Smart Structures and Systems, *Vol. 33, No. 4 (2024) 253-262*
https://doi.org/10.12989/sss.2024.33.4.253

Dual-loss CNN: A separability-enhanced network for current-based fault diagnosis of rolling bearings

Lingli Cui [1], Gang Wang [1], Dongdong Liu*[1], Jiawei Xiang [2] and Huaqing Wang [3]

[1] *Key Laboratory of Advanced Manufacturing Technology, Beijing University of Technology, Beijing 100124, China*
[2] *College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China*
[3] *College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China*

(*Received April 3, 2023, Revised March 11, 2024, Accepted March 18, 2024*)

Abstract. Current-based mechanical fault diagnosis is more convenient and low cost since additional sensors are not required. However, it is still challenging to achieve this goal due to the weak fault information in current signals. In this paper, a dual-loss convolutional neural network (DLCNN) is proposed to implement the intelligent bearing fault diagnosis via current signals. First, a novel similarity loss (SimL) function is developed, which is expected to maximize the intra-class similarity and minimize the inter-class similarity in the model optimization operation. In the loss function, a weight parameter is further introduced to achieve a balance and leverage the performance of SimL function. Second, the DLCNN model is constructed using the presented SimL and the cross-entropy loss. Finally, the two-phase current signals are fused and then fed into the DLCNN to provide more fault information. The proposed DLCNN is tested by experiment data, and the results confirm that the DLCNN achieves higher accuracy compared to the conventional CNN. Meanwhile, the feature visualization presents that the samples of different classes are separated well.

Keywords: convolutional neural network; fault diagnosis; feature representation; information fusion; rolling bearing

1. Introduction

Rolling bearings are key mechanical components and are widely used in a variety of industrial applications, including automobiles, airplanes, and wind turbines (Bae *et al*. 2022). However, they are prone to faults due to harsh operating conditions, such as severe impacts, high rotational speeds, and heavy loads. Therefore, automatic and accurate fault diagnosis of rolling bearings is vital to ensure the safety of rotating machinery which can prevent accidents and reduce the maintenance cost of the mechanical system (Liu *et al*. 2023).

Intelligent diagnosis can learn the fault information from collected data and produce accurate results. Usually, they include feature extraction and feature classification (Li *et al*. 2020). Shen *et al*. (2014) proposed a two-stage method that first extracts statistical features, and then uses those features to recognize bearing faults. Lessmeier *et al*. (2016) utilized wavelet packet decomposition (WPD) to extract features and then developed an ensemble learning classifier to recognize bearing faults. Toma *et al*. (2020) calculated the time-frequency domain features and then the k-nearest neighbor algorithm is used to identify bearing faults. Although those shallow learning methods have been studied extensively, the traditional methods cannot provide a satisfactory fault diagnosis and heavily rely on expert experience and domain knowledge (Shao *et al*. 2022).

Deep learning methods have emerged as a compelling alternative to shallow learning methods, as they have the ability to autonomously learn health information without relying on prior empirical knowledge. Convolutional neural networks (CNN) have gained popularity in the field of fault diagnosis (Wang *et al*. 2020), and are one of the dominant models for extracting diagnostic knowledge from machinery condition monitoring signals (Wang and Xiang 2021). A multi-head attention mechanism-based CNN model was developed to recognize various health states using vibration signals (Wang *et al*. 2019). Kumar *et al*. (2021) developed a new trigonometric cross-entropy function and introduced it into the CNN model to identify various bearing health conditions. In addition, Shao *et al*. (2018) proposed a locomotive bearings diagnosis method that uses CNN and a deep belief network to identify various faults based on raw vibration signals. Liu *et al*. (2022a) developed a rolling bearing fault recognition method that combines CNN with a matrix profile based impulse extraction strategy. Those CNN-based diagnosis methods have shown promising results in bearing fault diagnosis, since they depend on vibration signals that contain rich health information (Randall and Antoni 2011).

However, additional acceleration sensors must be installed near the monitored bearings to measure the vibration signals (Cui *et al*. 2022, Guan and Feng 2021). The installation of external measuring equipment is not available in some specific working environments, such as centrifugal pumps and cryogenic pumps (Singh *et al*. 2014). Different from the vibration signals, current signals can be

*Corresponding author, Ph.D.,
E-mail: liudd@bjut.edu.cn

ISSN: 1738-1584 (Print), 1738-1991 (Online)

measured using the existing frequency inverter in the motor (Leite *et al*. 2017). Alternatively, the external current sensor can be attached non-intrusively to the motor (Elbouchikhi *et al*. 2017). Hence, the measurement of current signals is convenient and cost-effective (Hsueh *et al*. 2019). Considering the success of CNN models in vibration-based diagnosis, they have the potential to provide promising diagnostic performance in current-based diagnosis.

Although CNN models exhibit powerful feature learning capabilities, it is challenging to extract a separable representation from current signals in bearing fault diagnosis. This is because the fault information is weak in the current data, as the bearing fault information is indirectly transmitted to the drive motor (Herold *et al*. 2013, Kim *et al*. 2022). The empirical wavelet transform (EWT) method was used to process the raw current signals and extract distinguishing features, which were then used to predict fault labels using a CNN model (Hsueh *et al*. 2019). To minimize the noise components in the input signals, Sadoughi *et al*. (2018) proposed a pre-processing method based on fast kurtogram analysis. Although signal preprocessing can remove noise to a certain extent and improve fault recognition, the selection of appropriate preprocessing methods requires domain knowledge. Hoang and Kang (2020) employed a CNN model to predict fault labels based on two-phase current signals and combined the predictions using a support vector machine (SVM). However, as the health information of each phase current signal was analyzed independently, the correlations between phases were neglected. In addition, common CNN models focus on minimizing prediction errors on individual samples, without considering the similarity between samples. This results in an inseparable representation, particularly in current-based diagnosis where the fault information is weak.

