Front. Mech. Eng. 2021, 16(2): 340–352
https://doi.org/10.1007/s11465-021-0629-3

RESEARCH ARTICLE

Xin ZHANG, Tao HUANG, Bo WU, Youmin HU, Shuai HUANG, Quan ZHOU, Xi ZHANG

Multi-model ensemble deep learning method for intelligent fault diagnosis with high-dimensional samples

Abstract Deep learning has achieved much success in mechanical intelligent fault diagnosis in recent years. However, many deep learning methods cannot fully extract fault information to recognize mechanical health states when processing high-dimensional samples. Therefore, a multi-model ensemble deep learning method based on deep convolutional neural network (DCNN) is proposed in this study to accomplish fault recognition of high-dimensional samples. First, several 1D DCNN models with different activation functions are trained through dimension reduction learning to obtain different fault features from high-dimensional samples. Second, the obtained features are constructed into 2D images with multiple channels through a conversion method. The integrated 2D feature images can effectively represent the fault characteristic contained in raw high-dimension vibration signals. Lastly, a 2D DCNN model with multi-layer convolution and pooling is used to automatically learn features from the 2D images and identify the fault mode of the mechanical equipment by adopting a softmax classifier. The proposed method, which is validated using the bearing public dataset of Case Western Reserve University, USA and a one-stage reduction gearbox dataset, has high recognition accuracy. Compared with other classical deep learning methods, the proposed fault diagnosis method has considerable improvements.

Keywords fault intelligent diagnosis, deep learning, deep convolutional neural network, high-dimensional samples

Received September 6, 2020; accepted January 8, 2021

Xin ZHANG, Tao HUANG, Bo WU, Youmin HU (✉), Shuai HUANG, Xi ZHANG
School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
E-mail: youmhwh@163.com

Quan ZHOU
School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China

1 Introduction

Intelligent fault diagnosis is a vital technology in today's manufacturing systems because effective detections can considerably improve the reliability of equipment [1–7]. Traditional intelligent fault diagnosis methods, such as support vector machine (SVM) [8], artificial neural network [9], and the Markov model [10], mostly complete the fitting of functions in the model structure of one or two layers, which belong to the algorithm structure of "shallow learning." However, shallow learning models cannot fully mine the fault characteristics contained in high-dimensional data, resulting in unreliable diagnosis results. As an alternative method, deep learning [11–15] is a promising strategy for fault diagnosis and has been extensively applied in fault diagnosis and other fields, such as image identification and speech recognition. In fault diagnosis, deep learning models are suitable for application to the identification of mechanical fault types; they can automatically learn abstract representation features from raw data and avoid obtaining abstract features by handcrafted features designed by engineers. Consequently, the limitation of expert knowledge can be avoided considerably by utilizing deep learning models to solve the problem of fault classification [16,17].

As one of the most effective deep learning methods, deep convolutional neural network (DCNN) can effectively extract abstract fault features through multi-layer nonlinear mapping [18–20]. The DCNN structure is composed of several convolution and pooling layers, which can extract key features in the process of mapping the raw input data layer by layer. The recognition results can be obtained by constructing and implementing a classifier, such as a softmax or SVM classifier, at the end of the DCNN model. Given that time-domain signals are the most common data type, one-dimensional (1D) DCNNs are usually applied in the real-time fault diagnosis of rotating machinery by inputting the 1D signals [21]. In comparison with two-dimensional (2D) DCNN, 1D DCNN extracts less effective information because the

original target of convolutional neural network (CNN) is used for image recognition. Therefore, 1D raw data were presented in a 2D image format in several studies [22,23], and fault classification was accomplished with image identification methods. However, 2D DCNN models need to be supported by experts' experience because they are not completely data-driven. To eliminate the reliance on experts' experience as much as possible, Ref. [24] introduced a data pre-processing method that can translate 1D raw signals into 2D gray images without handcrafted parameter settings. Data-driven DCNN models can intelligently mine the fault characteristics of high-dimensional complex samples and are suitable for real-time fault diagnosis [25]. Nevertheless, the applications of the DCNN method in fault diagnosis generally have three drawbacks. First, although the network structure of DCNN models can theoretically fit various complex nonlinear mappings, several obscure fault characteristics hidden in high-dimensional space are still easily ignored [26]. Second, due to the single model type and invariant network hyperparameter settings, the universalities of DCNN models need to be improved [27]. Third, the classification accuracies of DCNNs are usually influenced by the unbalanced number of different fault type samples [28].

This study focuses on solving such drawbacks, and a multi-model ensemble deep learning (MMEDL) approach based on DCNN is proposed to learn the hidden features from high-dimensional samples. The main contributions of this work are summarized as follows. First, the MMEDL model can directly utilize high-dimensional complex raw data as the model input without manual feature extraction. Second, the presented approach can extract abundant overall feature information by fusing multi-source features from different 1D DCNN models. Third, as a result, the presented approach can achieve excellent performance even with unbalanced datasets.

The remainder of this paper is organized as follows. Section 2 introduces common CNN contents and the process of the proposed approach. Section 3 demonstrates the effectiveness of the MMEDL model by using two different cases. Section 4 presents the conclusions.

2 Methodology

2.1 Convolutional neural networks

Generally, a DCNN model is constructed by a hierarchy of convolutional and pooling layers followed by one or more fully connected layers to extract features from input data; it is a typical feedforward network. A typical architecture of 2D DCNN is shown in Fig. 1. In this study, 1D and 2D DCNN models are used for feature extraction, and their principles are similar, as shown in Fig. 1. The convolutional layer employs a series of filters to extract the feature map and transfer the information layer by layer. Every filter is convolved with its input, and nonlinear mapping can be implemented by calculating with an activation function. In addition, one filter corresponds to one feature map, and using numerous filters can generate numerous feature maps in the next layer. In general, sufficient convolution operations can remarkably improve the fitting precision, but they also increase the calculation burden of the model. The process of the convolutional layer is described as follows:

$$(x * k)_i = \sigma\left(\sum_{j=0}^{d_k-1} k_j x_{i+j} + b\right), \tag{1}$$

where x is the input of the convolution layer, k denotes several trainable filters with length d_k, b expresses bias, and σ is a nonlinear activation function.

Nonlinear activation function σ is introduced to improve the expressive capability of the entire network with nonlinear mappings. In this study, three different activation functions shown in Table 1 that can express abundant mapping relationships are applied in the 1D DCNN models. The choice of the combination of activation functions can influence the model's recognition accuracy. However, the main objective of this study is to prove that deep neural network models integrating multi-type activation functions can better deal with high-dimensional samples than single models. Therefore, the selection of the best combination of activation functions needs to be investigated further in the future.

