Measurement 228 (2024) 114338

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Diesel engine fault diagnosis for multiple industrial scenarios based on transfer learning

Junhong Zhang [a,b], Guobin Pei [a], Xiaolong Zhu [a], Xin Gou [a], Linlong Deng [a], Lang Gao [a], Zewei Liu [a], Qing Ni [c,*], Jiewei Lin [a,*]

[a] State Key Laboratory of Engines, Tianjin University, Tianjin 300350, China
[b] Mechanical Engineering Department of Tianjin Ren'Ai College, Tianjin 301636, China
[c] School of Mechanical and Mechatronic Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:
Transfer learning
Diesel engine
Fault diagnosis
Small sample
Cross-domain

ABSTRACT

Fault diagnosis based on data-driven intelligence has recently attracted extensive interest owing to the rapid development of big data and deep-learning algorithms. However, when the amount of faulty data is limited, deep learning training is prone to overfitting. When the application scenario is changed, the generalization ability of the trained network is affected. In this study, a fault diagnosis architecture based on deep transfer learning is proposed to work with limited data and transfer between multiple scenarios. A wide convolution kernel convolutional long short-term memory neural network (WCL) was used to improve the feature extraction ability of fault data from a diesel engine with a low signal-to-noise ratio. A multiple transfer learning scheme based on WCL was further adopted to transfer the well-trained diagnostic knowledge of large-scale labeled source domain data to the target domain with limited samples. In addition, for diesel engines for various purposes, the knowledge transferability between different scenarios was studied. The algorithm evaluates the transfer performance of four different domains when the sample is insufficient, including the cross-fault type, cross-equipment type, cross-fault degree, and cross-working conditions. The results show the proposed method is proven with high noise immunity improves the accuracy of small sample cross-domain diagnosis and provides an optimal transfer scheme suitable for diesel engine fault signals.

1. Introduction

Diesel engines are a driving force in industries, agriculture, nuclear power, and other fields. The mechanical components of diesel engines are prone to faults because of their complex structures and poor working environments. The key to ensuring the safety of diesel engines is establishing a reliable fault diagnosis system. Traditional fault-diagnosis methods are based on signal processing for feature extraction and classification. Bi [1] proposed a novel diagnostic method based on variational mode decomposition (VMD) and kernel-based fuzzy c-means clustering (KFCM). Liu [2] proposed a novel approach based on improved intrinsic time-scale decomposition (ITD) and relevance vector machine (RVM) for the identification of diesel engine valve train faults. Xu [3] proposed an integrated pattern recognition algorithm, including an artificial neural network (ANN) model, a belief rule-based inference (BRB) model, and an evidential reasoning (ER) rule model. However, traditional fault diagnosis methods require expert experience, time-consuming design, and cannot guarantee versatility. In addition, it is difficult to characterize the complex mapping relationship between the measured signals and faults, which limits diagnostic accuracy.

Deep learning, one of the latest developmental directions and research trends in the field of machine learning, has brought revolutionary progress to the intelligent diagnosis of diesel engines. With the help of sufficient historical fault data, intelligent fault diagnosis establishes and trains a deep neural network model, mines the high-dimensional features contained in the original data, and reduces dependence on expert knowledge [4,5]. Chen [6] designed a neural network model called multiscale convolutional neural network-long short-term memory (CNN-LSTM) and a deep residual learning model that combined a multiscale-wide CNN-LSTM module and a deep residual module for rolling bearing fault diagnosis. Jiang [7] proposed a diesel engine operating condition recognition method based on a one-

* Corresponding authors.
E-mail addresses: zhangjh@tju.edu.cn (J. Zhang), bingopei@tju.edu.cn (G. Pei), zhuxl@tju.edu.cn (X. Zhu), gouxin@tju.edu.cn (X. Gou), dengll@tju.edu.cn (L. Deng), lang_gao@tju.edu.cn (L. Gao), liuzewei224@163.com (Z. Liu), Qing.Ni@outlook.com.au (Q. Ni), linjiewei@tju.edu.cn (J. Lin).

https://doi.org/10.1016/j.measurement.2024.114338
Received 3 November 2023; Received in revised form 29 December 2023; Accepted 15 February 2024
Available online 16 February 2024

dimensional convolutional long short-term memory network (1D-CLSTM) with a recognition accuracy of 99.08 %. Wang [8] presented a novel fault diagnosis method called a graph convolutional network (GCN) based on the distance and probability topological graph (DPGCN) model to solve the problem of imbalanced classification. However, in actual diesel engine engineering scenarios, the collected real-time data has two characteristics: (1) imbalance. Most of the time, they work in a healthy state and faults seldom occur. Therefore, the collected data were imbalanced, and the fault data were insufficient. (2) Inconsistent. The industrial application of diesel engines is complex and changeable, and fault samples of a single scenario are difficult to apply to the training of diagnostic models in other scenarios. Although simulating faults on an engine bench is a way to quickly obtain fault data, the cost of man-made faults under different operating conditions is extremely high, and running under long-term fault conditions is very risky. In particular, the environmental differences between the engine bench test and the actual operation cannot be ignored. When label samples are scarce, deep neural networks (DNNs) are prone to overfitting. When the industrial scenario was changed, the generalization performance was poor, resulting in reduced diagnostic accuracy. Therefore, in an actual diagnosis task, the key problem is how to use the historical data of different engineering scenarios for effective mining and correlation, realizing the transfer and reuse of knowledge, and improving the classification and generalization capabilities of networks under the scarcity of fault samples.

Transfer learning is one of the most effective methods for solving small-sample cross-domain fault diagnoses. Scholars at home and abroad have carried out basic research on small samples based on transfer learning, especially in rotating machinery such as bearings and gears. Zhuang [9,10] proposed an adversarial domain generalization framework with regularization learning (ADGR) and a two-stage transfer alignment (TSTA) methodto complete fault diagnosis of bearings under transfer tasks. Dong [11] proposed a diagnostic model for a marine low-speed diesel engine fuel-injection system based on the TrAdaBoost transfer-learning algorithm. Zhang [12] fine-tuned a pretrained diagnostic model based on samples under target operating conditions for motor bearing transfer scenarios under different operating conditions. Liu [13] used AlexNet and ResNet-18 convolutional networks as pretrained models to fine-tune diesel engine time–frequency graph samples. Hou [14] proposed a new transfer learning method based on simulation data focusing on the no fault data problems. Bai [15] proposed a diagnostic approach utilizing intelligent methods of optimized variational mode decomposition and deep transfer learning to perform fault diagnosis. Shao [16] directly used the pretrained AlexNet model, used a time–frequency graph to optimize the high level of the network, and improved the diagnostic accuracy of bearings for small samples. Han [17] proposed a deep adversarial transfer-network diagnosis model with good classification accuracy and generalization ability for transfer tasks with different loads. However, pre-trained models mostly apply typical existing networks. These networks have advantages in image processing, which is not a low-SNR noise signal; therefore, it is difficult to effectively solve the problem of the high noise interference of diesel engine signals. In addition, most studies focused on diagnostic scenarios with slight cross-domain differences. In practical engineering problems, the information contained in the source domain is not entirely related to that of the target domain. Therefore, the cross-domain diagnosis of diesel engines for multiple scenarios deserves more attention.