To address these challenges, the paper proposes a dual-loss convolutional neural network (DLCNN) that performs end-to-end current-based rolling bearing fault diagnosis. In DLCNN, a similarity loss (SimL) function is developed and introduced into a basic CNN model to maximize the intra-class similarity and minimize the inter-class similarity, which allows the DLCNN to learn a more separable and accurate representation of current signals. To leverage the performance of SimL, a weight parameter is used to balance the effect of the sample similarity on SimL. Furthermore, a synchronous two-phase current signal fusion (STPCSF) method is presented to fuse two-phase current signals measured synchronously at the data end, enabling models to explore the relationship between different current phases. The experimental evaluation validated the DLCNN method can achieve higher accuracy than the comparison methods. Moreover, due to the separable representation, our DLCNN method demonstrated a better performance in scenarios with limited training sample sizes compared to basic CNN.

The paper is structured as follows: Section 2 provides an overview of the basic principle of CNN. Section 3 analyzes the representation separability of various feature distributions and the effect of separability on the diagnostic results, and then the proposed DLCNN method and a fusion technique of two-phase current signals are provided. The experiments are given in Section 4, and Section 5 concludes this paper.

2. Principle of CNN

A typical CNN model consists of two parts: feature extractor g(·) and classifier c(·). Feature extractor is composed of convolutional layers, pooling layers, and nonlinear activation layers (Lecun *et al*. 1998). The classifier is composed of fully connected layers and nonlinear activation layers.

Convolutional layer: The task of feature learning is mainly realized by the convolutional layer. The feature extractor contains multiple convolutional layers. The convolutional layer can be represented as

$$Y_i^k = \varphi\left(\sum_{m=0}^{M-1} W_m^k * X_{i+m} + b^k\right) \tag{1}$$

where Y_i^k represents the i-th element of the k-th channel in the output, X_i denotes the i-th element of all channels present in the input, W_m^k is the m-th element of the k-th channel in the convolution kernel whose size is M, b^k is bias of k-th output channel, and $\varphi(\cdot)$ stands for the activation function.

Activation function: In this paper, the rectified linear unit (ReLU) is adopted as the activation function and its mathematical equation is

$$y = \begin{cases} x, & if\ x > 0, \\ 0, & else. \end{cases} \tag{2}$$

Fully connected layer: Fully connected layer (FCL) is responsible for mapping the features learned by feature extractor to a vector that represents the predicted fault labels. It can be mathematically expressed as follows

$$x^l = \varphi(w^T x^{l-1} + b^l) \tag{3}$$

where x^{l-1} is the inputs of the lth layer, w^T is the weight matrix used to connect the inputs and outputs, b^l represents the biases, $\varphi(\cdot)$ represent the activation function, and x^l stand for the lth layer outputs.

Softmax: The softmax function converts the outputs of the classifier into a vector that represents the probability of each health state. The softmax layer can be described as

$$y_i = \frac{exp(x_i)}{\sum_{j=1}^{C} exp(x_j)} \tag{4}$$

where x_i stand for the ith value of the inputs, C is the number of the health conditions, and y_i represents the ith outputs.

Gradient descent is the basic idea of neural network optimization. Before optimizing, a loss function must be specified to describe the optimization task. In the rolling bearing fault classification task, the cross-entropy loss is widely utilized to describe the gap between the predicted

labels and the real labels. The cross-entropy loss function can be expressed as

$$CEL = -\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{C} y_{k,i}^{T}\, log(y_{k,i}^{P}) \tag{5}$$

where $y_{k,i}^{T}$ and $y_{k,i}^{P}$ represent the true and predicted labels, respectively, N represents the sample number, C is the number of health states, and CEL means the cross-entropy loss.

3. The proposed DLCNN

In this paper, a dual-loss CNN method is presented to learn a separable representation of motor current signals in fault pattern recognition of rolling bearings. Additionally, the STPCSF method is developed for fusing two-phase current signals at the data end.

3.1 Separability analysis of representations

Three different representations are shown in Fig. 1 to analyze the effect of representation separability on diagnostic results. The feature distributions in the training set are represented by the solid coils and the possible feature distributions in the unseen testing set are represented by the dashed coils. The linear classifier $c(\cdot)$ is adopted to identify the health conditions according to the representations.

For a given class, the feature distribution of testing samples is expected to be consistent with the corresponding class in the training samples. In fact, there is a distributional shift between the training set and testing set, resulting in a wider possible feature distribution area for the testing samples than for the training samples. In Fig. 1(a), intra-class features are dispersed, and inter-class features are overlapped. Such distribution brings difficulty to feature classification, and the linear classifier cannot differentiate varieties of health conditions. In Fig. 1(b), intra-class features are clustered, but the cluster centers are close to each other. The distribution of inter-class features in the testing set is still overlapped, making it challenging to accurately classify the testing samples. In Fig. 1(c), the intra-class features are clustered, and the cluster centers of different classes are distant from each other. Furthermore, the feature distributions of different classes in the testing set do not overlap. This separable representation ensures safe inter-class distances and clear classification boundaries. Hence, the classifier trained on the training set can accurately classify unseen testing samples.

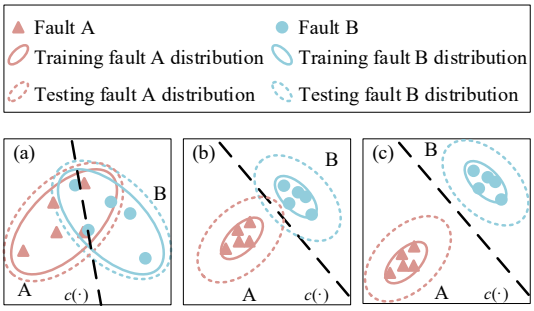


Fig. 1 Three different fault feature distribution

Therefore, to accurately identify the health conditions of testing samples, the representation of signals must be separable (the intra-class features are clustered and the cluster centers of different classes are distant from each other) to address the gap between the feature distributions of the training and testing sets.