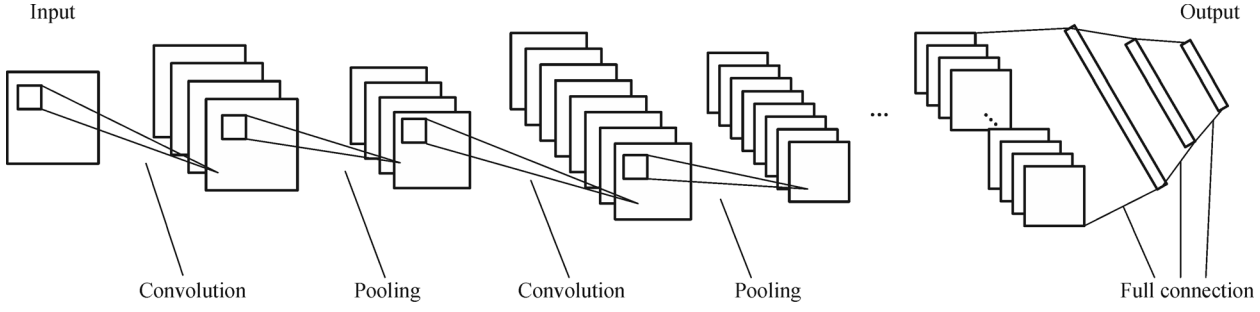

Fig. 1 Typical architecture of 2D deep convolutional neural network.

Table 1 Three different types of activation functions used in this study

Function	Expression
relu	$f(x)=\begin{cases} x, & x\geqslant 0 \\ 0, & x<0 \end{cases}$
elu	$f(\alpha, x)=\begin{cases} x, & x\geqslant 0 \\ \alpha(e^{x}-1), & x<0 \end{cases}$
tanh	$f(x)=2/(1+\exp(-2x))-1$

A convolution layer is generally followed by a pooling layer, which is used to decrease the size of the output feature maps from the previous convolution layer. Max-pooling and average pooling are typically employed to compress the feature space, extract pooling features, and improve the robustness of the deep neural network.

The output of the last pooling layer is flattened and connected to the fully connected layers. The classification module of 2D DCNN is composed of fully connected layers and a classifier. In the fully connected layers, each neuron is connected to every neuron in the next layer and achieves nonlinear mapping through an activation function. The softmax regression model is often placed at the end of the deep neural network structure as a classifier to classify the final output features.

In our study, a gradient descent-based backpropagation algorithm is applied to train the above-mentioned DCNN structure with cross-entropy as the loss function.

2.2 Flowchart of the proposed method

In this work, an MMEDL method is proposed for mechanical intelligent health identification with high-dimensional samples. In our method, high-dimensional raw data are preliminarily extracted using several 1D DCNN models with different activation functions to enrich feature diversity and achieve dimensionality reduction. The 1D features extracted by the 1D DCNN models are translated to feature images. The feature images from different sources are gathered and fused as the input of 2D DCNN, which can construct an authentic fault feature space to improve fault diagnosis accuracy.

The procedure of the proposed method is shown in Fig. 2. The procedure can be summarized in four main processes, namely, sample acquisition, feature extraction in 1D DCNNs, feature fusion, and fault recognition in 2D

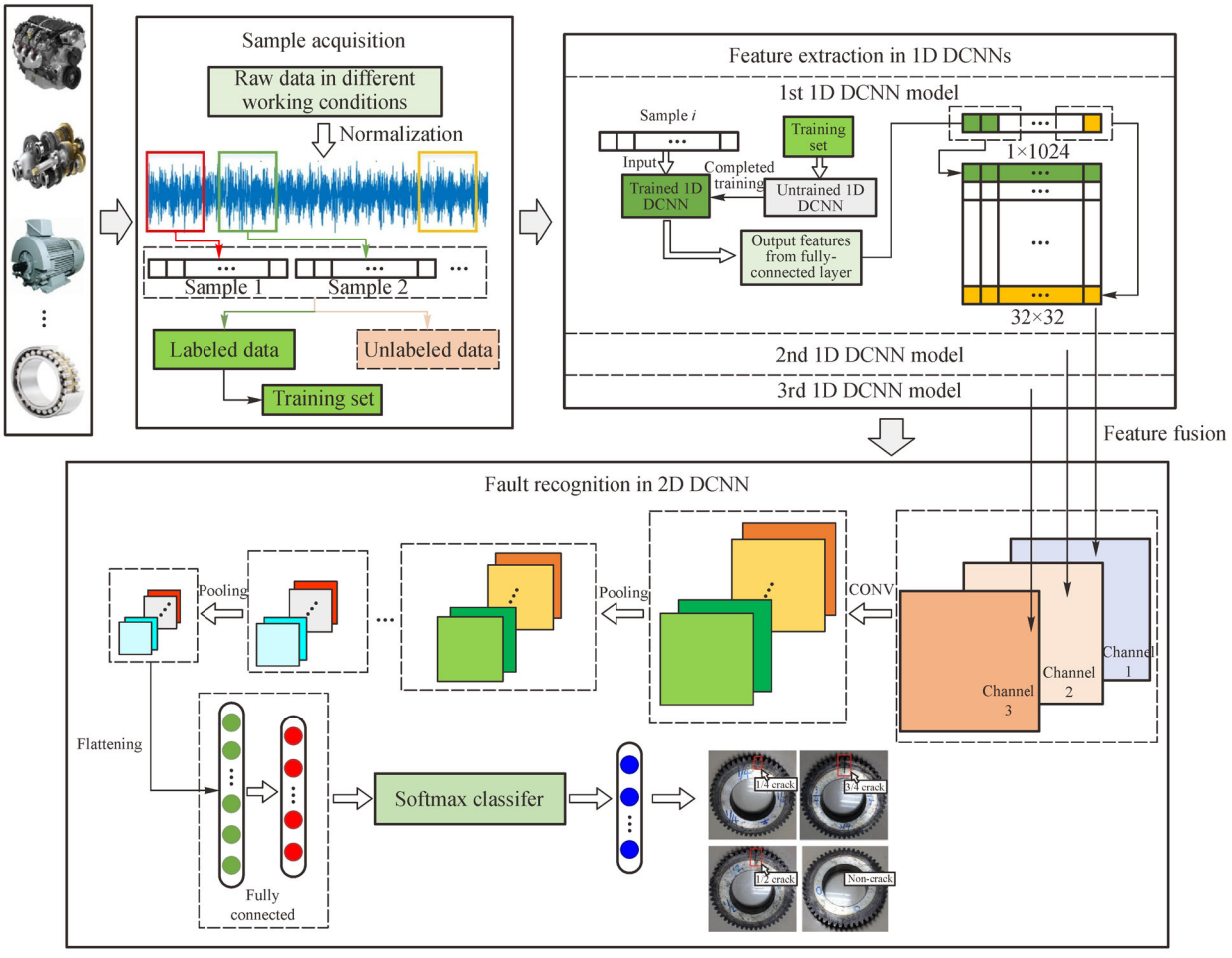

Fig. 2 Main procedure of the proposed multi-model ensemble deep learning method.

DCNN. First, the vibration signals in various working conditions are collected using accelerometers. The raw signals are sliced with interval sampling, where the labeled samples are gathered as the training set. A raw signal of length L can be divided into n samples according to

$$n = L/(N+I), \tag{2}$$

where each sample has N data points and I is the interval between two samplings.

Second, the training set is used to train several 1D DCNN models, and each of the models is constructed by convolution layers, pooling layers, a fully-connected layer, and a softmax classifier. After training, the outputs from the fully connected layers of these 1D DCNN models are collected separately as the result of feature extraction and dimensionality reduction. Then, the 1D features are translated into 2D images (i.e., 1×1024 features to 32×32 images in this study).

Third, the feature images are fused into multi-channel images as the input of the following 2D DCNN models. Then, the 2D DCNN models automatically implement secondary feature extraction by several convolutional and pooling layers. The obtained feature maps are transformed into 1D features through a flattening operation, and the fully connected layers further map the features into a lower dimension. Lastly, fault pattern recognition is accomplished with the softmax classifier to acquire fault diagnosis results.

