Neurocomputing 417 (2020) 36–63

Contents lists available at ScienceDirect

Neurocomputing

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

A comprehensive review on convolutional neural network in machine fault diagnosis

Jinyang Jiao [a], Ming Zhao [a], Jing Lin [b,*], Kaixuan Liang [a]

[a] State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[b] School of Reliability and Systems Engineering, Beihang University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 25 February 2020
Revised 18 May 2020
Accepted 7 July 2020
Available online 1 August 2020
Communicated by Bin Zhang

Keywords:
Convolutional neural network
Machine fault diagnosis
Classification
Prediction
Transfer learning

ABSTRACT

With the rapid development of manufacturing industry, machine fault diagnosis has become increasingly significant to ensure safe equipment operation and production. Consequently, multifarious approaches have been explored and developed in the past years, of which intelligent algorithms develop particularly rapidly. Convolutional neural network (CNN), as a typical representative of intelligent diagnostic models, has been extensively studied and applied in recent five years, and a large amount of literature has been published in academic journals and conference proceedings. However, there has not been a systematic review to cover these studies and make a prospect for the further research. To fill in this gap, this work attempts to review and summarize the development of the Convolutional Network based Fault Diagnosis (CNFD) approaches comprehensively. Generally, a typical CNFD framework is composed of the following steps, namely, data collection, model construction, and feature learning and decision making, thus this paper is organized by following this stream. Firstly, data collection process is described, in which several popular datasets are introduced. Then, the fundamental theory from the basic CNN to its variants is elaborated. After that, the applications of CNFD are reviewed in terms of three mainstream directions, i.e. classification, prediction and transfer diagnosis. Finally, conclusions and prospects are presented to point out the characteristics of current development, facing challenges and future trends. Last but not least, it is expected that this work would provide convenience and inspire further exploration for researchers in this field.

1. Introduction

Powered by the integration innovation of intelligent manufacturing, industrial big data and industrial 4.0, the modern industry is experiencing a new revolution from the traditional manufacturing industry to intelligent manufacturing industry [1,2]. Mechanical equipment, as one of the most significant roles in this revolution, is evolving to continuously promote production and improve economic benefits. Unfortunately, various faults will inevitably be exposed during the tireless operation of machine, once the fault appears, it will cause unscheduled downtime, economic loss, even catastrophic accidents and casualties [3,4]. However, big data generated from modern industry also affords an unprecedented opportunity to obtain an in-depth understanding of machine condition. Therefore, it is vital to seize this opportunity and advance diagnostic methods for accurate judgment and timely response on machine degradation and failure.

Over the years, a variety of approaches have been developed for machine fault diagnosis through the wisdom and efforts of researchers and engineers. Review them briefly, existing approaches can be roughly divided into four categories according to the development process, i.e. physical model-based methods, signal processing-based methods, machine learning-based methods and their hybrid [5–9]. Physical model-based methods usually require a thorough understanding for mechanisms of the machine, and it is difficult to build accurate physical systems for modern complex mechanical equipment, especially in dynamic and noisy working environment. In addition, most of physical models are inflexible and inefficient since they are unable to be updated with real monitoring data. Different from these approaches, the signal processing-based approach aims to explore advanced signal denosing and filtering technologies to emphasize fault characteristic information. However, it usually requires related equipment knowledge for feature frequency calculation, moreover, the solid fault representation theory and mathematical basis are also the

* Corresponding author.
E-mail addresses: jjy2015@stu.xjtu.edu.cn (J. Jiao), zhaomingxjtu@mail.xjtu.edu.cn (M. Zhao), linjing@buaa.edu.cn (J. Lin), lkxcarrot@stu.xjtu.edu.cn (K. Liang).

https://doi.org/10.1016/j.neucom.2020.07.088
0925-2312/

premise of this method. Another family named machine learning-based method, as a typical representative of data-driven approaches, has been active and brilliant with the development industrial modern industry. Although classical machine learning models, such as support vector machine (SVM) and *k*-nearest neighbor, have achieved remarkable progress over the past years, some drawbacks still exist when facing the higher industrial requirements [10]. For example, i) These methods generally require manual extraction and selection of features, which is limited in complex big data analysis. In addition, it is also difficult to effectively mine high-dimensional features due to the shallow structure; ii) Feature mining and decision making are separately designed, in which the unsynchronized optimization will consume considerable time and restrict the performance; iii) As the diversities of sensor and complexity of machine, data dimension and dynamics increase, it is difficult to obtain satisfactory diagnostic results using traditional algorithms.