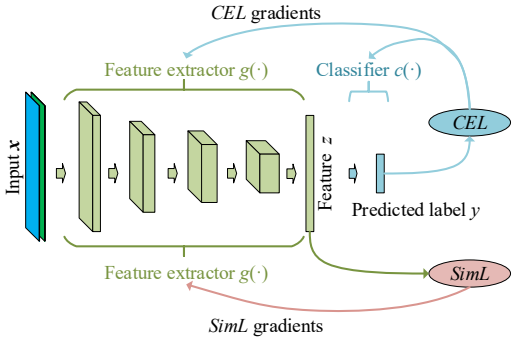


Fig. 2 The architecture of the proposed DLCNN

Algorithm 1: Learning algorithm of DLCNN

Input: steps S, batch size N, constant τ, $g(\cdot)$, $c(\cdot)$
for *step* $\in$ {1, 2, ..., S} **do**
  sample minibatch: $\{x_k, y_k\}_{k=1}^{N}$
  for $k \in$ {1, 2, ..., N} **do**
    $z_k = f(x_k)$, $y_k^P = f(f_k)$
  end for
  # the cross-entropy loss
  $CEL = -\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{C} y_{k,i}^{T}\, log(y_{k,i}^{P})$
  Z={}, Y={}, Z' ={}, Y' ={}
  for $k \in$ {2, 3, ..., N} **do**
    append $\{z_k\}_{k=1}^{N}$ to Z
    append $\{y_k\}_{k=1}^{N}$ to Y
    append $\{z_k, z_{k+1}, \ldots, z_N, \ldots, z_{k-1}\}$ to Z'
    append $\{y_k, y_{k+1}, \ldots, y_N, \ldots, y_{k-1}\}$ to Y'
  end for
  for $k \in$ {1, 2, ..., N(N-1)} **do**
    $s_k = \frac{z_k \cdot z_k'}{\|f_k\| \|f_k'\|}$
    $l_k = -log\frac{exp(s_k/\tau)}{\sum_{i=1}^{N(N-1)} exp(s_i/\tau)}$
  end for
  # the similarity loss
  $SimL = \frac{\sum_{k=1}^{N} \lambda_{y_k=y_k'}\, l_k}{\sum_{k=1}^{N} \lambda_{y_k=y_k'}}$
  $L = CEL + SimL$
  update networks $g(\cdot)$ and $c(\cdot)$ to minimize L
end for
return feature extractor $g(\cdot)$, classifier $c(\cdot)$

3.2 Design of the dual-loss CNN

The CNN is optimized by iteratively reducing prediction errors. Traditional CNN models struggle to learn a separable representation for current data, as the fault information in current signals is weak. To overcome this inherent limitation, a novel SimL function is designed to improve the representation separability. Based on the SimL and CEL, the dual-loss CNN is further developed and is shown in Fig. 2. Moreover, the complete pseudo-code of the DLCNN is provided in algorithm 1.

In Fig. 2, x represents the input data of the DLCNN, which is the fusion data of two-phase current signals. $g(\cdot)$ stands for the feature extractor, and z means the features extracted from x, $c(\cdot)$ represents the classifier and y is the predicted label. In addition, *CEL* provides an optimization object by evaluating the label prediction error. The more accurate the prediction, the smaller the value of *CEL*. *SimL* is expected to describe the separability of features. As the improvement of representation separability, the *SimL* decreases.

N sample and label pairs $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ are randomly selected as a minibatch, where x_i and y_i represent a bearing fault sample and its label. Then, N features $\{z_1, z_2, z_3, \ldots, z_N\}$ can be obtained by the feature extractor from the samples. And, the classifier predicts the health conditions based on the extracted features. Finally, the prediction error *CEL* of the minibatch samples can be obtained by Eq. (5).

To calculate the *SimL*, there are two feature lists need to be generated, as shown in Fig. 3. On the one hand, the original features are repeated (N-1) times to get the list Z including $N(N\text{-}1)$ features. On the other hand, (N-1) new arrangements of features are obtained by rearranging the original features. Finally, the list Z' including $N(N\text{-}1)$ features is generated by concatenating (N-1) new arrangements. Therefore, there are total $N(N\text{-}1)$ feature pairs according to the feature lists Z and Z'. Hereby, the two features of each feature pair come from different samples.

The cosine similarity of the two features at the corresponding position of F and F' are calculated in order. The cosine similarity of the two vectors is calculated as follows

$$sim(u, v) = \frac{u \cdot v}{\| u \| \| v \|} \tag{6}$$

In Eq. (6), u and v represent the feature vectors at the corresponding position of Z and Z' which are extracted from two different samples, $u \cdot v$ represents the dot production of u and$v$, $\|\cdot\|$ means l2-normoperation, and $sim(u, v)$ represents the cosine of u and v. For convenience, the cosine similarity of two features at the corresponding positions of F and F' are represented as

$$s_k = sim(z_k, z'_k), \qquad k = 0,1,2,\ldots,N(N-1). \tag{7}$$

where z_k and z'_k stand for the two feature vectors for different samples.

Let l_k

$$l_k = -log \frac{exp(s_k/\tau)}{\sum_{i=1}^{N(N-1)} exp(s_i/\tau)} \tag{8}$$

where τ is a constant, and it is used to weight the cosine value of z_k and z'_k. Then the similarity loss function is

$$SimL = \frac{\sum_{k=1}^{N} \lambda_{y_k=y'_k} l_k}{\sum_{k=1}^{N} \lambda_{y_k=y'_k}} \tag{9}$$

In which *SimL* represents similarity loss, and $\lambda_{y_k=y'_k}$ is shown as following equation

$$\lambda_{y_k=y'_k} = \begin{cases} 1, & y_k = y'_k \\ 0, & y_k \neq y'_k \end{cases} \tag{10}$$

The smaller the value of *SimL* in Eq. (9), the more separable the representation. Specifically, when the *SimL* is small enough, it indicates that features of the same class are clustered in the feature space, while features of different classes are separated.

