) combined with extracted manually feature parameters are used to diagnose the fault category of bearings. Without loss of generality, 10 test runs are conducted by these methods and the average results are shown in Table 2. It can be seen that the diagnosis accuracy of the proposed DMN-Yu model is 96.76%. The average diagnosis accuracy of DMN-Euc model is 95.72%, which is slightly lower than that of the DMN-Yu method. It is demonstrated that diagnosis accuracy of the DMN-Yu based on the Yu norm-based similarity measure criterion is higher than that of DMN-Euc based on the Euclidean distance. Meanwhile, the table shows that the average diagnosis accuracy of the traditional DBN is 94.10%, and the accuracy of

#### Shock and Vibration

1	210	0	0	0	1	0	0	0	2	0	98.6%
	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	1.4%
2	0 0.0%	195 9.3%	0	0	2 0.1%	0	0	0	0	0	99.0% 1.0%
3	0 0.0%	4 0.2%	201 9.6%	0	2 0.1%	3 0.1%	0	2 0.1%	0	0	94.8% 5.2%
4	0	2	0	207	3	0	0	0	2	0	96.7%
	0.0%	0.1%	0.0%	9.9%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	3.3%
5 5	0 0.0%	2 0.1%	0	0 0.0%	198 9.4%	0	0 0.0%	0	1 0.0%	0 0.0%	98.5% 1.5%
edictive sta	0	5	1	1	1	199	1	0	0	0	95.7%
	0.0%	0.2%	0.0%	0.0%	0.0%	9.5%	0.0%	0.0%	0.0%	0.0%	4.3%
Dre	0	0	0	0	0	0	208	0	2	0	99.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9.9%	0.0%	0.1%	0.0%	1.0%
8	0	0	0	0	0	0	0	207	0	0	100%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9.9%	0.0%	0.0%	0.0%
9	0	0	0	0	3	0	0	0	203	0	98.5%
	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	9.7%	0.0%	1.5%
10	0 0.0%	2 0.1%	8 0.4%	2 0.1%	0	8 0.4%	1 0.0%	1 0.0%	0 0.0%	210 10.0%	90.5% 9.5%
	100%	92.9%	95.7%	98.6%	94.3%	94.8%	99.0%	98.6%	96.7%	100%	97.0%
	0.0%	7.1%	4.3%	1.4%	5.7%	5.2%	1.0%	1.4%	3.3%	0.0%	3.0%
	1	2	3	4	5	6 Actual state	7 e	8	9	10	

FIGURE 8: Confusion matrix of fault recognition for the rolling bearing test set based on DMN-Yu model.

TABLE 2: Average diagnosis results of 10 experiments.

Method	Average diagnostic accuracy (%)	Method	Parameter value	
DMN-Yu	96.76		The number of neural nodes in each layer 512-100-100-	
DMN-Euc	95.72	DMN-	10; $a = 4.0$ ; the maximum number of iterations $T = 10$ ;	
DBN	94.10	Euc	$\gamma = 0.5$ ; $\tau = 0.2$ ; learning rate decay was 0.95; $k_1 = 5$ ;	
Common statistical	05.05		$k_2 = 10$	
features + SVM	85.05	DBN	The number of neural nodes in each layer 512-100-1	
			10; learning rate $\tau = 0.0001$	

the SVM combined with nine time-domain statistical features which are mean, standard deviation, root mean square, skewness, kurtosis, crest factor, margin factor, shape factor, and impact factor is 85.05%. It is obvious that the accuracies of the DBN and the traditional SVM are lower than that of DMN-Yu and DMN-Euc, and the accuracy of DMN-Yu is the highest. All these demonstrated that the performance of DMN-Yu and DMN-Euc is superior to DBN and SVM, and the DMN-Yu is the best. In addition, the parameter value of these methods is set as described in Table 3.

TABLE 3: Parameter settings of several methods.