Deep learning, as the hottest branch of machine learning, has been witnessed the proliferation and prosperity in various fields, including image identification, speech processing and so on [11,12]. This is not only due to some subjective factors, such as powerful data processing and feature learning capabilities as well as architecture innovation, but also several external factors cannot be ignored, such as explosive increase of industrial big data, breakthrough of hardware, stimulation from multifarious competitive task requirements. Naturally, deep learning has also raised the wave of intelligent fault diagnosis over the past five years. The popular deep learning based diagnostic models include deep autoencoder [13], deep belief network [14], recurrent neural network [15], and convolutional neural network (CNN) [16,17]. Among them, convolutional network [18] has become the leading architecture and achieved state of the art performance in many benchmarks [11]. Similarity, fault diagnosis approaches using convolutional network have also developed most rapidly and a lot of research work has been published. Given the popularity of Convolutional Network based Fault Diagnosis (CNFD), a systemic review and summary is necessary to help tease out current work and make prospects for the further research.

The CNFD framework can generally be summarized in three steps as shown in Fig. 1, including data collection, model construction, as well as feature learning and decision making. In the first step, tremendous monitoring data are collected and prepared from the concerned mechanical equipment. Next, convolutional network models are designed and constructed according to task requirements. Finally, the hierarchical and high-dimensional features can be learned for characterizing machinery condition. Meanwhile, the decision, such as fault classification and remaining useful life (RUL) prediction, is carried out based on the extracted features. Several merits can be clearly revealed from this framework, i) It is able to exploit the in-depth and intrinsic characteristics adaptively while alleviate the requirements of human labor as well as expert knowledge; ii) The model can flexibly update itself according to the real-time monitoring data for more practical diagnostic requirements; iii) This diagnostic framework integrates the feature extraction and decision making together and constructs an end-to-end intelligent diagnostic model.

Prior to our work, there are also several excellent review articles in the field of machine fault diagnosis. For instance, Liu et al. [19] summarized five artificial intelligent algorithms for fault diagnosis of rotating machinery. However, their work mainly focused on the traditional machine learning models, and the review on deep learning based methods is insufficient, especially for convolutional neural networks. Zhao et al. [20] presented a work to review several deep learning models and their applications to machine health monitoring. Although the convolutional network has also been described in their work, it was treated equally with other models and the review about CNN was not enough comprehensive. Hoang et al. [21] presented a survey on deep learning based bearing fault diagnosis, in which the literature only referred to bearing fault diagnosis. In addition, the summarization about CNN based methods is also incomplete. To push further along this line and remedy existing research, we follow the flow of intelligent diagnosis, i.e. from data to model then to application, for a systematical review of the convolutional network based fault diagnosis. Furthermore, it is worth noting that transfer learning technologies using CNN as backbone have begun to attract increasing attention since they are able to address more practical diagnostic issues. However, these applications have not been systemically summarized in existing review on CNFD. With these points in mind, this paper intends to review fault diagnosis methods based on convolutional networks more comprehensively, meanwhile, to provide a reference for those who want to understand and promote the development of CNN technologies for machine fault diagnosis.

The rest of this paper is organized as follows. According the flow of the Fig. 1, data collection process and several popular public datasets are described in Second 2. After that, the concept and theory of CNN and its variants are introduced in Section 3. In Section 4, the applications of convolutional network on machine fault diagnosis are comprehensively reviewed. In Section 5, some conclusions are drawn based on above literature. Finally, prospects are summarized in Section 6.

2. Data collection

As shown in Fig. 1, high-quality data are the premise and foundation for successfully training convolutional neural networks. In brief, there are two steps to acquire mechanical data, including sensor selection and layout as well as data sampling and storage. With the development of the sensor technology, various sensors have been applied to mechanical condition monitoring, such as accelerometer, current, built-in encoder [22,23], etc. Based on these sensors, comprehensive monitoring information can be captured for machine fault diagnosis. Among them, vibration analysis has become the most popular monitoring manner and been rapidly developed over past years. Although these vibration-based approaches have made impressive progress, there are still some restrictions for collecting vibration data in practical industry. For instance, vibration data are often plagued by the interference of transmission path and environment noise, thus the signal-to-noise ratio of data is usually low. In addition, vibration data are not sensitive to low frequency response, thus it is not suitable for the condition monitoring of low speed machinery. Furthermore, the vibration sensor cannot even be installed in high temperature, high pressure or closed working environments. However, these drawbacks can be circumvented by using other sensors, for example, infrared imaging can provide a non-contact measurement method and built-in encoder signal has better signal-to-noise ratio and low frequency response. Therefore, it is of significance to comprehensively consider multiple factors, such as equipment type, working environment, monitoring object and operating condition, for selecting well-suited sensors. Next, the layout of sensor is also an important consideration since proper location can perceive much more health information and reduce the influence of transmission path and interference. Following this step, data sampling can be carried out using the data acquisition system and then data are stored by the hard disk or cloud platform for further analysis and use.