Finally, both CEL and SimL are adopted to optimize the trainable parameters in the DLCNN

$$L = CEL + SimL \tag{11}$$

3.3 Information fusion of two-phase current signals

The synchronous measurement of two-phase current signals provides two distinct perspectives on the bearing health condition. To extract correlations between the current phases and further improve the accuracy of fault diagnosis, the STPCSF method is proposed and shown in Fig. 4. The main steps of the STPCSF method are listed below:

1) Two sub-signals are obtained from the two-phase current signals, and each sub-signal contains L samples.
2) Each sub-signal is transformed into a matrix with dimensions (M, M), where L = M × M.
3) The two matrices are concatenated to obtain fused data with dimensions (M, M, 2).
4) Repeat steps 1 to 3 until the remaining signal length is less than L.

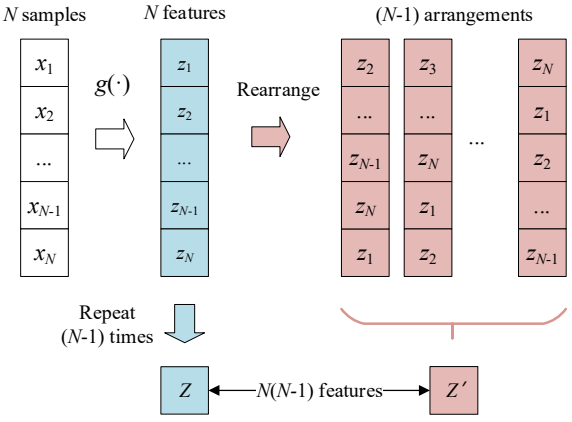


Fig. 3 Generation of two feature lists Z and Z'

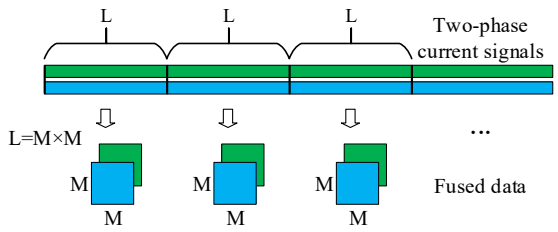


Fig. 4 The proposed synchronous two-phase current signal fusion method

Within fused data, two feature maps represent synchronous measurements of two-phase currents. CNN models can capture the correlation between current phases via weighted summation of all channels at a local area on the feature maps. Therefore, bearing health information from two-phase current signals can be leveraged to accurately recognize health conditions.

4. Experimental validation

4.1 Experimental setup and dataset description

The experimental validation is performed on a public dataset of the rolling bearing provided by the Kat-Data Center of the Chair of Design and Drive Technology, Paderborn University, Germany (Lessmeier *et al*. 2016). As shown in Fig. 5, the test rig mainly consists of five parts: an electric motor (permanent magnet synchronous motor), a torque-measurement shaft, a rolling bearing test module, a flywheel, and a load motor. The test rig is driven by a permanent magnet synchronous motor with a nominal speed of 3000 rpm. To generate the inertia of the rotating machine, a flywheel is introduced into the system. Additionally, a load motor is utilized to emulate the load imposed by the driven equipment.

In the experiments, two-phase current signals of the motor are measured by the LEM CKSR 15-NP current sensor, and vibration signals of the bearing test module are recorded using a piezoelectric accelerometer (type: 336C04) directly above the housing. The two-phase current signals and the vibration data are measured synchronously at a sampling frequency of 64 kHz. A total of 15 bearings (type: 6203) are tested in the experiment. The details of the health condition and damage of each bearing are listed in Table 1. To ensure the diversity of collected signals, each bearing is tested 20 times under each condition. Table 2 displays the bearings tested under four working conditions, with variations in motor rotational speed, load torque, and radial force. In addition, the health conditions of the tested bearing include normal condition (NC), outer ring fault (OF), and inner ring fault (IF). For each health condition, the collected two-phase current signals are converted into 1000 samples through the presented STPCSF method, and each sample containing 2×6400 data points is a fused data of two sub-signals.

To evaluate the effectiveness of DLCNN, the samples are divided into training and testing samples. First, the bearings with health conditions NC, OF and IF are labeled as 0, 1 and 2, respectively. Second, the labeled bearing samples are randomly shuffled. Finally, a given training sample proportion (TSP) is applied to select some samples to optimize the network, while the rest are utilized for testing.

The DLCNN is developed using Python 3.9, and a set of performance diagnostic tests are conducted to demonstrate its capabilities. The neural networks involved in the experiments have been constructed and optimized using Tensorflow version 2.6. Moreover, the inference of the neural networks is accelerated using an Nvidia Geforce 2060 graphics card. Furthermore, Table 3 provides comprehensive information on the detailed configurations of the proposed DLCNN, which consists of two sub-models: a convolution-based feature extractor $g(\cdot)$ and a classifier $c(\cdot)$.

As shown in Fig. 6(a), a normal sample (x_{NC}), an outer ring fault sample (x_{OF}), and an inner ring fault sample (x_{IF}) are displayed, respectively. The untrained CNN-based feature extractor is utilized to represent these samples as f_{NC}, f_{OF}, and f_{IF}, as shown in Fig. 6(b). According to row 9 of Table 3, the feature extractor output is a 750-element vector. To facilitate visualization, we only display the first 100 elements. It can be found that the representations obtained by the untrained CNN are almost indistinguishable, regardless of the health status of the input signal.