In this study, an intelligent fault diagnosis method based on deep transfer learning was proposed to solve the problem of small samples and multiple industrial scenarios. First, a wide convolution kernel convolutional long- and short-term memory neural network is constructed. Subsequently, based on the model transfer in inductive transfer learning, a small-sample fault can be accurately classified. Specifically, the source domain data were used to fully train the WCL in a strong noise environment to obtain a pretrained network with high diagnostic accuracy and excellent noise immunity. Next, we input the rare target domain data, freeze the specific layer to fine-tune the pre-training model, and make it suitable for new tasks in the target domain.

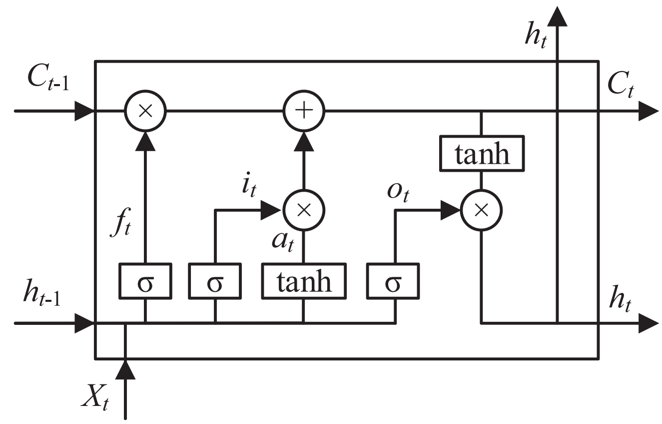


Fig. 1. LSTM Gating architecture.

The remainder of this paper is organized as follows. Section 2 briefly introduces the basic principles of CNN and LSTM. Section 3 proposes a method for solving the problem of small samples and noisy signals in a diesel engine fault diagnosis. Section 4 presents the diesel engine fault simulation experiment and describes the datasets used in this study. Section 5 presents four cross-domain diagnostic tasks and six transfer schemes to illustrate the transferability of diesel engines in different industrial scenarios, followed by conclusions in Section 6.

2. A brief introduction to CNN and LSTM

2.1. Convolutional neural network

A CNN is a multilayer deep neural network that combines low-level features to form a more abstract high-level representation. CNN have four basic properties: local feature extraction, nonlinear mapping, weight-sharing, and feature pooling. Compared with fully connected networks, CNN significantly reduce the number of trainable parameters and promote effective training without loss of expression ability. A typical CNN structure includes convolutional, activation, pooling, and fully connected layers. Among them, the convolutional and pooling layers are the unique structure of the CNN and are the key to realizing the above four basic characteristics. This study only shows the operation process of the convolutional and pooling layers in a simple form, as shown in Eq. (1).

$$\begin{cases} a^l_{\text{conv}} = \sigma\left(a^{l-1}_{\text{pool}} * W^l + b^l\right) \\ a^l_{\text{pool}} = p\left(a^l_{\text{conv}}\right) \end{cases} \tag{1}$$

where l is the number of layers, σ is the activation function, * is the convolution operation, W is the convolution kernel matrix, b is the bias, and p is the pooling operation.

2.2. Long short term memory

A recurrent neural network (RNN) is suitable for processing time series, which requires consideration of input order. LSTM is an improved RNN that solves the problems of gradient disappearance and gradient explosion during the training process of long sequences with a clever gating structure and hidden cell states. A typical LSTM architecture is shown in Fig. 1. The figure shows that, in addition to the same hidden state h_t as the RNN, another hidden state is propagated forward at each sequence index position t. This hidden state is called the Cell State, and is denoted as C_t. The LSTM also has several gates. The gates of the LSTM at each sequence index position t generally include three types: forget, input, and output gates.

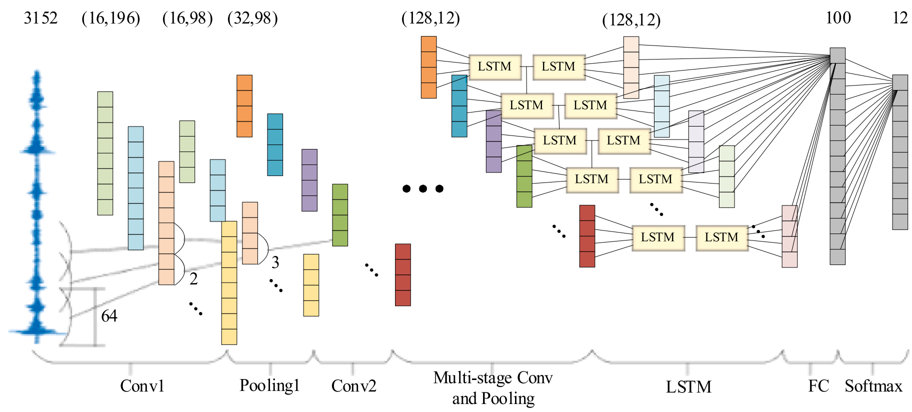


Fig. 2. The architecture of the proposed WCL.

Table 1
Structural parameters of WCL.

Block	Layer	Size of convolution kernel	Number of convolution kernels	Stride	Padding	Output dimension
B1	Conv	64 × 1	16	16	Yes	16 × 196
	BN	–	–	–	–	16 × 196
	LeakyReLU	–	–	–	–	16 × 196
	Pool	2 × 1	16	2	No	16 × 98
B2	Conv	3 × 1	32	16	Yes	32 × 98
	BN	–	–	–	–	32 × 98
	LeakyReLU	–	–	–	–	32 × 98
	Pool	2 × 1	32	2	No	32 × 49
B3	Conv	3 × 1	64	16	Yes	64 × 49
	BN	–	–	–	–	64 × 49
	LeakyReLU	–	–	–	–	64 × 49
	Pool	2 × 1	64	2	No	64 × 24
B4	Conv	3 × 1	128	16	Yes	128 × 24
	BN	–	–	–	–	128 × 24
	LeakyReLU	–	–	–	–	128 × 24
	Pool	2 × 1	128	2	No	128 × 12
B5	LSTM	–	–	–	–	128 × 12
	FC	128 × 12	–	–	–	100
	BN	–	–	–	–	100
	LeakyReLU	–	–	–	–	100
	Dropout(0.5)	–	–	–	–	100
	FC	100			–	12
	LeakyReLU	–	–	–	–	12
	Softmax	12	–	–	–	12

3. Proposed transfer learning method

3.1. Structure of the proposed WCL

Inspired by the classic CNN [18] and LSTM [19] models, a WCL was constructed to improve noise immunity. The overall architecture of the proposed WCL was similar to that of an ordinary 1D-CNN, as shown in Fig. 2. The input of the network was the original vibration signal of the diesel engine in the time domain. Four sets of convolutional and pooling layers are established to extract the local features of the original signal. The number of convolution kernels was gradually increased (16, 32, 64, and 128) to obtain more discriminative features in the higher layers of the network. After each group of convolution operations, batch normalization (BN) [20] is performed to reduce the difference in feature learning at each layer of the network. The activation function then chooses LeakyReLU [21] to solve the problem of the ReLU function not learning in the negative interval. Subsequently, 2 × 1 maximum pooling is performed to reduce the output feature size after pooling by half to reduce the complexity of the network. The network output layer was softmax, which was used to obtain the category probability output of each input sample. Finally, the parameters of each layer are updated by backpropagation based on the cross-entropy loss function. The advantages of the WCL network are as follows:1) the first convolution layer uses a wide convolution kernel and 2) an LSTM network is added after the last pooling layer. The structural parameters of the WCL are listed in Table 1.