Although the process of data collection is clear and intuitive, there are still difficulties for acquiring high-quality data in real industrial scenarios [24]. For example, i) The fault data are hard to be acquired than health data since the machine is usually not

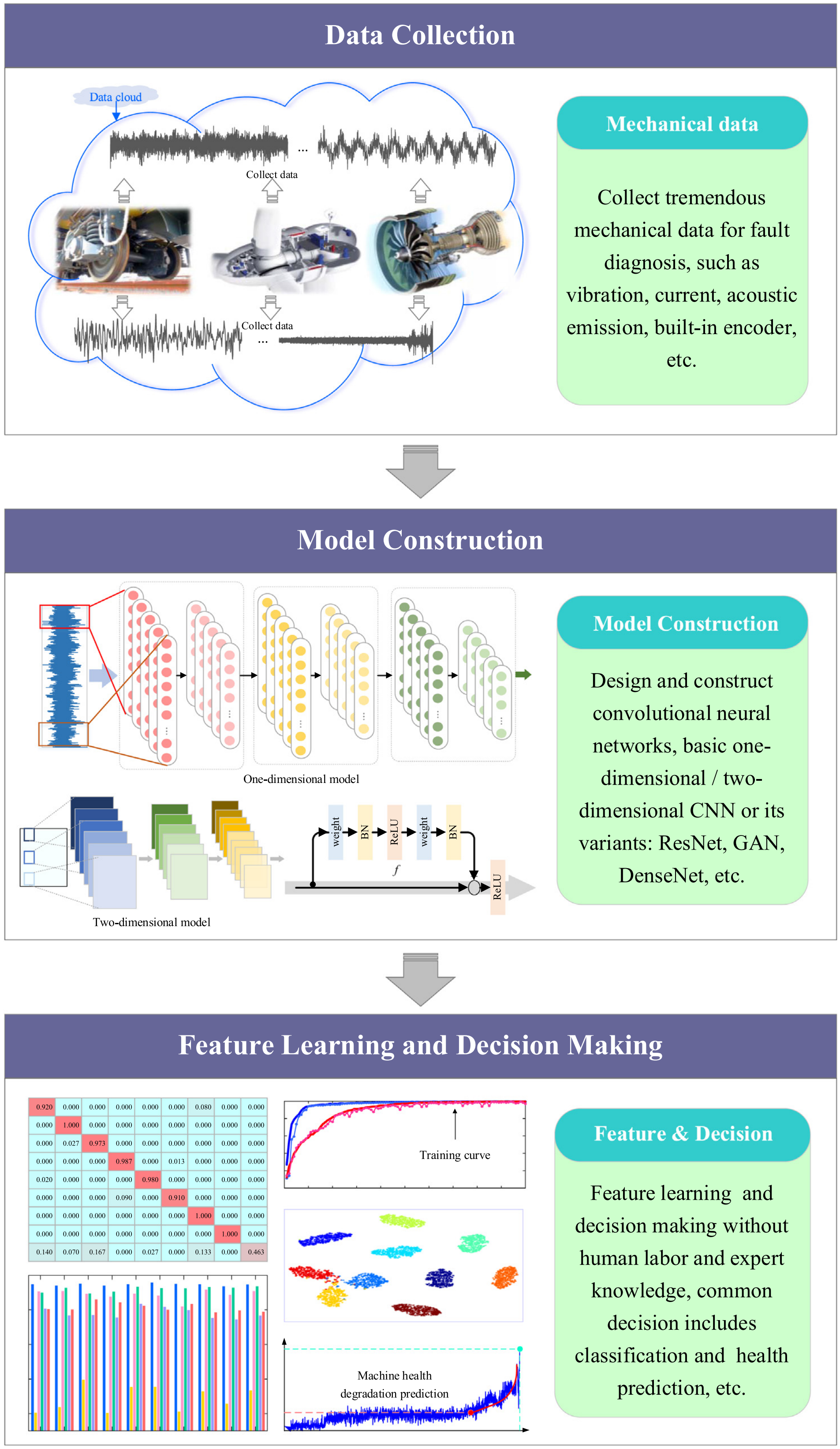


Fig. 1. Framework of general CNFD.