To quantitatively evaluate the separability, two evaluation metrics are introduced. The features similarity of the same class (SS) can be described as

$$SS = \frac{\sum_{k=1}^{N} \lambda_{y_k=y'_k} s_k}{\sum_{k=1}^{N} \lambda_{y_k=y'_k}} \tag{12}$$

The features similarity of different classes (SD) is expressed as

Table 1 The health state and damage extent of each tested bearing

Bearing	Health state	Damage
K001	NC	None
K002	NC	None
K003	NC	None
K004	NC	None
K005	NC	None
KA04	OF	0-2 mm
KA15	OF	0-2 mm
KA16	OF	2-4.5 mm
KA22	OF	0-2 mm
KA30	OF	0-2 mm
KI04	IF	0-2 mm
KI14	IF	0-2 mm
KI16	IF	4.5-13.5 mm
KI18	IF	2-4.5 mm
KI21	IF	0-2 mm

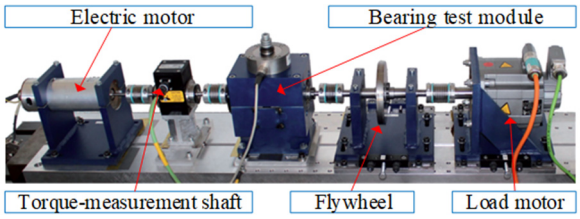


Fig. 5 Overview of the test rig

Table 2 Four working conditions

Working condition	Rotational speed [rpm]	Load torque [Nm]	Radial force [N]	Number of samples
A	1500	0.7	1000	3×1000
B	900	0.7	1000	3×1000
C	1500	0.1	1000	3×1000
D	1500	0.7	400	3×1000

$$SD = \frac{\sum_{k=1}^{N} \lambda_{y_k \neq y_k'} s_k}{\sum_{k=1}^{N} \lambda_{y_k \neq y_k'}} \tag{13}$$

The SS and SD metrics are dimensionless quantities with a range of 0 to 1. SS metric represents the aggregation degree of the same class features, and SD metric represents the separability of different class features. Therefore, the larger the SS, the smaller the SD, which means that the representation is more separable. The SS and SD metrics for untrained feature extractors are approximately 0.95, which demonstrates that untrained feature extractors cannot separately represent the current signal.

4.2 Experimental results and comparisons

As shown in Eq. (8), the parameter τ is used to assign weight to the cosine similarity of two feature vectors, and therefore it directly affects the results of the SimL. Several tests are conducted on the DLCNN using different values of τ to determine the appropriate parameter value. The experiments are conducted on the current signals collected under working condition B. In the tests, the parameter τ is set to 0.05, 0.5, and 5.0, respectively. In addition, 80% of the samples are randomly selected to optimize the network, and the remaining 20% are used for testing.

In the experiments, a basic CNN model and the DLCNN model (with varying parameter τ) are trained using the Adam optimizer (Kingma and Ba 2014) with a learning rate of 0.0001 and a batch size of 32. To obtain the optimal model, each model is iteratively optimized 550 times on training samples. Finally, the networks are evaluated using metrics of Accuracy, SS, and SD. To avoid contingency, each experiment is performed five times, and the average and standard deviation of the metrics are shown in Table 4 and Fig. 7.

The first row in Table 4 is the untrained CNN model, and the second row is a conventional CNN model trained solely using CEL. The remaining rows show the test results of DLCNN under varying parameter τ. In addition, the variation curves of prediction error (i.e., CEL), classification accuracy, SS, and SD on the testing samples during training are given in Fig. 8, in which "None" means

Table 3 Configurations of the proposed DLCNN

Model	Layer	Kernel size	Kernel number	Output size	Activation
$g(\cdot)$	C1	(9, 9)	5	(80, 80, 5)	ReLU
$g(\cdot)$	P1	(2, 2)	5	(40, 40, 5)	
$g(\cdot)$	C2	(7, 7)	10	(40, 40, 10)	ReLU
$g(\cdot)$	P2	(2, 2)	10	(20, 20, 10)	
$g(\cdot)$	C3	(5, 5)	15	(20, 20, 15)	ReLU
$g(\cdot)$	P3	(2, 2)	15	(10, 10, 15)	
$g(\cdot)$	C4	(3, 3)	30	(10, 10, 30)	ReLU
$g(\cdot)$	P4	(2, 2)	30	(5, 5, 30)	
$g(\cdot)$	Flatten			750	
$c(\cdot)$	FCL			3	Softmax

(a) Phase 1 of two current signals (b) Features represented by untrained g(·)

Fig. 6 Visualization of three current signal samples

Table 4 Diagnostic performance of DLCNN with different parameter τ

Loss	τ	Accuracy (%)	SS	SD
-	-	-	0.96 ± 0.012	0.95 ± 0.007
CEL	-	93.13 ± 2.82	0.63 ± 0.038	0.55 ± 0.014
DL	0.05	97.02 ± 1.14	0.93 ± 0.006	0.76 ± 0.010
DL	0. 5	98.39 ± 0.35	0.69 ± 0.055	0.19 ± 0.134
DL	5.0	96.29 ± 2.87	0.63 ± 0.122	0.52 ± 0.202