Method	Parameter value
	The number of neural nodes in each layer 512-100-100-
DMN-	10; $a = 4.0$ ; the maximum number of iterations $T = 10$ ;
Euc	$\gamma = 0.5; \tau = 0.2;$ learning rate decay was 0.95; $k_1 = 5;$
	$k_2 = 10$
DBN	The number of neural nodes in each layer 512-100-100-
DDN	10; learning rate $\tau = 0.0001$
SVM	Polynomial kernel function and the order is set as 2; 1
	vs all classification mechanism

#### 6. Conclusion

A novel deep metric learning based on Yu norm is proposed to diagnose the fault of the rolling bearings, which can measure the similarities as well as differences between data samples and improve diagnosis ability by reducing intraclass scatter and interclass similarity. Due to the fuzziness of the boundary of the fault categories, the data samples at the boundary are easy to be misclassified. To solve this problem, the similarity measure based on Yu norm is used to compute the similarity between different samples, and the marginal Fisher analysis method based on Yu norm was used to optimize features. And in order to integrate feature extraction and classification recognition into a model based on the deep metric learning framework, BPNN classifier is used to fine tune the entire network and complete the fault classification. The effectiveness of the proposed method was verified by the diagnosis analysis of the bearing dataset. The results testify that the proposed method can diagnose the bearing fault effectively and it is superior to the deep metric learning model based on Euclidean distance (DMN-Euc), DBN fault, and SVM.

#### Abbreviations

DMN-Yu:	Deep metric learning method based on Yu
	norm
DMN:	Deep metric learning neural network
BPNN:	Backpropagation neural network
DMN-Euc:	Deep metric learning method based on
	Euclidean distance
DBN:	Deep belief network
SVM:	Support vector machine
DML:	Deep metric learning
DDML:	Discriminative deep metric learning
DTML:	Deep transfer metric learning
ART:	Adaptive resonance theory
MFA:	Marginal Fisher analysis
t-SNE:	t-distributed stochastic neighbor embedding

#### **Data Availability**

The data used to support the findings of this study are available at http://csegroups.case.edu/bearingdatacenter/ pages/download-data-file.

#### **Conflicts of Interest**

The authors declare that there are no conflicts of interest.

#### Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 51775391), Open Research Foundation of State Key Lab. of Digital Manufacturing Equipment and Technology in Huazhong University of Science and Technology (Grant no. DMETKF2017010), Equipment Pre-Research Foundation (Grant no. 6142223180312), WUST National Defense Pre-Research Foundation (Grant no. GF201902), and Wuhan Science and Technology Plan Project (Grant no. 2019010702011280).

#### References

 D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.

- [2] Y.-G. Lei, F. Jia, W. Zhou et al., "A health monitoring method for mechanical equipment big data based on deep learning theory," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 49–56, 2015.
- [3] X. Su, Y.-Y. Wu, H.-J. Pei et al., "Progress in the application of big data technology in process industry," *Chemical Industry Progress*, vol. 35, no. 6, pp. 1652–1659, 2016.
- [4] B. Hopkins, B. Evelson, S. Leaver et al., Expand Your Digital Horizon with Big Data, Forrester, Cambridge, MA, USA, 2011.
- [5] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4137–4145, 2013.
- [6] V. Muralidharan and V. Sugumaran, "A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis," *Applied Soft Computing*, vol. 12, no. 8, pp. 2023–2029, 2012.
- [7] A. Sadeghian, Z. Zhongming Ye, and B. Bin Wu, "Online detection of broken rotor bars in induction motors by wavelet packet decomposition and artificial neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 7, pp. 2253–2263, 2009.
- [8] J. Zheng, Z. Dong, H. Pan, Q. Ni, T. Liu, and J. Zhang, "Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing," *Measurement*, vol. 143, pp. 69–80, 2019.
- [9] K. Shao, W. Fu, J. Tan, and K. Wang, "Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris Hawks optimization-based SVM for fault diagnosis of rolling bearing," *Measurement*, Article ID 108580, 2020.
- [10] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [11] D.-T. Hoang and H.-J. Kang, "A survey on deep learning bearing fault diagnosis," *Neurocomputing*, vol. 335, pp. 327– 335, 2019.
- [12] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering & System Safety*, vol. 115, pp. 124–135, 2013.
- [13] Y.-H. Li, W.-P. Shan, and X.-Q. Zeng, "Classification and identification of bearing faults based on deep belief network," *Journal of Vibration Engineering*, vol. 29, no. 2, pp. 340–347, 2016.
- [14] J. O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, no. 337, pp. 331–345, 2016.
- [15] H. Shi, Y. Yang, X. Zhu et al., "Embedding deep metric for person re-identification: a study against large variations," in *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, October 2016.
- [16] H.-X. Zhi, H.-T. Yu, and S.-M. Li, "A video classification method based on depth metric learning," *Electronics & Information Technology Journal*, vol. 40, no. 11, pp. 2562–2569, 2018.
- [17] B. Wu, Z. Chen, J. Wang, and H. Wu, "Exponential discriminative metric embedding in deep learning," *Neurocomputing*, vol. 290, pp. 108–120, 2018.
- [18] W. Yang, Y. Yan, and S. Chen, "Adaptive deep metric embeddings for person re-identification under occlusions," *Neurocomputing*, vol. 340, pp. 125–132, 2019.
- [19] J. Hu, J. Lu, and Y. P. Tan, "Discriminative deep metric learning for face verification in the wild," in *Proceedings of the*