Deep learning networks are complex nonlinear mapping systems that rely on the activation functions of network layers to achieve the effect of nonlinear mapping. However, the loss of important information is inevitable in the process of feature extraction layer by layer. For raw high-dimensional samples with rich information in particular, a large amount of fault information is easily lost during feature extraction, and this loss affects the results of fault identification to a certain extent. In the proposed method, several 1D DCNN models with different activation functions can be used to extract features from the original high-dimensional samples by using different nonlinear mapping approaches. These features are combined to obtain redundant fusion features. Compared with the features extracted from a single 1D DCNN model, fusion features contain more abundant fault information. Therefore, the proposed ensemble model can partially reduce the loss of fault information contained in high-dimensional samples during nonlinear dimensionality reduction, an improved fault identification effect can be achieved.

In addition, several visualization methods or tools can be used to reveal the extracted feature of models and ensure the reliability of fault diagnosis results. For instance, t-distributed stochastic neighbor embedding (t-SNE) [29] is applied to project learned features onto 2D space and show the feature learning capability of models. The fault identification effects can be completely demonstrated by confusion matrices and related charts.

3 Validation of the proposed method

Rolling bearing and gear crack severity fault diagnosis experiments were performed to verify the effectiveness of the proposed method. The experiments were conducted with MATLAB R2014a by using a computer with an Intel® Core™ i7-8550U processor and 16 GB of RAM.

3.1 Case study 1

The public bearing dataset acquired from the Bearing Data Center of Case Western Reserve University (CWRU), USA was used in this experiment [30]. The rolling bearing experiment rig was composed of a 1491.4 W three-phase motor, a torque sensor, and a loading motor. The test bearing (6205-2RS JEM SKF) was mounted on the driving end, and the acceleration sensor was attached to the housing of the driving motor.

3.1.1 Dataset description

The data adopted in this experiment were obtained in the following working conditions: The motor loads were 0, 1, 2, and 3 HP (1 HP≈735 W), and the corresponding speeds were 1797, 1772, 1750, and 1730 r/min, respectively. The sampling frequency was 12 kHz. The types of samples covered normal (N), inner race (IR) fault, outer race (OR) fault, and rolling element (RE) fault. Three defect diameters (0.007, 0.014, and 0.021 inch (1 inch = 2.54 cm)) were used for each fault type. In accordance with the above-mentioned order, 10 health states were labeled as 0 to 9 (e.g., 0 represents the normal state and 9 represents a rolling element fault with a fault size of 0.021 inch). In this experiment, 10 health state samples under four working conditions (load of 0, 1, 2, and 3 HP) were utilized.

Sequential 4096 data points were used as one sample to construct the high-dimensional samples. Interval sampling was applied to simulate the sample distributions in the real scene, and the interval was set to 0 between two samples. The dataset contained 1000 samples established to test the MMEDL model's performance in identifying high-dimensional fault samples in different working conditions. Furthermore, the samples were scrambled and divided into training, validation, and test datasets at a ratio of 6:2:2.

3.1.2 Activation function selections

The high-dimensional samples contained complicated fault information, which is difficult to mine fully. Thus, this study selected three 1D DCNN models with different activation functions to achieve diversified nonlinear

mapping effects on the original high-dimensional samples and fully mine the fault characteristics. Then, the feature extraction results from the three models were fused as the input of the 2D DCNN model. Activation function selections of the 2D DCNN model can be determined according to the recognition effect of the 1D DCNN models. For instance, the 1D DCNN model using tanh as the activation function has a good recognition effect, so the convolutional layer activation functions of the 2D DCNN model also use tanh. The activation function settings of the three 1D DCNN models and one 2D DCNN model are shown in Table 2.

Table 2 Activation function settings of the four models

Layer	Activation function			
	1st 1D DCNN	2nd 1D DCNN	3rd 1D DCNN	2D DCNN
input	–	–	–	–
conv1	relu	elu	tanh	tanh
pool1	–	–	–	–
conv2	relu	elu	tanh	tanh
pool2	–	–	–	–
conv3	relu	elu	tanh	tanh
pool3	–	–	–	–
FC1	–	–	–	–
FC2	relu	relu	relu	relu
output	–	–	–	–

Note: conv1, conv2, and conv3, convolutional layers; pool1, pool2, and pool1, max pooling layers; FC1 and FC2, fully-connected layers; output, output layer.

3.1.3 Model design of MMEDL

Two kinds of network structures (1D DCNN and 2D DCNN) need to be designed in the proposed MMEDL method, and the design requirements are different. The network structure design of the 1D DCNN model is closely related to the effectiveness of preliminary feature extraction, which requires high training speed and avoids the loss of critical fault features. Therefore, the 1D DCNN models can adopt few convolution kernels and large convolution kernel size. For the 2D DCNN models in MMEDL, they need to perform deep secondary feature mining on the input feature images to obtain the hidden fault information. As a result, numerous filters should be used to accomplish feature extraction and improve identification accuracy. To a certain extent, the deeper the structure of DCNN models is, the stronger the expression capability is. However, complex network structures often require sufficient training data to support model training, which is difficult to achieve in real fault diagnosis scenarios. Sophisticated structures require numerous hyper-parameters to be set, which makes it difficult to determine the most appropriate network structure. In addition, the selected hyper-parameters of the above-mentioned model still need to be tested and adjusted. In this case study, the detailed parameters of the two DCNN models are listed in Table 3. A dropout layer [31] with a ratio of 0.25 was set behind the full connection layer to improve the robustness of the models.

3.1.4 Results

The training process of the MMEDL model is divided into two parts: One is to directly train the three 1D DCNN models by using raw samples, as displayed in Figs. 3(a)–3(c), and the other is to train the 2D DCNN model by using the output features of the 1D DCNN models, as displayed in Fig. 3(d). During the two training parts, the procedure wherein a model is trained on the entire training set is called one “epoch.” In the training process of the three 1D DCNN models, significant differences existed between training and validation accuracies, which prove the

Table 3 Details of the 1D DCNN and 2D DCNN models in case study 1

Layer	1D DCNN		2D DCNN	
	Parameter size	Output size	Parameter size	Output size
input	–	4096×1	–	32×32×3
conv1	8×1×4	4096×4	4×4×8	32×32×8
pool1	4×1	1024×4	2×2	16×16×8
conv2	4×1×8	1024×8	4×4×16	16×16×16
pool2	4×1	256×8	2×2	8×8×16
conv3	4×1×16	256×16	2×2×32	8×8×32
pool3	4×1	64×16	2×2	4×4×32
FC1	–	1024×1	–	512×1
FC2	1024×1024	1024×1	512×256	256×1
DR	–	1024×1	–	256×1
output	1024×10	10×1	256×10	10×1

Note: conv1, conv2, and conv3, convolutional layers; pool1, pool2, and pool1, max pooling layers; FC1 and FC2, fully-connected layers; DR, dropout layer; output, output layer.