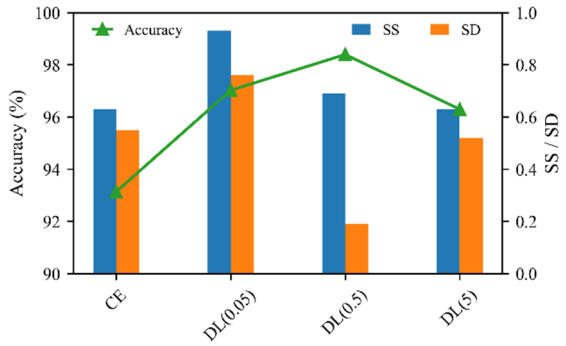


Fig. 7 Diagnostic performance of DLCNN with different parameter τ

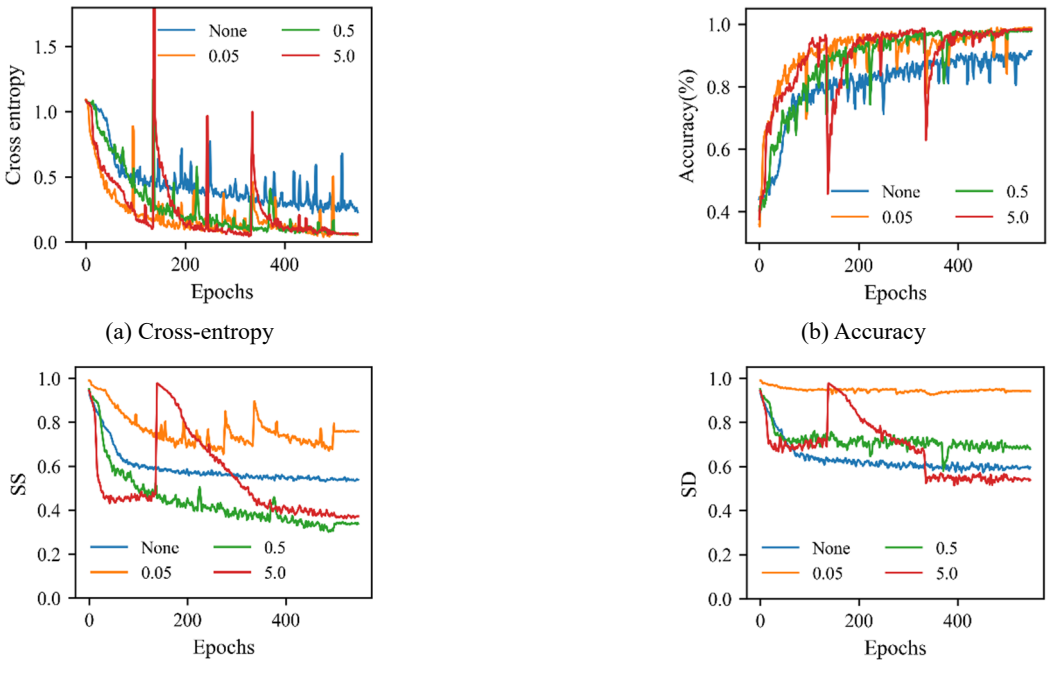
that the proposed SimL is not used during training. Furthermore, the t-distributed Stochastic Neighbor Embedding algorithm (Van and Hinton 2008) is introduced to transform the representations into a two-dimensional space, as presented in Fig. 9.

For the conventional CNN method, the accuracy, SS, and SD are about 93%, 0.63, and 0.55, respectively. As shown in Fig. 9(a), the representations learned by CNN suffer from serious inseparability due to the dispersed intra-class features and overlapped inter-class features. For the DLCNN model, the accuracy is higher, SS is larger, and SD is smaller. If τ is set to 0.5, the DLCNN model achieves the highest diagnostic accuracy of 98.39%, with the minimum SD metric of 0.19. This indicates that the feature similarity of the different classes has a greater effect on fault recognition than the feature similarity of the same class. Additionally, as shown in Fig. 9(c), the same class features are clustered, and the cluster centers are distant from each other. It can be found that the proposed SimL significantly improves the representation separability and diagnosis.

To test the effectiveness of the STPCSF approach, our DLCNN method and a basic CNN model are tested on phase 1 (C1), phase 2 (C2) of two-phase current signals, and the fused data of two phases (C1&2) collected from working conditions A, B, C, and D respectively. In addition, the basic architecture of CNN model is the same as DLCNN in the tests, as shown in Table 3. The parameter τ of SimL is 0.5, and the TSP is 0.8. Furthermore, the optimizer and other hyperparameters used to optimize the model are consistent with previous experiments. The detailed results are listed in Table 5.

Some conclusions can be drawn from the tests as follows: Firstly, DLCNN improves diagnostic accuracy based on the fused data of two-phase current signals acquired via the STPCSF technique. Second, DLCNN performs better than CNN under most working conditions due to the separable representation.

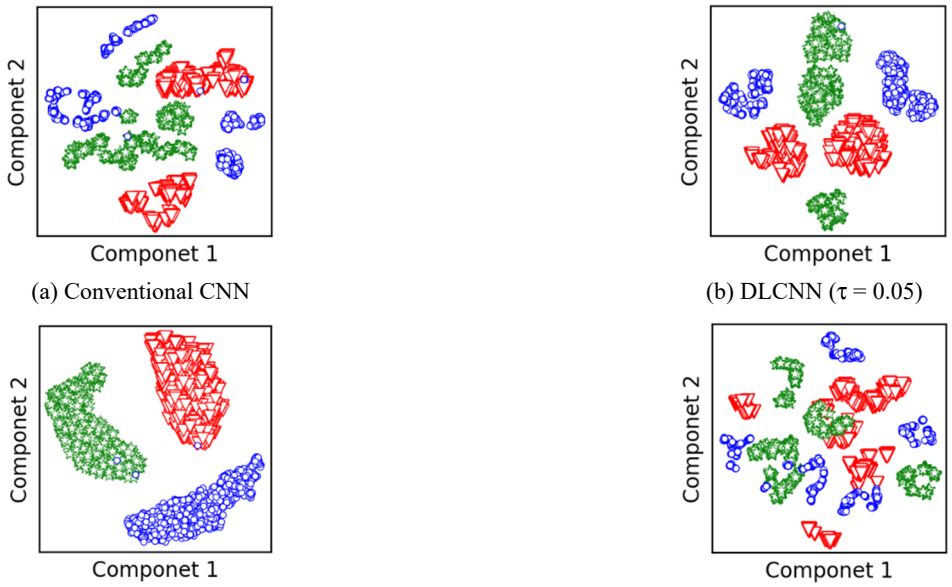
Additionally, the diagnostic performance of deep learning models relies on sufficient fault data (Gao *et al.* 2020). To evaluate the diagnostic performance of DLCNN with limited training samples, the DLCNN and CNN are