3.1.1. Wide convolution kernel

The convolutional layer uses convolutional kernels to perform convolution operations on the local area of the input signal (or features) and generate the corresponding features. Its function is similar to that of short-time Fourier transform. A diesel engine vibration signal has a low signal-to-noise ratio, and it is difficult for a small convolution kernel to capture the characteristics of medium and low frequencies, and is

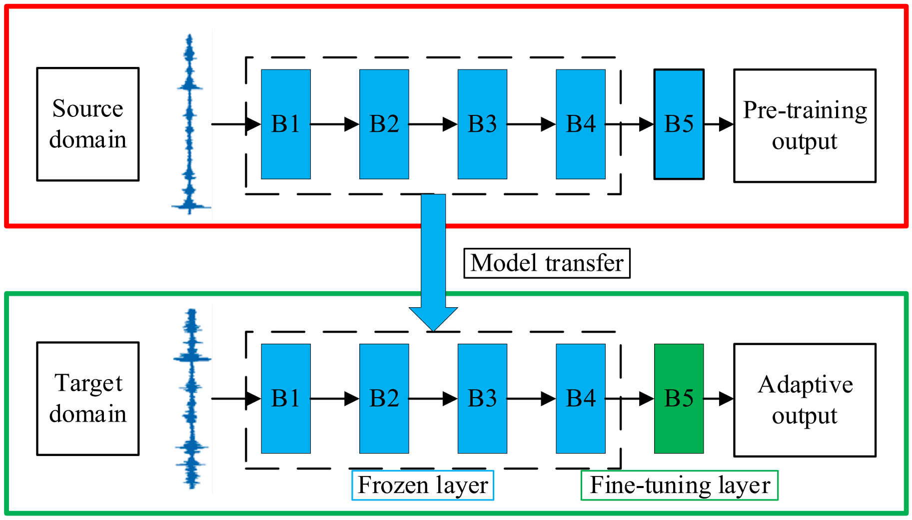


Fig. 3. Model Transfer Scheme.

susceptible to high-frequency noise. When the noise is large, choosing a larger convolution kernel is beneficial for improving the noise immunity of the model [22]. In this study, the first layer of the convolution step length was set to 16, the width of the convolution kernel was four times the step length, and the size was 64×1. Except for the first layer, the size of the convolution kernel of the other convolutional layers was 3×1 and the step size was 2. The smaller convolution kernel in the upper layer enhances the network's ability to learn the detailed features. In addition, the number of convolution cores in the last convolution layer is increased to 128 to deal with the long sample sequence.

3.1.2. LSTM parameter settings

In this study, the time-domain signal, which has an obvious time correlation, was used as the input. Feature learning is the main focus of the front part of a network to ensure comprehensiveness of fault features. In the latter part of the network, additional consideration was given to the sequence of the appearance of features to improve the fault tolerance of diesel engine diagnosis in harsh noise environments. So an LSTM network is added after the last pooling layer. The basic parameters of LSTM were set to an input size of 128, hidden size of 128, and number of layers of two.

3.2. Model transfer scheme

Model transfer based on deep neural networks is a type of inductive transfer-learning method. Problems that are good at handling have the following requirements: (1) A large number of source domain datasets accompanied by labels. (2) A small number of target domain datasets are accompanied by labels. (3) The source and target domain data come from different but similar distributions. The purpose of model transfer is to transfer the knowledge learned from the source domain to the target domain with a small amount of data to improve the small-sample classification performance of the target domain task.

The fault diagnosis process of the model transfer based on WCL is shown in Fig. 3. The first step was to fully pre-train the proposed WCL based on the constructed large-scale labeled source domain dataset using the traditional supervised deep neural network training method. The parameters of each layer of the pre-trained model obtained diagnostic knowledge of the source domain dataset. The pre-training output is guaranteed to meet the accuracy of the source domain samples. The second step is to freeze the low-level parameters of the pre-training network and use a small number of target domain training samples to fine-tune the high-level parameters, because the features learned by the deep neural network are more general in the first few layers and more specific in the high level. Finally, with the help of the model transfer, a diagnostic model suitable for the target domain is established. The adaptive output can satisfy the precision of the target domain sample.

Table 2
Parameters of the diesel engine.

Item	Parameter
Number of cylinders	6
Shape	I-shape
Firing sequence	1–5-3–6-2–4
Idle speed	950 rpm
Max output power	112 kW

Table 3
Fault simulation of 12 different operating conditions.

No.	Source domain	No.	Target domain
1	Normal	7	Exhaust pipe clogged
2	Intake valve clearance is 0.1 mm larger	8	The fuel advance Angle is increased by 3°
3	Intake valve clearance is 0.2 mm larger	9	The fifth cylinder misfired
4	Exhaust valve clearance is 0.1 mm larger	10	The sixth cylinder misfired
5	Exhaust valve clearance is 0.2 mm larger	11	The sixth cylinder performed poorly
6	The air intake filter is clogged	12	The sixth cylinder performed badly

4. Data description

4.1. Dataset A: Diesel engine a fault

The test data were obtained from a six-cylinder diesel engine, the main technical parameters are listed in Table 2.

In the test, a PCB company ICP 356A26 three-way piezoelectric acceleration sensor was used to obtain the cylinder head signal under idling and no-load conditions. The sampling frequency was 25 kHz, and the working cycle contains 3152 sampling points. Based on the above settings, the test simulated 11 common faults in the fuel system and valve train as listed in Table 3. Twelve different state single-cycle time-domain waveforms were marked as Normal, Fault1-Fault11. The layout of the test bench and fault simulation scheme are detailed in [23].