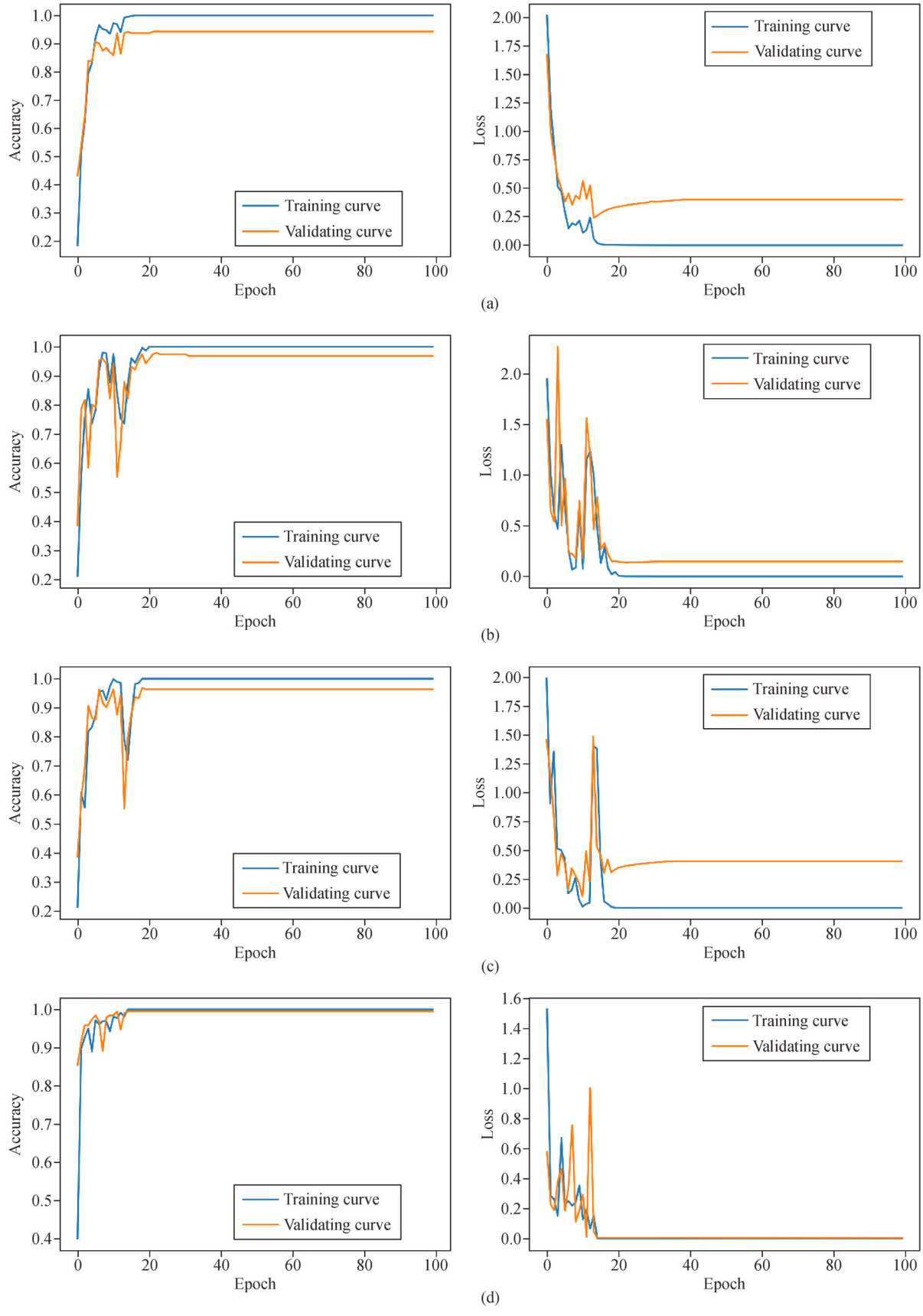

Fig. 3 Training and validation accuracies/loss curves of the multi-model ensemble deep learning model: 1D DCNN using (a) relu, (b) elu, and (c) tanh; (d) 2D DCNN.

phenomenon of overfitting. Furthermore, the 1D DCNN models appeared to have insufficient generalization capability in the process of fault identification of high-dimensional samples. By contrast, the 2D DCNN model integrated the features extracted by the three models, and its training and verification accuracies were nearly 100%.

Three different deep learning methods, namely, 1D DCNN, 2D DCNN, and deep convolutional autoencoder (DCAE), were used to compare with the MMEDL method [32]. For the 1D DCNN comparison model, the input dimension is 4096×1, and the other network hyperparameter settings are similar to those in Table 3. The inputs of 2D DCNN and DCAE are images with 64×64 dimensions and are transformed from 4096×1 raw data. Visualization of the learned features by t-SNE was performed with the test dataset, as shown in Fig. 4, to demonstrate the feature extraction capability of MMEDL and the three comparison methods. The horizontal and vertical coordinates in Fig. 4 respectively referred to the first and second principal component features obtained after dimensionality reduction by t-SNE. And the features used for visualization were extracted from the full connection layer of the four models. The results showed that most of the samples in different health conditions could not separate from each other when 1D DCNN or DCAE was used. When 2D DCNN was adopted, three clusters were difficult to distinguish clearly. Evidently, the three comparison methods could not extract the most critical features to achieve the best visualization effect. By contrast, the features of 10 different health states were completely separated when the MMEDL method was used. The features of the same health state were also clustered well.

The confusion matrices in Fig. 5 further demonstrate that the MMEDL method exhibited excellent fault identification and generalization capability. The test accuracies of 1D DCNN, 2D DCNN, and DCAE were 95%, 95.5%, and 96%, respectively, indicating that the three models were relatively good in processing high-dimensional samples. By contrast, the MMEDL method ensembled four DCNN models to mine the fault features deeply from high-dimensional sample space, and its test accuracy could reach 99.48%. Only one fault sample of the rolling element