allowed to run under fault condition; ii) The obtaining of life-cycle data is time consuming, expensive and even prohibitive since machine generally has a long running time from the health to failure. Fortunately, a few institutions have published their datasets for public study and application. Therefore, several public datasets are introduced in following subsections, which aims to offer a guideline for researchers and engineers who intend to select these data for performance evaluation of their approaches.

2.1. CWRU bearing fault dataset

2.1.1. Description of the dataset

The Case Western Reserve University (CWRU) [25] bearing dataset has become one of the most popular datasets for machine fault diagnosis since it was made public. The experimental rig is shown as Fig. 2, which is consisted of an electric motor, a torque transducer/encoder, a dynamometer, and control electronics. The single point motor bearing faults simulated by the electro-discharge machining were tested in this platform, including inner race fault, outer race fault and ball fault. Each fault contains different sizes, i.e. 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils (1mil = 0.001 in.). The accelerometers attached to the drive end and fan end of the motor housing were used to collect vibration data with respective sampling frequencies, i.e. 12 kHz and 48 kHz. In addition, this dataset was constructed under four operating conditions, including 0 hp/1797 rpm, 1 hp/1772 rpm, 2 hp/1750 rpm, and 3 hp/1730 rpm.

2.1.2. Characteristics of the dataset

1) CWRU dataset contains multiple bearing fault types, therefore, this dataset has been widely applied to fault identification tasks.

2) The data come from different sensor positions, i.e. the drive end and fan end. Hence, it can be used to study the diagnostic scenario across sensors at different positions.

3) There are different damage sizes for each fault type in this dataset. Thus this dataset can be applied to the diagnostic scenario where the model is respectively trained and tested using data with different fault sizes.

4) This dataset includes four different working conditions and is suitable for the transfer diagnosis research of different operating conditions.

2.2. PHM 09 gearbox fault dataset

2.2.1. Description of the dataset

This database was shared by the IEEE International Conference on prognostics and health management (PHM) 2009 [26]. The schematic of the experimental rig is shown in Fig. 3, where the test industrial gearbox contains three shafts, four gears and six bearings. Two types of gears, i.e. spur gear and spiral cut (helical) gear, were used for experimental test. The spur gear dataset contains eight health conditions while helical gear dataset contains six health conditions as described in Table 1 and 2. In these experiments, the vibration data were sampled with a sampling frequency of 66.67 kHz by two accelerometers mounted on both the input and output shaft retaining plates. Meanwhile, the tachometer signals were collected by 10 pulse per revolution. Therefore, each data file contains three columns, in which the first two columns are vibration signals and the third column is tachometer signal. Moreover, the experimental operating conditions involve five speeds (30 Hz, 35 Hz, 40 Hz, 45 Hz and 50 Hz) and two loads (high and low).

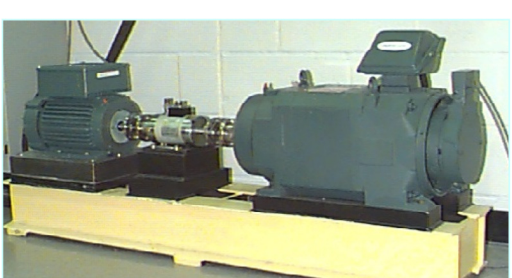


Fig. 2. Experimental platform of CWRU.

2.2.2. Characteristics of the dataset

1) Eight spur gears and six helical gears with different health conditions are used to build this dataset, thus it is suitable for the multi-category classification research.
2) This dataset is built on different working conditions, thus it can be applied to the transfer diagnosis research across different speeds and loads.
3) Two accelerometers were used to synchronously sample vibration data of different positions. Therefore, this dataset is applicable for the information fusion research of double-sensor or transfer diagnosis research between sensors.
4) There are multiple hybrid faults of gears, bearings and shafts in this database, thus it is a typical case for studying hybrid fault diagnosis.