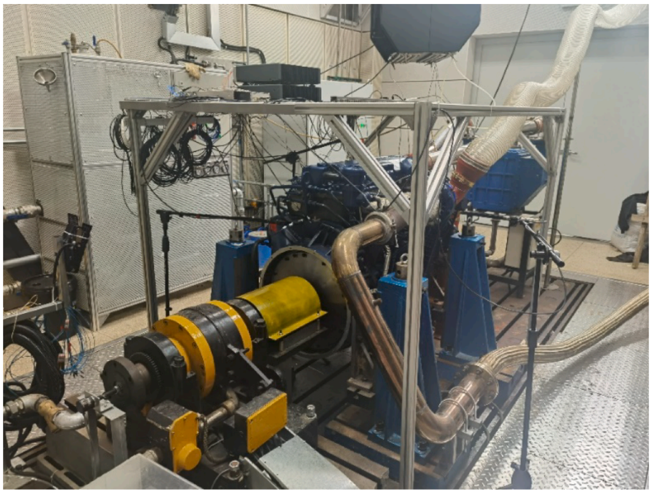


Fig. 4. Test bench layout.

4.2. Dataset B: Diesel engine B fault

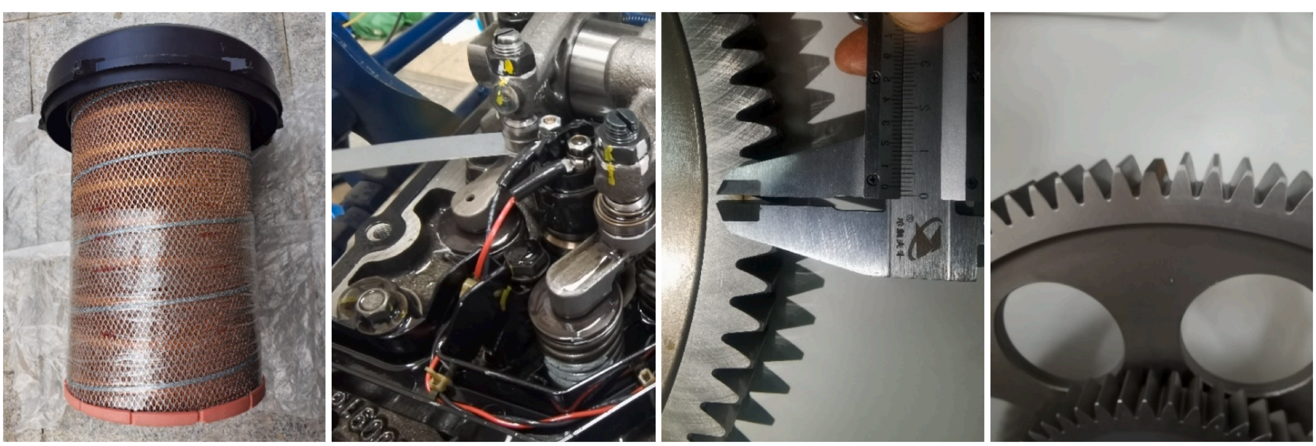
The test data were derived from another six-cylinder diesel engine that simulated various faults to varying degrees, including air intake filter blockages, abnormal valve clearances, gear cracks, gear tooth breakages, and misfires. The sampling frequency was 51.2 kHz, and the working cycle contains 5120 sampling points. The test bench is shown in Fig. 4.

The artificial fault simulation scheme is illustrated in Fig. 5. Fig. 5(a) illustrates the blocking method for the air intake filter. The filter was blocked using adhesive tape, and the blocking degrees were set to 20 %, 40 %, 50 %, and 60 %. Fig. 5(b) shows the adjustment mode of intake valve clearance. The normal clearance of this diesel engine model is approximately 0.4 mm, and two types of abnormal clearances are set in the test, 0.19 mm, and 0.59 mm. Fig. 5(c) shows a simulated gear-crack fault. The cracks were manually cut at the root of the teeth with lengths of 1, 2, and 3 mm. Fig. 5(d) shows the test gear after the artificial tooth broke. The misfire fault was controlled directly by the ECU of the diesel engine, and two fault degrees of a single cylinder and a double cylinder were set.

The data collected in the experiment were divided into one source-domain dataset and two target-domain datasets, as shown in Table 4. Compared with the source domain, the fault degree of target domain A is different and minor or rare in the actual operation. The target domain B and the source domain are only different in the operating conditions.

4.3. Dataset C: Bearing fault

The bearing failure dataset of Western Reserve University was used as the target domain dataset to verify the transfer learning effect of the proposed method between different devices [24]. There are six types of bearing faults: normal, ball, inner-race, and three outer-race faults at different locations. This study selected data with motor speeds of 1797r/min and 1772r/min under a fault diameter of 0.1778 mm for a total of 12 groups, and the labels are marked as 0–11. Table 5 presents the results.



(a) Air intake filter blockage (b) Abnormal valve clearance (c) Gear crack (d) Gear tooth breakage

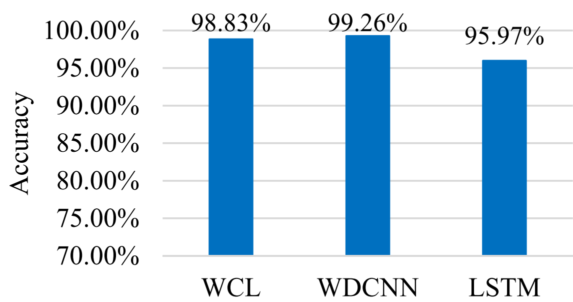
Fig. 5. Artificial fault simulation.

Table 4
Description of diesel engine B datasets.

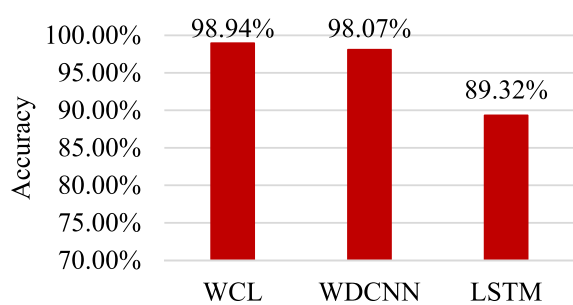
Source domain		Target domain A		Target domain B	
Working condition	Fault type	Working condition	Fault type	Working condition	Fault type
Speed:1200 rpm	Normal	Speed:1200 rpm	Normal	Speed:1000 rpm	Normal
Load:25 %	Blocking50	Load:25 %	Blocking20	Load:25 %	Blocking50
	Blocking60		Blocking40		Blocking60
	Clearance 0.19 mm		Clearance 0.36 mm		Clearance 0.19 mm
	Clearance 0.59 mm		Clearance 0.44 mm		Clearance 0.59 mm
	Gear tooth breakage		Gear crack 2 mm		Gear tooth breakage
	Gear crack 3 mm		Gear crack 1 mm		Gear crack 3 mm
	Misfire #1		Misfire #1&6		Misfire #1
	Misfire #2		Misfire #2&4		Misfire #2
	Misfire #3		Misfire #2&6		Misfire #3
	Misfire #4		Misfire #3&6		Misfire #4
	Misfire #5		Misfire #4&6		Misfire #5
	Misfire #6		Misfire #5&6		Misfire #6

Table 5
Description of rolling element-bearing datasets.