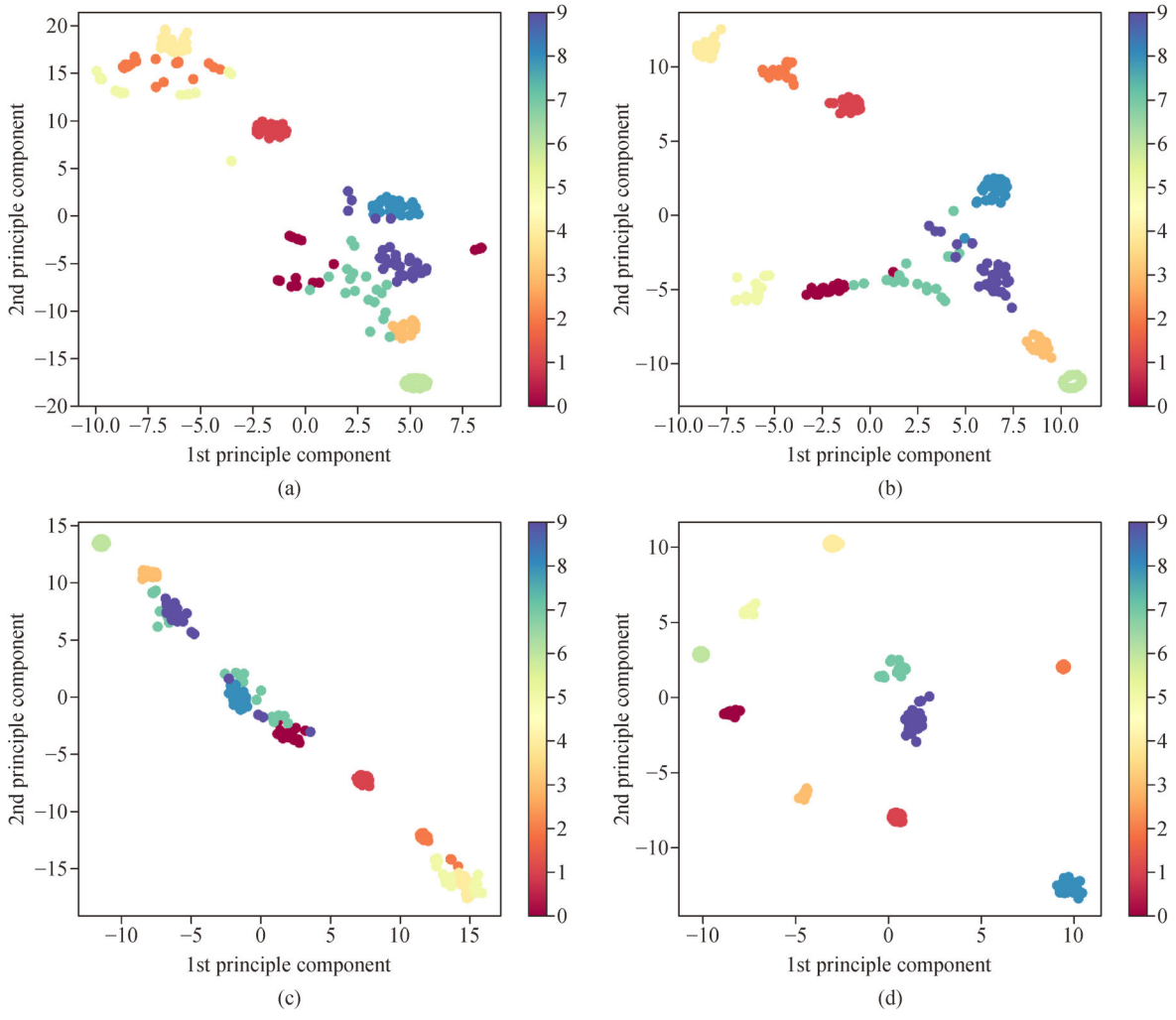

Fig. 4 Visualization effect by t-SNE of the learned features under the test dataset: (a) 1D DCNN, (b) 2D DCNN, (c) DCAE, and (d) MMEDL.

with a fault size of 0.021 inch was incorrectly identified as 0.014 inch. We conclude that the proposed MMEDL method has remarkable superiority in processing high-dimensional samples compared with the other three methods.

3.2 Case study 2

A gear dataset with different crack severities was used to further validate the effectiveness of the MMEDL method. In this experiment, the gear dataset was obtained from a condition monitoring platform for a one-stage reduction gearbox, which was set up by Huazhong University of Science and Technology, China. References [4,28] utilized the gear dataset collected from this platform to conduct their case studies.

3.2.1 Experimental setup and dataset description

The experimental platform comprised a one-stage reduction gearbox, a servo motor, a magnetic power brake, a torque sensor, and a brake controller, as shown in Fig. 6. Three triaxial accelerometers (PCB-356A16) were installed on the experimental platform at different locations, as shown in Fig. 7. The key parameters of the driven and driving gears used in the one-stage reduction gearbox are listed in Table 4. Wire-electrode cutting technology was used to construct four kinds of gear crack conditions (non-crack, 1/4 crack, 1/2 crack, and 3/4 crack), as shown at the end of the flowchart in Fig. 2. The length of the gear crack can be calculated by $L_c = i \times (R_c - r_h)/4$, $i = 0, 1, \ldots, 3$, where R_c and r_h are the radius of the root circle of the main driving wheel (27.5 mm) and the radius of the center hole (47.5 mm), respectively.

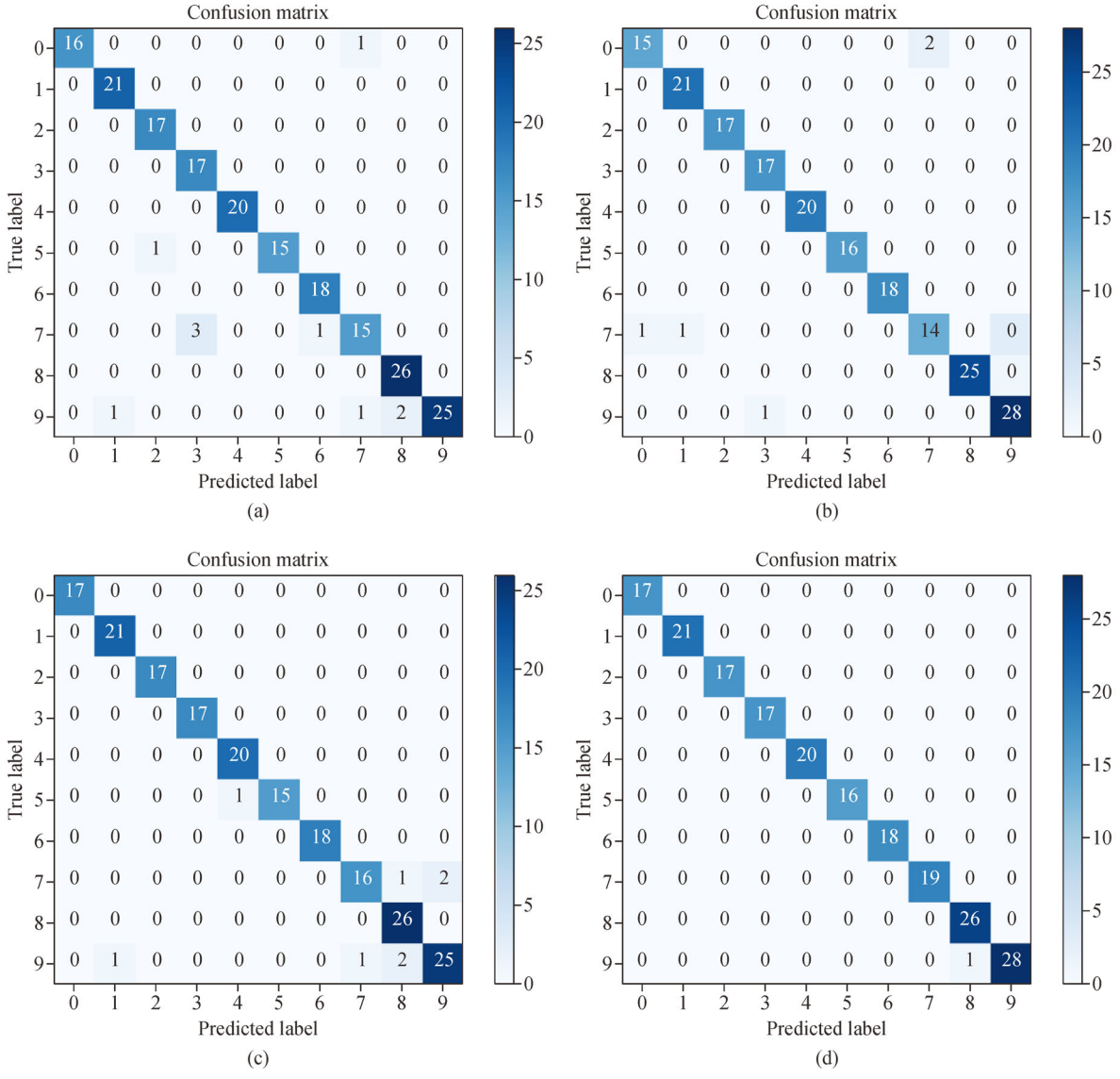

Fig. 5 Confusion matrices of the fault diagnosis results under the test dataset: (a) 1D DCNN, (b) 2D DCNN, (c) DCAE, and (d) MMEDL.