(a) Cross-entropy (b) Accuracy

(c) SS (d) SD

Fig. 8 The metric curves during training



(a) Conventional CNN (b) DLCNN ($\tau = 0.05$)

(c) DLCNN ($\tau = 0.5$) (d) DLCNN ($\tau = 5.0$)

Fig. 9 Representation visualization of current signals by CNN and DLCNN

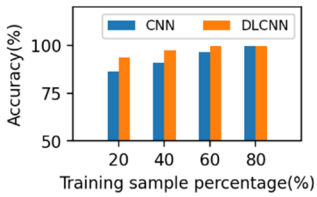
Table 5 Diagnostic results of CNN and DLCNN based on different signals

Method	Data	A	B	C	D
CNN	C1	90.26%	86.40%	85.32%	86.01%
CNN	C2	88.15%	85.30%	92.55%	92.13%
DLCNN	C1	94.41%	91.95%	89.31%	89.19%
DLCNN	C2	91.33%	87.76%	94.31%	95.61%
DLCNN+STPCSF	C1&2	99.52%	98.27%	98.73%	99.46%

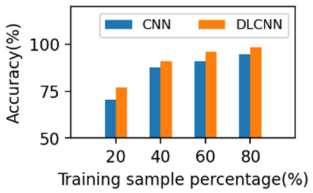
Table 6 Diagnostic results under different training sample percentages

Method	Condition	20%	40%	60%	80%
CNN	A	86.45%	90.84%	96.44%	99.70%
	B	70.57%	87.64%	90.85%	94.50%
	C	69.35%	80.81%	94.66%	96.70%
	D	79.85%	83.81%	94.47%	97.17%
	Mean	76.56%	85.78%	94.11%	97.02%
DLCNN	A	93.73%	97.42%	99.57%	99.52%
	B	77.03%	90.95%	95.73%	98.27%
	C	79.36%	93.24%	97.44%	98.73%
	D	86.17%	97.46%	95.39%	99.46%
	Mean	84.07%	94.77%	97.03%	99.00%
Improved accuracy		7.51%	8.99%	2.92%	1.98%

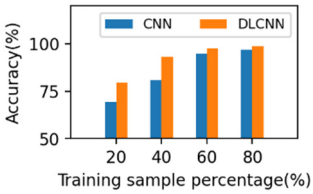
tested under four different TSPs (20%, 40%, 60%, and 80%). To reduce randomness, five tests are conducted for each TSP. The averaging testing accuracies are provided in Table 6 and Fig. 10. The diagnostic accuracy increases with the increase in training samples for both CNN and DLCNN. However, DLCNN outperforms conventional CNN in terms of fault classification accuracy in different TSP cases. The comparison results confirm that our DLCNN method outperforms the conventional CNN method in current-based fault diagnosis. Furthermore, as the proportion of training samples decreases from 80% to 20%, the superiority of DLCNN in accuracy over CNN increases from 2% to 7.5%. This suggests that the diagnostic performance of DLCNN is less sensitive to the training sample size compared to the basic CNN method, and DLCNN exhibits the potential to address the issue of insufficient training samples.



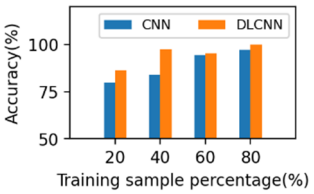
(a) Tested on working condition A



(b) Tested on working condition B



(c) Tested on working condition C



(d) Tested on working condition D

Fig. 10 Diagnosis accuracy under different training sample percentages

5. Conclusions

To boost the current-based diagnostic performance of rolling bearings, we presented the DLCNN method in which a novel SimL is constructed to improve the representation separability for current signals in rolling bearing fault diagnosis. The experiments confirm that the diagnostic accuracy of DLCNN is superior to the conventional CNN method. In particular, the DLCNN method achieved a 99% accuracy of fault recognition based on the fused data of two-phase current signals. Furthermore, the diagnostic accuracy of DLCNN in the scenarios with insufficient training samples is significantly higher than that of the basic CNN. Owing to the novel SimL, the DLCNN achieves more separable representation and accurate diagnostic results via raw current signals, which is of great significance when the vibration-based method is not feasible. In further work, we will explore the potential of the proposed SimL in cross-domain fault diagnosis tasks. The code of SimL and DLCNN is publicly available at: *https://github.com/OneBugMaker/DLCNN*.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 52075008 and Grant 52305086.

References

Bae, J., Jung, W. and Park, Y.H. (2022), "Normal data based rotating machine anomaly detection using CNN with self-labeling", *Smart Struct. Syst., Int. J.*, **29**(6), 757-766. https://doi.org/10.12989/sss.2022.29.6.757

Cui, L. Li, W., Wang, X., Zhao, D. and Wang, H. (2022), "Comprehensive remaining useful life prediction for rolling element bearings based on time-varying particle filtering", *IEEE Trans. Instrum. Meas.*, **71**, 1-10. https://doi.org/10.1109/TIM.2022.3163167

Elbouchikhi, E., Choqueuse, V., Auger, F. and Benbouzid, M.E.H. (2017), "Motor current signal analysis based on a matched subspace detector", *IEEE Trans. Instrum. Meas.*, **66**(12), 3260-3270. https://doi.org/10.1109/TIM.2017.2749858

Gao, Y., Liu, X. and Xiang, J. (2020), "FEM simulation-based generative adversarial networks to detect bearing faults", *IEEE Trans. Ind. Inform.*, **16**(7), 4961-4971. https://doi.org/10.1109/TII.2020.2968370

Guan, Y. and Feng, Z. (2021), "Adaptive linear chirplet transform for analyzing signals with crossing frequency trajectories", *IEEE Trans. Ind. Electron.*, **69**(8), 8396-8410. https://doi.org/10.1109/TIE.2021.3097605

Herold, T., Mboo, C.P. and Hameyer, K. (2013), "Evaluation of the use of an electrical drive as a sensor for the detection of bearing damage", *Proc. Conference on Acoustics, AIA-DAGA.*, pp. 1538-1540.