Fault diameter	Speed (r/min)	Normal	Ball fault	InnerRace fault	OuterRace fault at 6:00	OuterRace fault at 3:00	OuterRace fault at 12:00
0.1778 mm	1797	0	1	2	3	4	5
	1772	6	7	8	9	10	11



(a) Diagnostic accuracy of dataset A



(b) Diagnostic accuracy of dataset B

Fig. 6. Diagnostic accuracy of datasets A and B without noise interference.

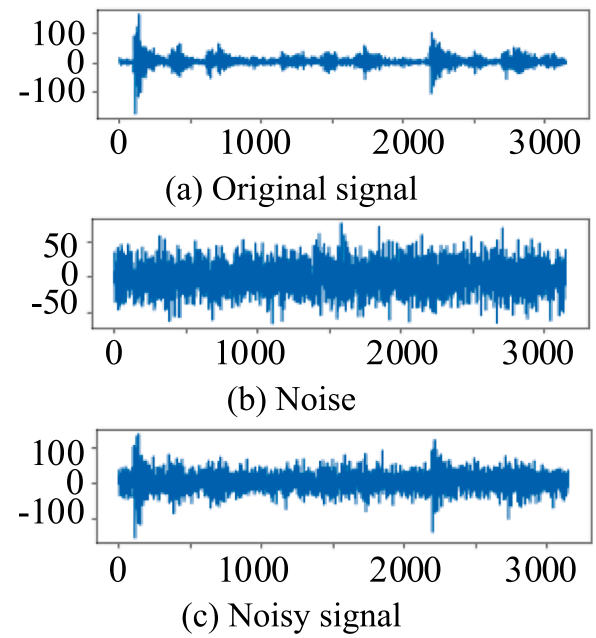


Fig. 7. Signals (a) in a normal state, (b) with an additive white Gaussian noise, and (c) with the composite noisy signal with SNR = 0 dB.

5. Validation and discussion

The deep learning framework was built using PyTorch and developed using Facebook. The computer was configured with a CPU i7 10,700 and 16 GB memory. Five patients were included in the study. In the first case, the noise immunity of the WCL model was tested. The second through fifth cases tested the diagnostic ability of the transfer model in different domains. The test divided the samples into training, validation, and test sets at a ratio of 0.5:0.25:0.25. The validation and test sets do not participate in model training and is used to monitor whether the model is overfitted to determine whether to stop training and adjust the hyperparameters. The test results for all cases, the number of samples in the training set, and the average accuracy of the test set are listed.

5.1. Case study 1: Performance of the pretrained WCL

The precision of WCL pre-training model is firstly compared and verified, which is one of the steps of transfer learning and the basis for ensuring the transfer effect. The accuracy of the proposed WCL was tested with datasets A and B and compared with three other deep learning methods: WDCNN and LSTM. The results are shown in Fig. 6.

The results showed that the average accuracies of WDCNN, WCL, and LSTM exceeded 89 %, and WDCNN and WCL were quite close in the two tests. Regardless of random noise interference, the advantage of the WCL is not obvious. To verify the noise immunity of the network further, additive white Gaussian noise [25] was added to the original signal, as shown in Eq. (2), to simulate the noise pollution of the diesel engine working environment.

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{2}$$

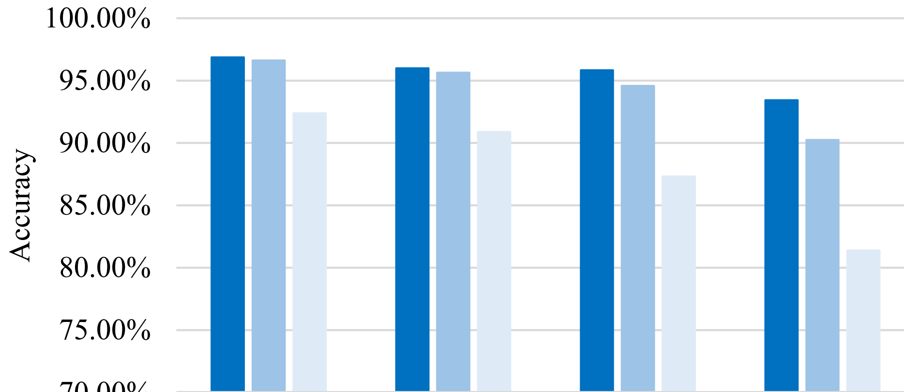
where P_{noise} and P_{signal} represent signal energy and noise, respectively. If the discrete signal is $S = \{s_1, s_2, \dots, s_n\}$, then $P_{signal} = \frac{1}{n}\sum_{k=1}^{n} S_k^2$. The greater the noise contained in the signal, the smaller the SNR value. When the SNR value was 0 dB, the energies contained in the signal and noise were equal.

The original signal, noise, and the signal with noise in the normal state are shown in Fig. 7.

The diagnostic accuracies of the three methods for different signal-to-noise ratios are shown in Fig. 8. The sample size of the three methods is 6000. It can be observed that the WCL has stronger noise immunity, and the lower the signal-to-noise ratio, the more obvious the advantage of the WCL.

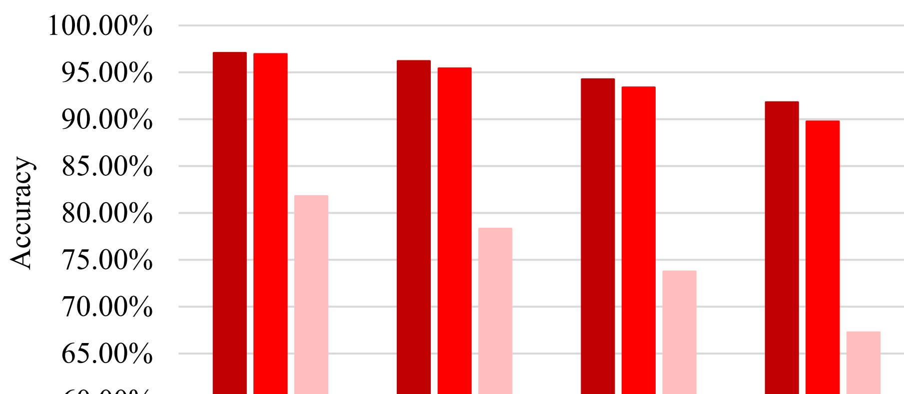
The proposed WCL network can achieve superior classification performance owing to the large amount of labeled training data. However, in actual industrial applications, it is difficult to obtain sufficient labeled data for certain research tasks. As shown in Fig. 9, in the same noise environment (SNR = 0), the diagnostic ability of the DNN decreased exponentially as the label data of the target domain task decreased. In particular, when the number of samples in each state decreased to 1000, the diagnostic accuracy of the CNN and LSTM methods was only approximately 60 %. The proposed method performed slightly better; however, its accuracy was still less than 80 %.

To solve the aforementioned sample shortage problem, the latter four cases will discuss in detail the advantages of cross-domain diagnosis based on transfer learning in small samples.