2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, June 2014.

- [20] J. Hu, J. Lu, and Y. P. Tan, "Deep transfer metric learning," in Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Boston, MA, USA, June 2015.
- [21] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep metric learning via lifted structured feature embedding," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [22] X. Guo and X. Ma, "Fault diagnosis feature subset selection using rough set," *Computer Engineering and Application*, vol. 43, no. 1, pp. 221–224, 2007.
- [23] Z.-B. Xu, Y.-Y. Li, Z.-G. Wang, and J. Xuan, "A novel clustering method combining ART with Yu's norm for fault diagnosis of bearings," *Journal of Vibration and Shock*, vol. 2016, Article ID 5468716, 2016.
- [24] J. Lu, J. Hu, and J. Zhou, "Deep metric learning for visual understanding: an overview of recent advances," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 76–84, 2017.
- [25] X. Cao, Y. Ge, R. Li, J. Zhao, and L. Jiao, "Hyperspectral imagery classification with deep metric learning," *Neurocomputing*, vol. 356, pp. 217–227, 2019.
- [26] I. W. Tsang, J. T. Kwok, C. Bay, and H. Kong, "Distance metric learning with kernels," in *Proceedings of the ICANN (International Conference on Artificial Neural Networks)*, pp. 126– 129, Istanbul, Turkey, June 2003.
- [27] D.-Y. Yeung and H. Chang, "A kernel approach for semisupervised metric learning," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 141–149, 2007.
- [28] Y. D. Yu, "Triangular norms and TNF-sigma-algebras," Fuzzy Sets and Systems, vol. 2, no. 2, pp. 251–264, 1985.
- [29] P. Luukka, "Similarity classifier using similarity measure derived from Yu's norms in classification of medical data sets," *Computers in Biology and Medicine*, vol. 37, no. 8, pp. 1133–1140, 2007.
- [30] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, p. 40, 2007.
- [31] H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data-with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.
- [32] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [33] M.-R. Chen, B.-P. Chen, G. Q. Zeng, K. D. Lu, and P. Chu, "An adaptive fractional-order BP neural network based on extremal optimization for handwritten digit recognition," *Neurocomputing*, vol. 391, pp. 260–272, 2019.
- [34] The Case Western Reserve University Bearing Data Center, Bearing Data Center Fault Test Data, Case Western Reserve University, Cleveland, OH, USA, 1998, http://csegroups.case. edu/bearingdatacenter/pages/download-data-file.
- [35] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation," *Advances in Artificial Intelli*gence, vol. 4304, pp. 1015–1021, 2006.
- [36] L. Van Der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2580-2605, 2008.