Hoang, D.T. and Kang, H.J. (2020), "A motor current signal-based bearing fault diagnosis using deep learning and information fusion", *IEEE Trans. Instrum. Meas.*, **69**(6), 3325-3333. https://doi.org/10.1109/TIM.2019.2933119

Hsueh, Y.M., Ittangihal, V.R., Wu, W.B., Chang, H.C. and Kuo, C.C. (2019), "Fault diagnosis system for induction motors by CNN using empirical wavelet transform", *Symmetry*, **11**(10), 1212. https://doi.org/10.3390/sym11101212 [EWT]

Kim, I., Lee, Y., Oh, J. and Kim, N. (2022), "Fault detection and classification of permanent magnet synchronous machine using signal injection", *Smart Struct. Syst., Int. J.*, **29**(6), 785-790. https://doi.org/10.12989/sss.2022.29.6.785

Kingma, D.P. and Ba, J. (2014), "Adam: A method for stochastic optimization", arXiv preprint. https://doi.org/10.48550/arXiv.1412.6980

Kumar, A., Vashishtha, G., Gandhi, C.P., Zhou, Y., Glowacz, A. and Xiang, J. (2021), "Novel convolutional neural network (NCNN) for the diagnosis of bearing defects in rotary machinery", IEEE Trans. Instrum. *Meas.*, **70**, 1-10. https://doi.org/10.1109/TIM.2021.3055802

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998), "Gradient-based learning applied to document recognition", *Proc. IEEE*, **86**(11), 2278-2324. https://doi.org/10.1109/5.726791

Leite, V.C.M.N., da Silva, J.G.B., Torres, G.L., Veloso, G.F.C., da Silva, L.E.B., Bonaldi, E.L. and de Oliveira, L.E.D.L. (2017), "Bearing fault detection in induction machine using squared envelope analysis of stator current", Bearing Technology; Rijeka, Croatia: InTech.

Lessmeier, C., Kimotho, J.K., Zimmer, D. and Sextro, W. (2016), "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification", *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, **3**(1), 5-8. https://doi.org/10.36001/phme.2016.v3i1.1577

Li, Y., Dai, W. and Zhang, W. (2020), "Bearing Fault Feature Selection Method Based on Weighted Multidimensional Feature Fusion", *IEEE Access*, **8**, 19008-19025. https://doi.org/10.1109/ACCESS.2020.2967537

Liu, D., Cui, L., Cheng, W., Zhao, D. and Wen, W. (2022a), "Rolling bearing fault severity recognition via data mining integrated with convolutional neural network", *IEEE Sens. J.*, **22**(6), 5768-5777. https://doi.org/10.1109/JSEN.2022.3146151

Liu, X., Liu, S., Xiang, J., Sun, R. and Wei, Y. (2022b), "An ensemble and shared selective adversarial network for partial domain fault diagnosis of machinery", *Eng. Appl. Artif. Intel.*, **113**, 104906. https://doi.org/10.1016/j.engappai.2022.104906

Liu, D., Cui, L. and Cheng, W. (2023), "Fault diagnosis of wind turbines under nonstationary conditions based on a novel tacholess generalized demodulation", *Renew. Energy*, **206**, 645-657. https://doi.org/10.1016/j.renene.2023.01.056

Randall, R.B. and Antoni, J. (2011), "Rolling element bearing diagnostics—A tutorial", *Mech. Syst. signal Process.*, **25**(2), 485-520. https://doi.org/10.1016/j.ymssp.2010.07.017

Sadoughi, M., Downey, A., Bunge, G., Ranawat, A., Hu, C. and Laflamme, S. (2018), "A deep learning-based approach for fault diagnosis of roller element bearings", In: *2018 Annual Conference of the Prognostics and Health Management Society*, **10**(1), 1-7.

Shao, H., Jiang, H., Zhang, H. and Liang, T. (2018), "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network", *IEEE Trans. Ind. Electron.*, **65**(3), 2727-2736. https://doi.org/10.1109/TIE.2017.2745473

Shao, H., Xia, M., Wan, J. and Silva, C.W. (2022), "Modified stacked autoencoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery", *IEEE-ASME Trans. Mech.*, **27**(1), 24-33. https://doi.org/10.1109/TMECH.2021.3058061

Shen, C., Wang, D., Liu, Y., Kong, F. and Tse, P.W. (2014), "Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines", *Smart Struct. Syst., Int. J.*, **13**(3), 453-471. https://doi.org/10.12989/sss.2014.13.3.453

Singh, S., Kumar, A. and Kumar, N. (2014), "Motor current signature analysis for bearing fault detection in mechanical systems", *Procedia Mater. Sci.*, **6**, 171-177. https://doi.org/10.1016/j.mspro.2014.07.021

Toma, R.N., Prosvirin, A.E. and Kim, J.M. (2020), "Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers", *Sensors*, **20**(7), 1884. https://doi.org/10.3390/s20071884

Van, V.M.L. and Hinton, G. (2008), "Visualizing data using t-SNE", *J. Mach. Learn. Res.*, **9**, 2579-2605.

Wang, G. and Xiang, J. (2021), "Railway sleeper crack recognition based on edge detection and CNN", *Smart Struct. Syst., Int. J.*, **28**(6), 779-789. https://doi.org/10.12989/sss.2021.28.6.779

Wang, H., Xu, J., Yan, R., Sun, C. and Chen, X. (2019), "Intelligent bearing fault diagnosis using multi-head attention-based CNN", *Procedia Manuf.*, **49**, 112-118. https://doi.org/10.1016/j.promfg.2020.07.005

Wang, H., Xu, J., Yan, R. and Gao, R.X. (2020), "A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN", *IEEE Trans. Instrum. Meas.*, **69**(5), 2377-2389. https://doi.org/10.1109/TIM.2019.2956332