	SNR=4dB	SNR=2dB	SNR=0	SNR=-2dB
WCL	96.86%	95.99%	95.83%	93.43%
WDCNN	96.62%	95.63%	94.58%	90.24%
LSTM	92.38%	90.88%	87.32%	81.39%

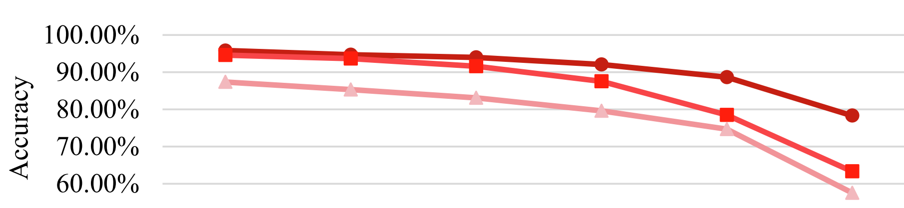
(a) Diagnostic accuracy of dataset A with noise interference.



	SNR=4dB	SNR=2dB	SNR=0	SNR=-2dB
WCL	97.08%	96.20%	94.26%	91.81%
WDCNN	96.96%	95.42%	93.39%	89.77%
LSTM	81.80%	78.32%	73.76%	67.25%

(b) Diagnostic accuracy of dataset B with noise interference.

Fig. 8. Diagnostic accuracy of datasets A and B with noise interference.



	Sample 6000×12	Sample 5000×12	Sample 4000×12	Sample 3000×12	Sample 2000×12	Sample 1000×12
WCL	95.83%	94.72%	93.94%	92.04%	88.69%	78.29%
CNN	94.58%	93.62%	91.59%	87.53%	78.50%	63.29%
LSTM	87.35%	85.30%	83.07%	79.60%	74.67%	57.54%

Sample size

Fig. 9. Diagnostic accuracy of different sample numbers.

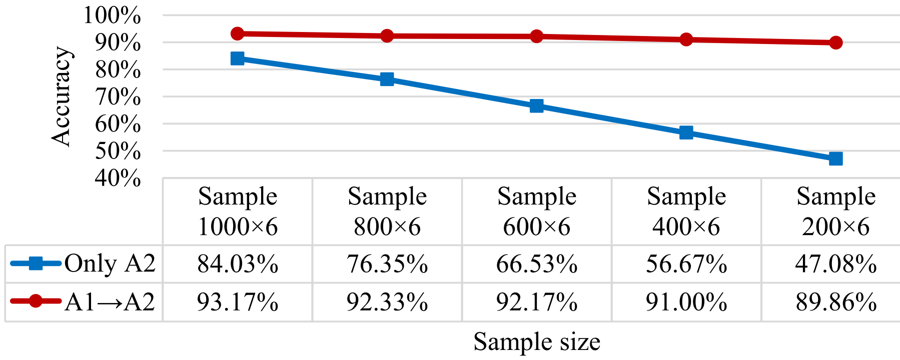


Fig. 10. Diagnosis accuracy before and after domain transfer for different fault types.

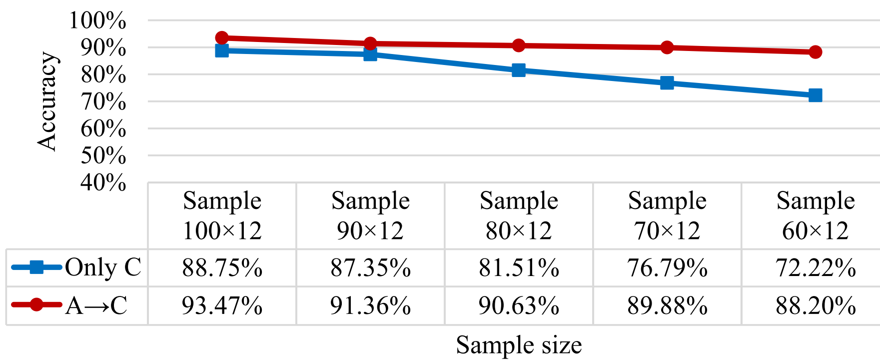


Fig. 11. Diagnosis accuracy before and after domain transfer for different types of equipment.

5.2. Case study 2: Performance across fault type

Dataset A was split into two parts to study the transfer effect of different fault type domains, named Dataset A1: Normal and fault1-5 and Dataset A2: Fault6-11. Assume that Dataset A1 represents a dataset that has a large failure rate and is easy to obtain in actual situations, and Dataset A2 represents a dataset that has a small failure rate and is difficult to obtain. To address the problem of insufficient samples in the target domain Dataset A2, Dataset A1 was used as the source domain data for the WCL model training. Dataset A2 was then used to train and fine-tune the optimization layer of the pre-training model and calculate it after adding AWGN with SNR = 0 to the dataset. The results are shown in Fig. 10.

The results showed that the effect of training A2 based only on the WCL was very poor, and the accuracy continued to decline as the number of samples decreased. In comparison, the deep transfer learning

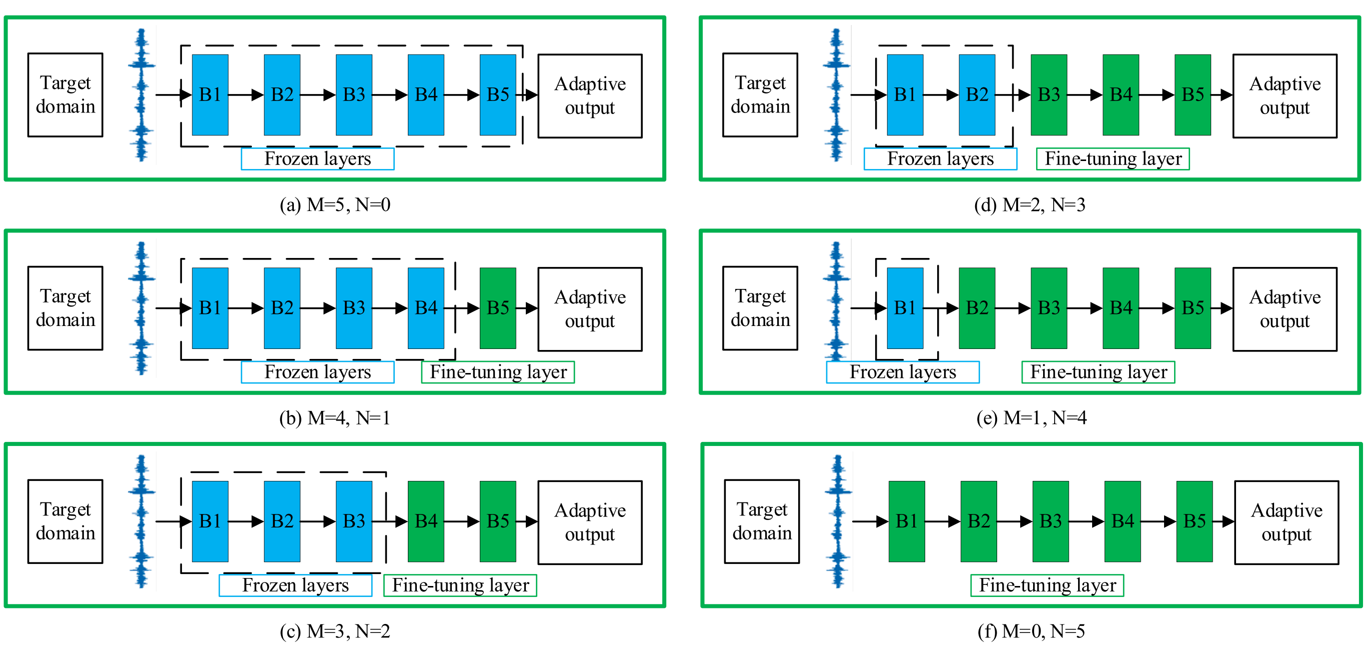
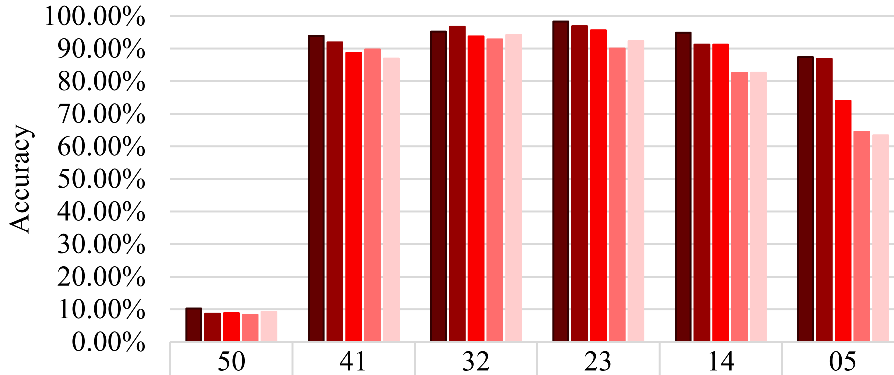