2.3. Paderborn dataset

2.3.1. Description of the dataset

The Paderborn dataset [27] was obtained by using the test bench shown in Fig. 4. This rig is composed of an electric motor, a torque measurement shaft, a rolling bearing test module, a flywheel and a load motor. The bearings of different states were installed into the bearing test module to acquire experimental data. In total, experiments with 26 faulty bearings and 6 healthy bearings were performed, in which the fault contains 12 artificial damages shown in Table 3 and Fig. 5 (a) as well as 14 real damages shown in Table 4 and Fig. 5 (b). The motor current and vibration signals of bearing housing were synchronously measured with a sampling rate of 64 kHz. Moreover, this test rig was operated under four different operating conditions shown in Table 5 by changing the rotational speed, load torque and radial force.

2.3.2. Characteristics of the dataset

1) This dataset contains various different damage states and can be used for the multi-category identification research. Besides, this dataset is more comprehensive than CWRU dataset since it takes into account the artificial and realistic bearing damages simultaneously.
2) In this dataset, the motor current and vibration signals were synchronously sampled. Therefore, on one hand, the motor current signal or the vibration signal can be independently used for bearing diagnosis and comparison study; on the other hand, multi-sensor information fusion based diagnosis can be studied using this dataset.
3) This dataset can be used for transfer diagnosis according to different fault formation mode, that is, the model is trained on the artificial fault data and tested on the real fault data.
4) As this dataset is collected under four different operating conditions, it is suitable for the cross-domain fault diagnosis.

2.4. IMS bearing dataset

2.4.1. Description of the dataset

This bearing dataset was established by the center for Intelligent Maintenance Systems (IMS) of University of Cincinnati [28]. The test rig is presented in Fig. 6, in which four Rexnord ZA-2115 double row bearings were installed on the shaft for testing. The rotation speed was kept constant at 2000 RPM by an AC motor

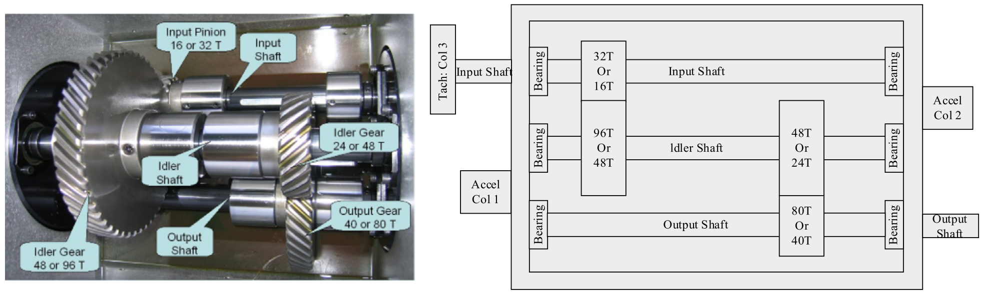


Fig. 3. Experimental rig of PHM 2009.

Table 1
Health conditions of spur gear in PHM 2009 dataset.

	Gear				Bearing						Shaft	
	32T	96T	48T	80T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
Spur 1	G	G	G	G	G	G	G	G	G	G	G	G
Spur 2	C	G	E	G	G	G	G	G	G	G	G	G
Spur 3	G	G	E	G	G	G	G	G	G	G	G	G
Spur 4	G	G	E	Br	B	G	G	G	G	G	G	G
Spur 5	C	G	E	Br	In	B	O	G	G	G	G	G
Spur 6	G	G	G	Br	In	B	O	G	G	G	Im	G
Spur 7	G	G	G	G	In	G	G	G	G	G	G	KS
Spur 8	G	G	G	G	G	B	O	G	G	G	Im	G

IS: Input Shaft; ID: Idler Shaft; OS: Output Shaft; IS: Input Side; OS: Output Side; G: Good; C: Chipped; E: Eccentric; Br: Broken; B: Ball; In: Inner; O: Outer; Im: Imbalance; KS: Keyway Sheared.

Table 2
Health conditions of helical gear in PHM 2009 dataset.

	Gear				Bearing						Shaft	
	16T	48T	24T	40T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
Hel 1	G	G	G	G	G	G	G	G	G	G	G	G
Hel 2	G	G	C	G	G	G	G	G	G	G	G	G
Hel 3	G	G	Br	G	G	G	G	Co	In	G	BS	G
Hel 4	G	G	G	G	G	G	G	Co	B	G	Im	G
Hel 5	G	G	Br	G	G	G	G	G	In	G	G	G
Hel 6	G	G	G	Br	In	B	O	G	G	G	BS	G

Hel: Helical; Co: Combination; BS: Bent Shaft.