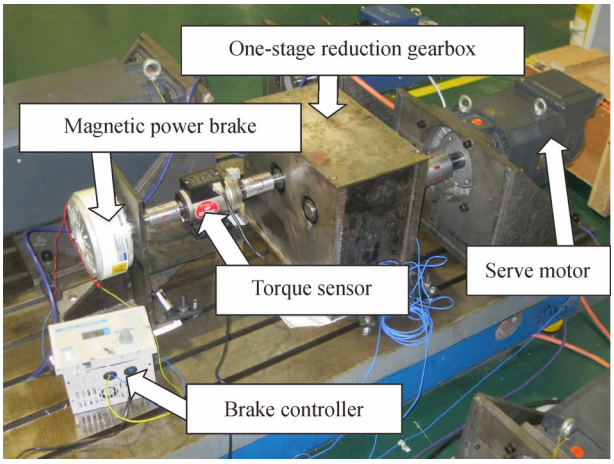

Fig. 6 Composition of the monitoring platform.

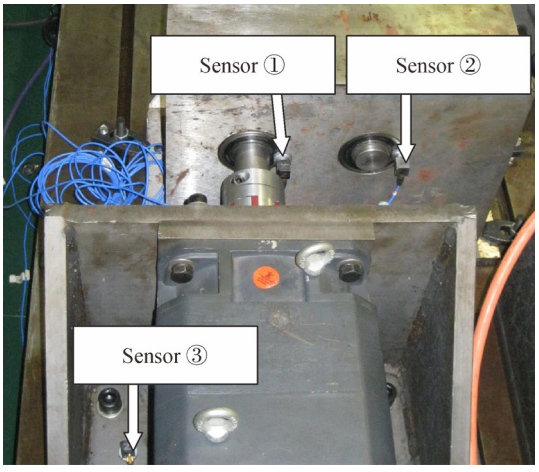

Fig. 7 Installation position of three triaxial sensors.

Table 4 Parameters of the experimental gears

Gear type	Teeth number	Gear module /mm	Teeth width /mm
Driving gear	50	2	20
Driven gear	80	2	20

In this experiment, the fault recognition capability for different gear crack severities under multiple working conditions was verified, as shown below:

Crack length: 0, 5, 10, and 15 mm;

Input shaft speed: 300, 600, 900, 1200, and 1500 r/min;

Load: 0 and 4 N·m;

Number of samples: 880;

Points of each sample: 2048.

Five input shaft speed conditions of the driving gear and two kinds of loads were used. Thus, condition monitoring was carried out under 10 operating working conditions for each gear crack severity. Sampling was implemented using a data recorder unit (NI PXI-1042) with a 5000 Hz sampling rate to acquire raw acceleration signals. Eighty-eight samples were obtained for each working condition, and each sample had 2048 data points. Thus, 880 samples were collected under 10 operating working conditions, and each kind of crack severity had 220 samples. The numbers of training, validation, and test samples were 560, 160, and 160, respectively.

3.2.2 Details of MMEDL

In this case study, the selected activation functions for the 1D DCNNs and 2D DCNN were similar to those in Table 2. The details of the three 1D DCNN models are shown in Table 5. For the 2D DCNN model in this case, the details of network structures were similar to those in Table 3, except that the output layer's size was set to 4×1.

Table 5 Details of the 1D DCNN models in case study 2

Layer	Parameter size	Output size
input	–	2048×1
conv1	8×1×4	2048×4
pool1	4×1	512×4
conv2	4×1×8	512×8
pool2	4×1	128×8
conv3	4×1×16	128×16
pool3	2×1	64×16
FC1	–	1024×1
FC2	1024×1024	1024×1
DR	–	1024×1
output	1024×4	4×1

Note: conv1, conv2, and conv3, convolutional layers; pool1, pool2, and pool1, max pooling layers; FC1 and FC2, fully-connected layers; DR, dropout layer; output, output layer.

3.2.3 Results

In this case study, fast Fourier transform was employed to transform the raw acceleration signals from the time domain to the frequency domain. Hence, the trained models could pay sufficient attention to the fault characteristic frequencies and filter out the external interference frequencies. In particularly, in order not to change the dimensions of the samples, the original time-domain signals were only treated with fast Fourier transform and its true frequency and amplitude did not need to be calculated. The fast Fourier transform results of the time domain signals of the four kinds of gear crack severity are shown in Fig. 8. Then, the MMEDL model was used to implement feature extraction and fault identification from the frequency samples. After the feature

extraction of the three 1D DCNN models, the 2D DCNN model was used to acquire the fault recognition results by feature learning. The training and validation accuracies/loss curves of the 2D DCNN model are shown in Fig. 9. The accuracy and loss curves converged after about 20 iterations of the 2D DCNN model. Moreover, the training and verification accuracies were relatively high (nearly 100% and 99.5%, respectively). The test accuracy could reach 98.75%.

The acceleration signals from channel 5 were used in the experimental process. The raw signals from the eight other channels were also used in this experiment to further verify the effectiveness of the MMEDL method. Meanwhile, the deep neural network [33], deep belief network [34], and 1D DCNN were compared with the proposed method. The signals of each channel were applied for 10 tests. The upper limits, lower limits, and median values of the test dataset identification accuracies are shown in Fig. 10. Significant differences in test accuracy were observed between the different channels because the signal quality of the nine channels differed. Overall, the MMEDL method had better accuracy than the three other methods under multi-channel signals.

4 Conclusions

Given the fact that the complicated fault features in high-dimensional samples are difficult to extract, a single deep learning method is generally unable to learn complete fault information. In view of this situation, this study proposed a novel fault identification method called MMEDL for high-dimensional samples. The presented method is ensembled by several DCNN models, which can enrich the feature spaces of high-dimensional raw samples and avoid the loss of critical information. Moreover, two experimental studies were conducted on the bearing public dataset of CWRU and a one-stage reduction gearbox dataset. The