Fig. 12. Description of several transfer schemes.



	50	41	32	23	14	05
90×13	10.26%	93.93%	95.21%	98.29%	94.87%	87.35%
80×13	8.65%	91.92%	96.73%	96.88%	91.23%	86.83%
70×13	8.79%	88.68%	93.74%	95.60%	91.21%	73.96%
60×13	8.33%	89.74%	92.82%	90.00%	82.56%	64.49%
50×13	9.23%	86.92%	94.15%	92.31%	82.62%	63.38%

Fig. 13. Diagnostic accuracy of different transfer schemes for different fault degrees.

method A1 → A2 has obvious advantages in terms of accuracy and is less dependent on the number of samples. It should be noted that the diagnostic accuracy of only A2 data at 1000 samples is higher than that in Section 5.1, because this section is a 6-classification problem, whereas the previous section is a 12-classification problem. This case shows that when a fault sample is limited, the diagnosis accuracy can be effectively improved by using other fault data of the same equipment based on the proposed model to transfer the learned knowledge to the target task.

5.3. Case study 3: Performance across equipment type

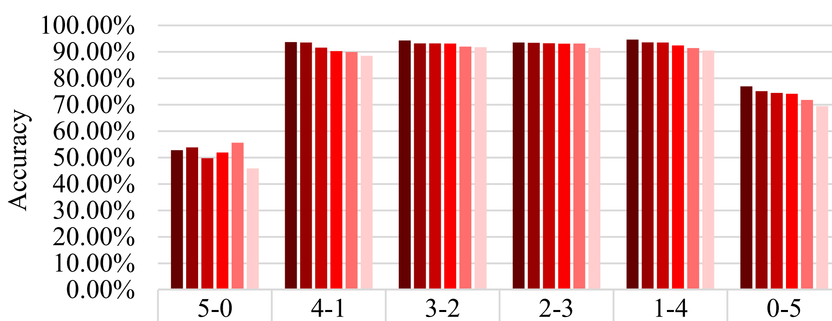
In this case, we evaluate the transfer capability of the proposed method between different devices. Fig. 11 shows the diagnosis results with diesel engine fault dataset A as the source domain, and bearing fault dataset C as the target domain. Similarly, the A → C model transfer method performed better than using only Dataset C. It is worth mentioning that, compared with the diesel engine fault data, the bearing fault data are cleaner and have a higher signal-to-noise ratio; therefore, the model requires fewer samples for the target domain after diesel engine data pre-training.

This case explains that the knowledge of diesel engine fault characteristics learned by the proposed WCL model through pre-training can also be transferred to the fault diagnosis task of the bearing. It can be proven that transfer learning across device domains provides another feasible approach for solving small-sample problems.

5.4. Case study 4: Performance across fault degree

Furthermore, the influence of different transfer schemes on the results was considered. As shown in Fig. 12, the transfer scheme can be divided into six cases according to the number of frozen and fine-tuned layers. If the number of freezing layers is M and the number of fine-tuning layers is N, the scheme is denoted as $M-N$.

This part of the analysis considered the fault degree as an example.



	5-0	4-1	3-2	2-3	1-4	0-5
190×13	52.83%	93.68%	94.29%	93.52%	94.64%	76.92%
180×13	53.85%	93.55%	93.21%	93.42%	93.59%	75.13%
170×13	49.77%	91.62%	93.17%	93.26%	93.53%	74.48%
160×13	51.92%	90.24%	93.12%	93.08%	92.40%	74.13%
150×13	55.64%	89.95%	92.00%	93.13%	91.44%	71.79%
140×13	45.88%	88.52%	91.76%	91.48%	90.49%	69.45%

Fig. 14. Diagnostic accuracy of different transfer schemes for different working conditions.

Specifically, the Source and target domains A in Dataset B were used. The results are shown in Fig. 13, where when $M = 5$ and $N = 0$, the model parameters are all frozen, indicating that the pretrained model is directly applied to the target domain dataset. Owing to the obvious distribution difference between the source and target domains, the diagnostic ability of the model is completely lost. When $M = 0$ and $N = 5$, the model parameters were retrained, and the accuracy was not satisfied by the limit of the number of samples. The results of these two special schemes indirectly illustrate the necessity of transfer learning. Although the middle four transfer schemes are all effective, schemes 3–2 and 2–3 perform better than 4–1 and 1–4. In contrast to the common use of a fully connected layer as a fine-tuning layer, it is better to appropriately add one or two fine-tuning layers in the face of low-signal-to-noise ratio signals.

5.5. Case study 5: Performance across working condition

Similarly, this part of the analysis considers the working conditions as an example. Source and target domains B in Dataset B were selected. The results are shown in Fig. 14 and are broadly similar to those of Case Study 4. The difference is that the distribution difference between different working conditions is small; therefore, more samples are required to achieve the desired accuracy compared with different fault degrees. For the same reason, the accuracy of scheme 5-0 shows a certain improvement.