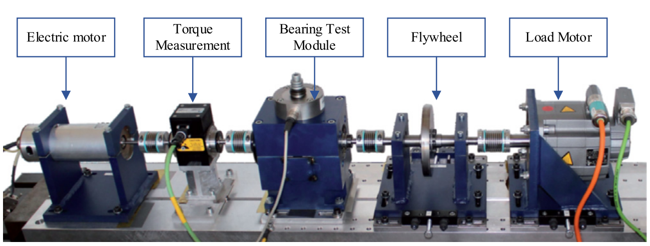


Fig. 4. Paderborn test rig for condition monitoring of rolling bearings.

coupled to the shaft via rub belts. Besides, a radial load of 6000 LBS was applied to the shaft and bearing by a spring mechanism. The vibration data were acquired by accelerometers attached on the bearing housings with a sampling rate of 20 kHz. In total, there are three experiments and each describes a test-to-failure task shown in Table 6. The inner race defect occurred in bearing 3 and roller element defect in bearing 4 at the end of the first experiment. In the second and third experiment, the outer race failure finally occurred in bearing 1 and bearing 3, respectively.

2.4.2. Characteristics of the dataset

1) The bearings of this dataset have four different health conditions, i.e. health, roller fault, outer race fault and inner race fault, which thus can be used to study the issue of fault classification.

Table 3
Description of artificial damages.

	Artificial Damage											
BC	OR	OR	OR	OR	OR	OR	OR	IR	IR	IR	IR	IR
ED	1	2	1	2	1	2	2	1	1	1	2	2
DM	EDM	EE	EE	EE	D	D	D	EDM	EE	EE	EE	EE

BE: Bearing Component; ED: Extent of Damage; DM: Damage Method; OR: Outer Ring; IR: Inner Ring; EDM: Electric Discharge Machining; D: Drilling; EE: Manual Electric Engraving.

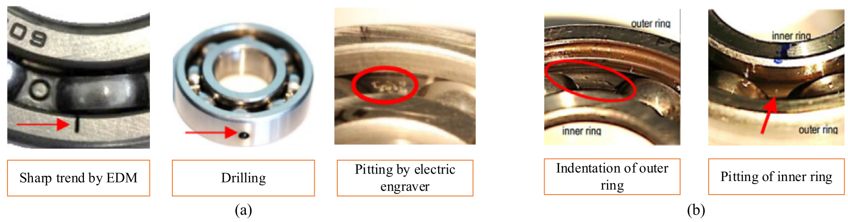


Fig. 5. Several examples of bearing damages. (a) Artificial; (b) real.

Table 4
Description of real damages.

	Real Damage													
D	FP	PI	FP	FP	PI	FP	FP	PI	FP	FP	FP	FP	FP	FP
BC	OR	OR	OR	OR	OR	IR(+OR)	IR(+OR)	IR + OR	IR	IR	IR	IR	IR	IR
Co	S	S	R	S	R	M	M	M	M	M	S	R	S	S
A	nr	nr	r	nr	r	r	nr	r	nr	nr	nr	r	nr	nr
ED	1	1	2	1	1	2	3	1	1	1	3	1	2	1
CD	SP	SP	SP	SP	Di	SP	Di	Di	SP	SP	SP	SP	SP	SP

D: Damage (main mode and symptom); BC: Bearing Component; Co: Combination; A: Arrangement; ED: Extent of Damage; CD: Characteristic of damage; FP: fatigue: pitting; PI: Plastic deformation: Indentations; OR: Outer Ring; IR: Inner Ring; S: Single Damage; R: Repetitive Damage; M: Multiple Damage; nr: no repetition; r: random; SP: Single Point; Di: Distributed.

Table 5
Four operating conditions.

No.	0	1	2	3
Rotational Speed (rpm)	1500	900	1500	1500
Load Torque (Nm)	0.7	0.7	0.1	0.7
Radial Force (N)	1000	1000	1000	400

2) Each data describes a run-to-failure experiment, thus researchers can employ this dataset to study the bearing RUL prediction.
3) The bearings of this experiment experienced an "increase–decrease-increase" degradation trend, in which the reason of "decrease" is the self-healing nature of the damage. As a result, selecting data during this period will increase the difficulty of fault diagnosis.
4) There is only one operating condition in this dataset, which limits the diversity of the data. In addition, the lifetime of each unit has distinct discrepancies, which increases the difficulty of RUL prediction.