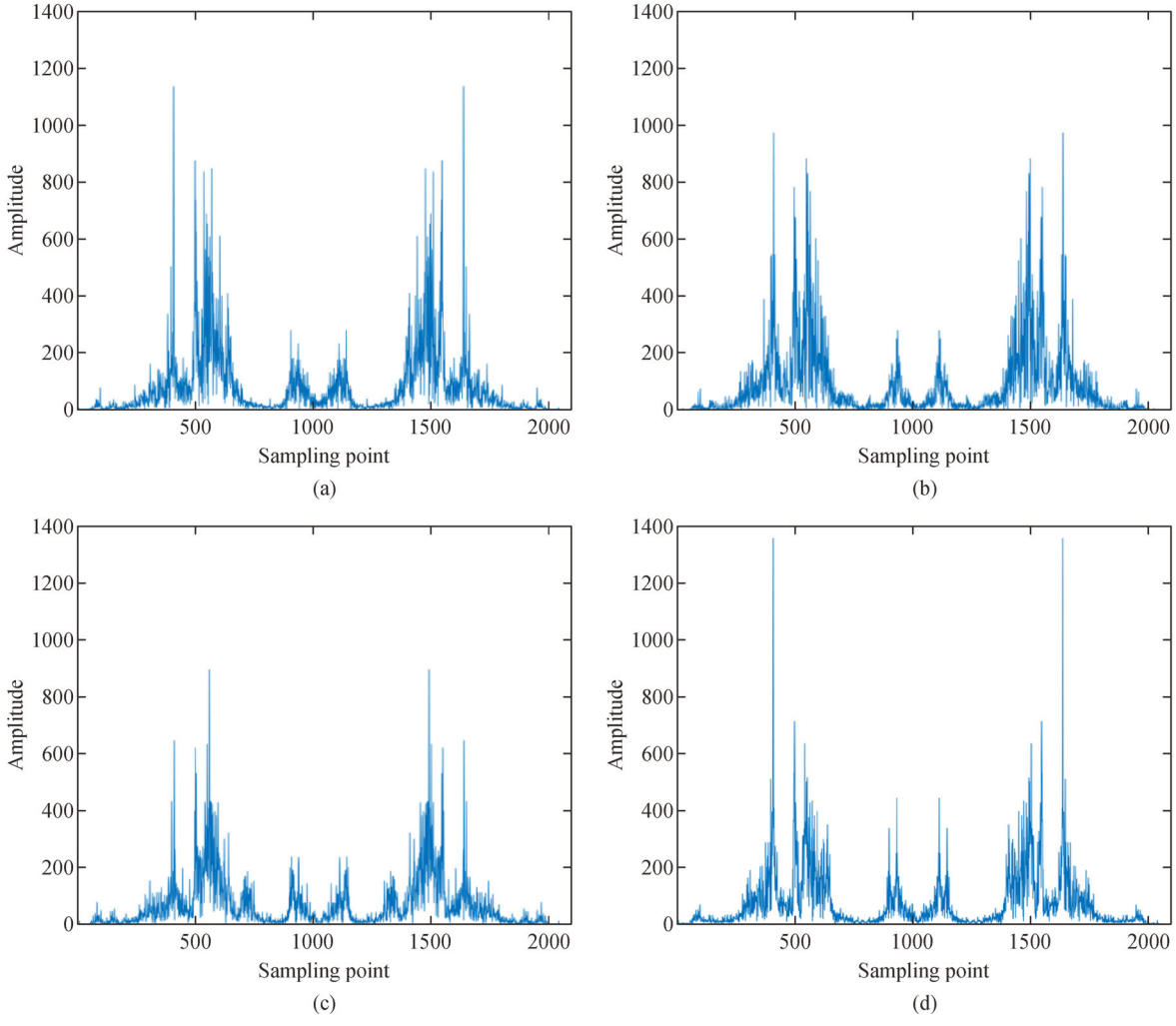

Fig. 8 Fast Fourier transform results the four gear crack severities: (a) Non-crack, (b) 1/4 crack, (c) 1/2 crack, and (d) 3/4 crack.

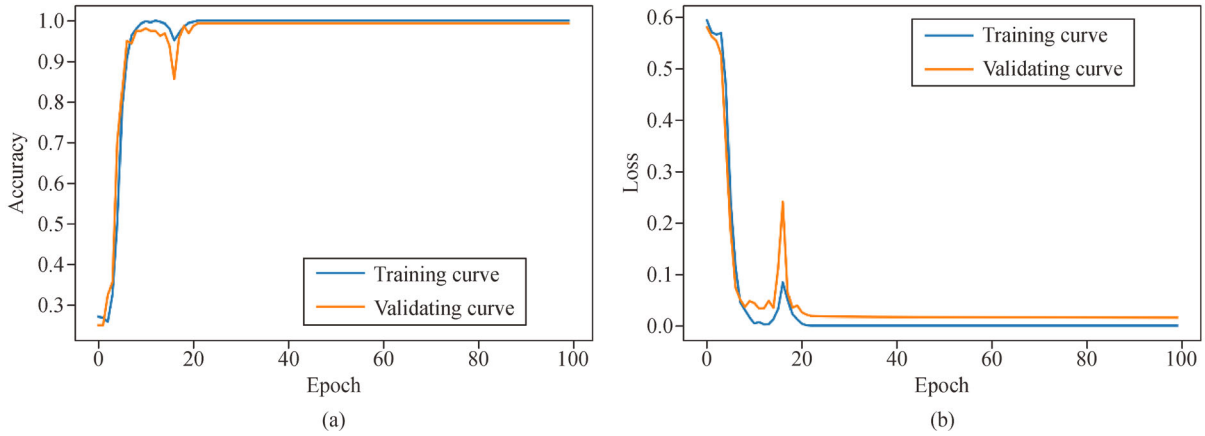

Fig. 9 (a) Training and validation accuracies and (b) loss curves of the 2D deep convolutional neural network in the proposed multi-model ensemble deep learning.

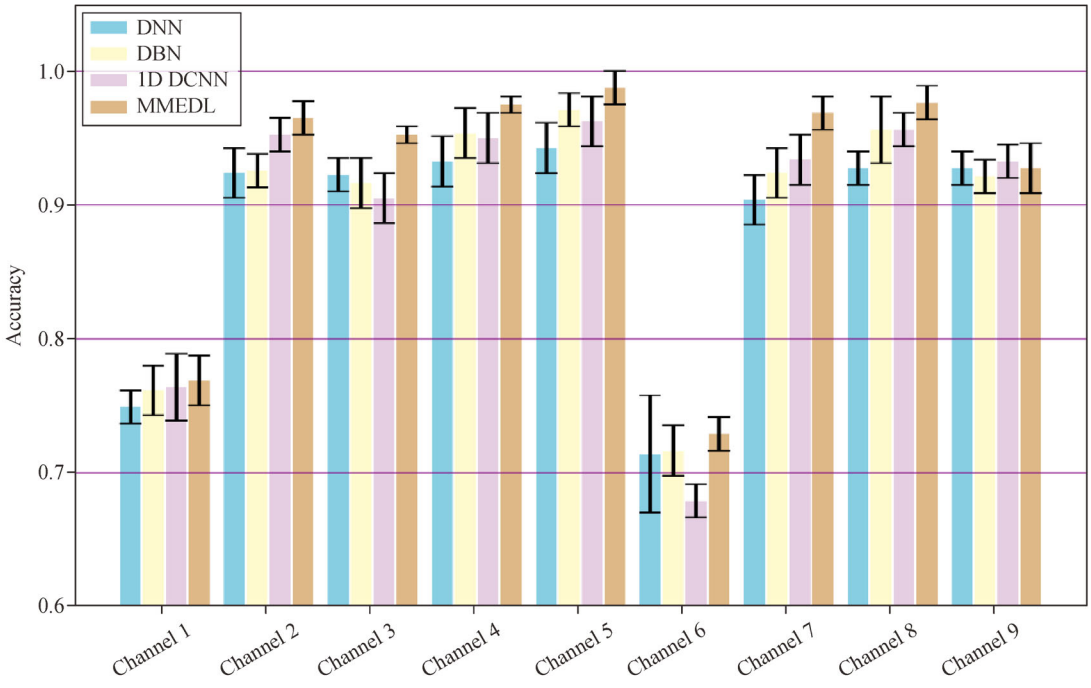

Fig. 10 Accuracy comparison of the different methods under the multi-channel signals. DNN: Deep neural network; DBN: Deep belief network; DCNN: Deep convolutional neural network; MMEDL: Multi-model ensemble deep learning.

experimental results showed that MMEDL has superior feature extraction and fault identification capability. Compared with the other deep learning methods, the MMEDL model can considerably improve the accuracy of fault diagnosis.