6. Conclusion

The core idea of the proposed method is to transfer the diagnostic knowledge obtained by training the deep neural network with historical data of different fault type domains, equipment type domains, fault degree domains, and working condition domains to the new target domain task to improve the classification ability of the target diagnosis task in the case of scarce samples. On the one hand, the proposed method constructs a multilayer anti-noise neural network to adaptively extract features from the original vibration signal automatically, which improves the anti-noise and effectiveness of feature extraction. On the other hand, the model transfer scheme is further applied to WCL. By fixing the specific layer parameters and tuning the remaining layers, the diagnosis knowledge of the large source domain data is effectively transferred to the target domain task, which promotes fast and effective training of the target diagnosis network and improves the target domain task diagnosis performance.

The experimental results show that, compared with other deep neural network methods, the proposed method has better anti-noise performance and calculation accuracy. However, with a decrease in the target domain-labeled data, the diagnostic ability of deep neural networks will still be greatly reduced. It was also found that when a certain fault sample was limited, the diagnosis accuracy could be effectively improved by using other fault data, other working condition data, and other fault degree data of the same equipment and transferring the learned knowledge to the target task. In addition, Different from the common fine-tune full connection layer, a modest increase in the number of fine-tuning layers may be beneficial for complex signals generated by diesel-engine-type equipment. Therefore, the performance of model transfer is related to the size of the target domain training data, similarity between the target and source domains, transfer scheme, and signal-to-noise ratio. The more training data in which the target domain task participates and is more similar to the source domain task, the better the transfer performance and the higher the classification accuracy. Therefore, in addition to building a powerful deep neural network diagnostic model, the quantity and quality of the training data are still two crucial factors for improving model transfer performance.

CRediT authorship contribution statement

Junhong Zhang: Writing – original draft. **Guobin Pei:** Methodology. **Xiaolong Zhu:** Methodology. **Xin Gou:** Methodology. **Linlong Deng:** Investigation. **Lang Gao:** Investigation. **Zewei Liu:** Writing – review & editing. **Qing Ni:** Supervision. **Jiewei Lin:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author upon reasonable request.

Acknowledgment

This research was supported by the National Key R&D Program of China (Grant No.2021YFD2000303) and Tianjin Graduate Research and Innovation Fund Project (Project No. 2021YJSB182). The authors are grateful to anonymous reviewers for their helpful comments and constructive suggestions.

References

[1] X. Bi, J. Lin, D. Tang, F. Bi, X. Li, X. Yang, T. Ma, P. Shen, VMD-KFCM algorithm for the fault diagnosis of diesel engine vibration signals, Energies. 13 (2020).
[2] Y. Liu, J. Zhang, F. Bi, J. Lin, W. Ma, A fault diagnosis approach for diesel engine valve train based on improved ITD and SDAG-RVM, Meas. Sci. Technol. 26 (2015).
[3] X. Xu, Z. Zhao, X. Xu, J. Yang, L. Chang, X. Yan, G. Wang, Machine learning-based wear fault diagnosis for marine diesel engine by fusing multiple data-driven models, Knowledge-Based Syst. 190 (2020) 105324.
[4] R. Huang, Y. Liao, S. Zhang, W. Li, Deep decoupling convolutional neural network for intelligent compound fault diagnosis, IEEE Access. 7 (2019) 1848–1858.
[5] C. Sun, M. Ma, Z. Zhao, X. Chen, Sparse deep stacking network for fault diagnosis of motor, IEEE Trans. Ind. Informatics. 14 (2018) 3261–3270.
[6] H. Chen, W. Meng, Y. Li, Q. Xiong, An anti-noise fault diagnosis approach for rolling bearings based on multiscale CNN-LSTM and a deep residual learning model, Meas. Sci. Technol. 34 (2023) 045013.
[7] Z. Jiang, Y. Lai, J. Zhang, H. Zhao, Z. Mao, Multi-factor operating condition recognition using 1D convolutional long short-term network, Sensors (switzerland). 19 (2019).
[8] R. Wang, H. Chen, C. Guan, DPGCN model: a novel fault diagnosis method for marine diesel engines based on imbalanced datasets, IEEE Trans. Instrum. Meas. 72 (2023) 1–11.
[9] J. Zhuang, Y. Cao, M. Jia, X. Zhao, Q. Peng, Fault diagnosis of bearings using a two-stage transfer alignment approach with semantic consistency and entropy loss, Expert Syst. Appl. 226 (2023) 120274.
[10] J. Zhuang, M. Jia, Y. Ding, X. Zhao, Health assessment of rotating equipment with unseen conditions using adversarial domain generalization toward self-supervised regularization learning, IEEE/ASME Trans. Mechatronics. 27 (2022) 4675–4685.
[11] G. Xiong, W. Ma, N. Zhao, J. Zhang, Z. Jiang, Z. Mao, Multi-type diesel engines operating condition recognition method based on stacked auto-encoder and feature transfer learning, IEEE Access. 9 (2021) 31043–31052.
[12] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer learning with neural networks for bearing fault diagnosis in changing working conditions, IEEE Access. 5 (2017) 14347–14357.
[13] Y. Liu, J. Kang, C. Guo, Y. Bai, Diesel engine small-sample transfer learning fault diagnosis algorithm based on STFT time–frequency image and hyperparameter autonomous optimization deep convolutional network improved by PSO–GWO–BPNN surrogate model, Open Phys. 20 (2022) 993–1018.
[14] W. Hou, C. Zhang, Y. Jiang, K. Cai, Y. Wang, N. Li, A new bearing fault diagnosis method via simulation data driving transfer learning without target fault data, Meas. J. Int. Meas. Confed. 215 (2023), https://doi.org/10.1016/j.measurement.2023.112879.
[15] H. Bai, X. Zhan, H. Yan, L. Wen, X. Jia, Combination of optimized variational mode decomposition and deep transfer learning: a better fault diagnosis approach for diesel engines, Electron. 11 (2022).
[16] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Informatics. 15 (2019) 2446–2455.
[17] T. Han, C. Liu, W. Yang, D. Jiang, A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults, Knowl.-Based Syst. 165 (2019) 474–487.

[18] B.A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM. 60 (2012) 84–90.
[19] H. Sak, A. Senior, F. Beaufays, Long short-term memory recurrent neural network architectures for large scale acoustic modeling, 15th Annu. Conf. Int. Speech Commun. Assoc. Singapore. 44 (2014) 338–342.
[20] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, Proc. AAAI Conf. Artif. Intell. 42 (2017) 11–24.
[21] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network,, ArXiv Prepr. ArXiv1505.00853. (2015).
[22] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors (switzerland). 17 (2017).
[23] Y. Liu, J. Zhang, K. Qin, Y. Xu, Diesel engine fault diagnosis using intrinsic time-scale decomposition and multistage Adaboost relevance vector machine, Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 232 (2018) 881–894.
[24] X. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mech. Syst. Signal Process. 18 (2004) 1077–1095.
[25] M. Amar, I. Gondal, C. Wilson, Vibration spectrum imaging: a novel bearing fault classification approach, IEEE Trans. Ind. Electron. 62 (2015) 494–502.