2.5. C-MAPSS dataset

2.5.1. Description of the dataset

C-MAPSS dataset [29] was provided by Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) to simulate the turbofan engine degradation. The diagram of simulated engine is shown in Fig. 7. This system has 14 inputs, i.e. the fuel flow and a set of 13 health parameters, which allows the user to simulate the effects of faults and deterioration in any one of the engine's five rotating components. There are 58 outputs including various sensor responses and operability margins, in which a total of 21 sensor variables are used to measure the health states of the engine. In total, five subsets are included in this dataset as listed in Table 7 and each trajectory has a specific initial wear level and degradation process.

2.5.2. Characteristics of the dataset

1) There are sufficient training samples in this dataset, which thus is suitable to train the deep convolutional network models for prediction research.
2) This dataset contains 21 different observation features, such as the temperature, pressure, speed, etc., which means that this dataset can be applied to multi-sensor information fusion research.
3) There are different operating conditions for the same fault mode. Thus, it can be used to evaluate the generalization capability of model.
4) It is worth noting that this dataset was generated from the simulation software, therefore, it has a certain difference with the experimental data or realistic industrial data.

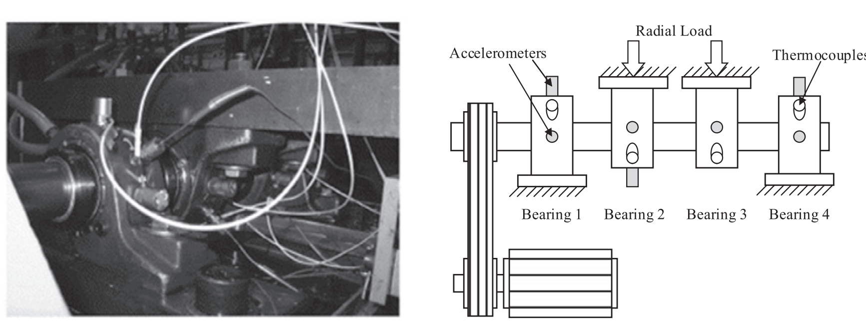


Fig. 6. IMS bearing test rig.

Table 6
Description of bearing conditions in three experiments.

Experiment	Bearing 1	Bearing 2	Bearing 3	Bearing 4
1	UD	UD	IRD	RED
2	ORD	UD	UD	UD
3	UD	UD	ORD	UD

UD: undamaged; IRD: inner race damage; RED: roller element damage; ORD: outer race damage.

2.6. PHM 10 CNC milling machine cutters dataset

2.6.1. Description of the dataset

This dataset was shared in the 2010 PHM Society Conference Data Challenge [30], which focused on RUL estimation of a high-speed CNC milling machine cutters. The experimental rig is shown in Fig. 8 and the operation parameters are listed in Table 8. A Kistler quartz 3-component platform dynamometer was mounted between the work piece and machining table to measure the cutting forces, three Kislter accelerometers were mounted on the work piece to measure the machine tool vibrations of cutting process in X, Y, Z direction, respectively, and a Kistler acoustic emission sensor was mounted on the workpiece to monitor the high frequency stress wave generated by the cutting process.

2.6.2. Characteristics of the dataset

1) It is worth noting that this dataset is conducted under the dry milling environment and has certain differences with the real milling process. However, real milling data are quite difficult to be acquired due to the cost or commercial competition. Therefore, this dataset is still a good choice for the study of RUL prediction of milling machine cutters.
2) Each data file is composed of three-dimensional cutting force data, three-dimensional vibration data and acoustic emission signal. Therefore, it can be used to explore single-sensor or multi-sensor fusion based prediction scenarios.
3) Although this dataset has six cutter data, there are only three cutters are labeled. Thus the amount of data may be insufficient for building complex deep diagnostic networks.
4) This experiment was only performed under one milling operating condition, which limits the diversity of data and restricts the construction of cross prediction scenarios among different conditions.