Selecting the most appropriate activation functions can make the MMEDL model as accurate as possible. In this work, the activation functions of three 1D DCNN models were selected by artificial settings. In the future, how to choose the most suitable activation functions in an intelligent manner should be further considered. Moreover, compared with several deep learning methods [20,24,35] for mechanical fault diagnosis in recent years, the presented MMEDL model generally has higher time consumption. Thus, another important future task is to explore how to reduce the training time consumption while ensuring model precision. In addition, whether the proposed method can achieve high identification accuracy under other fault diagnosis backgrounds remains unclear. Thus, additional experimental studies should be performed to further verify the robustness and generalization of the proposed method.

Nomenclature

1D	One-dimensional
2D	Two-dimensional
conv	Convolutional layer
CNN	Convolutional neural network
DBN	Deep belief network
DCAE	Deep convolutional autoencoder
DCNN	Deep convolutional neural network
DNN	Deep neural network
DR	Dropout layer
FC	Fully-connected layer
IR	Inner race fault state of the bearing in case study 1
MMEDL	Multi-model ensemble deep learning
N	Normal state of the bearing in case study 1
OR	Outer race fault state of the bearing in case study 1
pool	Max pooling layer
RE	Rolling element fault state of the bearing in case study 1
SVM	Support vector machine
t-SNE	t-distributed stochastic neighbor embedding
b	bias of the neurons in Eq. (1)
d_k	Length of the trainable filters in Eq. (1)
k	Trainable filters in Eq. (1)
L_c	Length of the gear crack in case study 2
R_c	Radius of the root circle of the main driving wheel in case study 2
r_h	Radius of the center hole of the main driving wheel in case study 2
x	Input of the convolution layer in Eq. (1)
s	Nonlinear activation functions in Eq. (1)

Acknowledgements This study was financially supported by the National Key R&D Program of China (Grant No. 2017YFD0400405).

References

1. Liu R, Yang B, Zio E, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing, 2018, 108: 33–47
2. Zhang H. Fault diagnosis and life prediction of mechanical equipment based on artificial intelligence. Journal of Intelligent & Fuzzy Systems, 2018, 37(3): 3535–3544
3. Jia F, Lei Y, Guo L, et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. Neurocomputing, 2018, 272: 619–628
4. Liu J, Hu Y, Wang Y, et al. An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis. Measurement Science & Technology, 2018, 29(5): 055103
5. Azamfar M, Li X, Lee J. Intelligent ball screw fault diagnosis using a deep domain adaptation methodology. Mechanism and Machine Theory, 2020, 151: 103932
6. Ainapure A, Li X, Singh J, et al. Enhancing intelligent cross-domain fault diagnosis performance on rotating machines with noisy health labels. Procedia Manufacturing, 2020, 48: 940–946
7. Zhang X, Guo S, Jiang L. Semi-supervised fault identification based on improved Laplace feature mapping and constraint seed K-means. Journal of Vibration and Shock, 2019, 38(16): 93–99 (in Chinese)
8. Widodo A, Yang B S. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 2007, 21(6): 2560–2574
9. Sun Y, Zhang S, Miao C, et al. Improved BP neural network for transformer fault diagnosis. Journal of China University of Mining and Technology, 2007, 17(1): 138–142
10. Liu J, Hu Y, Wu B, et al. A hybrid generalized hidden markov model-based condition monitoring approach for rolling bearings. Sensors (Basel), 2017, 17(5): 1143
11. Hoang D T, Kang H J. A survey on deep learning based bearing fault diagnosis. Neurocomputing, 2019, 335: 327–335
12. Li X, Li J, Qu Y, et al. Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning. Chinese Journal of Aeronautics, 2020, 33(2): 418–426
13. Zhang W, Li C, Peng G, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mechanical Systems and Signal Processing, 2018, 100: 439–453
14. Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks, 2015, 61: 85–117
15. Tang S, Yuan S, Zhu Y. Deep learning-based intelligent fault diagnosis methods toward rotating machinery. IEEE Access: Practical Innovations, Open Solutions, 2020, 8: 9335–9346
16. Jing L, Wang T, Zhao M, et al. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. Sensors (Basel), 2017, 17(2): 414
17. Zhu J, Hu T, Jiang B, et al. Intelligent bearing fault diagnosis using PCA-DBN framework. Neural Computing & Applications, 2020, 32(14): 10773–10781
18. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324
19. Zhao D, Wang T, Chu F. Deep convolutional neural network based planet bearing fault classification. Computers in Industry, 2019, 107: 59–66
20. Chen H, Hu N, Cheng Z, et al. A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. Measurement, 2019, 146: 268–278
21. Wu C, Jiang P, Ding C, et al. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. Computers in Industry, 2019, 108: 53–61
22. Kumar A, Zhou Y, Gandhi C P, et al. Bearing defect size assessment using wavelet transform based deep convolutional neural network (DCNN). Alexandria Engineering Journal, 2020, 59(2): 999–1012
23. Li Y, Du X, Wan F, et al. Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging. Chinese Journal of Aeronautics, 2020, 33(2): 427–438
24. Wen L, Li X, Gao L, et al. A new convolutional neural network-

based data-driven fault diagnosis method. IEEE Transactions on Industrial Electronics, 2018, 65(7): 5990–5998

25. Xue Y, Dou D, Yang J. Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine. Measurement, 2020, 156: 107571
26. Hu Z X, Wang Y, Ge M F, et al. Data-driven fault diagnosis method based on compressed sensing and improved multiscale network. IEEE Transactions on Industrial Electronics, 2020, 67(4): 3216–3225
27. Li H, Huang J, Ji S. Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network. Sensors (Basel), 2019, 19(9): 2034
28. Zhou Q, Li Y, Tian Y, et al. A novel method based on nonlinear auto-regression neural network and convolutional neural network for imbalanced fault diagnosis of rotating machinery. Measurement, 2020, 161: 107880
29. Pezzotti N, Lelieveldt B, Maaten L, et al. Approximated and user steerable tSNE for progressive visual analytics. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(7): 1739–1752
30. Smith W A, Randall R B. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. Mechanical Systems and Signal Processing, 2015, 64–65: 100–131
31. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1): 1929–1958
32. Agostini C E, Sampaio M A. Probabilistic neural network with bayesian-based, spectral torque imaging and deep convolutional autoencoder for PDC bit wear monitoring. Journal of Petroleum Science Engineering, 2020, 193: 107434
33. Jia F, Lei Y, Lin J, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mechanical Systems and Signal Processing, 2016, 72–73: 303–315
34. Zhang X, Guo S, Li Y, et al. Semi-supervised fault identification based on laplacian eigenmap and deep belief networks. Journal of Mechanical Engineering, 2019, 56(1): 69
35. Gong W, Chen H, Zhang Z, et al. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. Sensors (Basel), 2019, 19(7): 1693