2.7. FEMTO dataset

2.7.1. Description of the dataset

FEMTO dataset [31] was acquired from the PRONOSTIA experimental platform designed by Franche-Comté Electronics, Mechanics, Thermal Processing, Optics-Sciences and Technologies institute (FEMTO) and aimed to provide the experimental data of characterizing the degradation of bearings. This dataset was also used for the prognostic challenge in the IEEE International Conference on PHM 2012. The overview of the test rig is presented in Fig. 9, which is composed of a rotating part (the asynchronous motor with a gearbox and two shafts), a degradation generation part (with a radial force applied on the tested bearing) and a measurement part (sensors). All bearings are healthy and not seeded with any defects at the beginning of the test. Two types of sensor, i.e. thermocouple and accelerometers (horizontal and vertical) were used to collect the temperatures and vibration signals of the testing bearing, the sampling frequencies of vibration and temperature were set to 25.6 kHz and 10 Hz, respectively. The bearing life is considered terminated when the amplitude of the vibration signal exceeds 20 g, which aims to avoid propagation of damages to the whole test bed. Totally, this dataset contains 17 start-to-end data of bearings under three operation conditions.

2.7.2. Characteristics of the dataset

1) This is a full-life bearing dataset with real damage, thus it is becoming a popular choice for RUL prediction research.
2) Although two-direction vibration signals are collected, the vertical vibration signals provide less useful information than the horizontal ones for tracking the bearing degradations according to the related literature [32,33].
3) This dataset provides a natural degradation process since the bearings are healthy and not seeded with any defects at the beginning of the tests. But this dataset presents no prior information about the properties of the damages.
4) The train set is small while the spread of the life duration of all bearing is wide, which means that the RUL estimation is more challenging using this dataset. In addition, the degradation and fault patterns are discriminative for distinct bearings even under the same operating condition due to various factors, which thus increases the difficulty of RUL prediction.

2.8. Epilog

In this section, we simply introduced the data collection process from the sensor selection and layout to data sampling and storage. After that, we summarize seven popular public datasets to provide more convenience for readers, in which the overview and characteristics of each dataset are included. Furthermore, a concise summary is listed in Table 9, where the second column shows the main monitoring object of each dataset; the third column lists the data type of each dataset; the last three columns illustrated and compared the application scenarios, including classification, prediction, and transfer diagnosis. Furthermore, it is noteworthy that the datasets introduced in this work are just several popular ones and there are still some other available public datasets for the use and reference, such as MFPT fault dataset [34], XJTU-SY bearing dataset [35], University of Connecticut gear fault dataset [36]. In addition, the international conferences on PHM from the PHM society or the IEEE reliability society often provide some valuable datasets for researchers and engineers.

3. Convolutional neural network and its variants

Convolutional neural network, as a feedforward deep learning model, has become a milestone technique and achieved state-of-the-art performance in various computer vision and pattern recognition tasks. In recent five years, convolutional neural networks have also been successfully introduced to the field of machine fault diagnosis. From the perspective of completeness, the basic theory

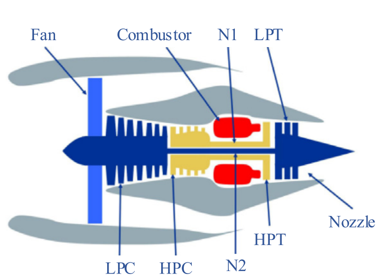


Fig. 7. Diagram of the simulated engine.

Table 7
Five subsets included in the turbofan engine degradation dataset.

No.	Train trajectories	Test trajectories	Conditions	Fault modes
1	218	218	/	/
2	100	100	One (sea level)	HPC Degradation
3	260	259	Six	HPC Degradation
4	100	100	One (sea level)	HPC and Fan Degradation
5	248	249	Six	HPC and Fan Degradation

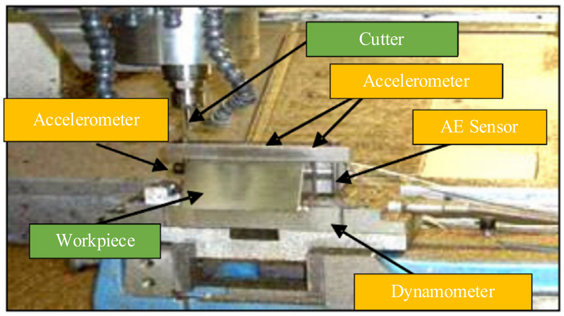


Fig. 8. High-speed CNC milling machine.

Table 8
Description of operation parameters.

Description	Value
Running speed of the spindle	10400 rpm
Feed rate in the $\times$ direction	1555 mm/min
Depth of cut in the *y* direction	0.125 mm
Depth of cut in the z direction	0.2 mm
Sampling frequency	50 kHz

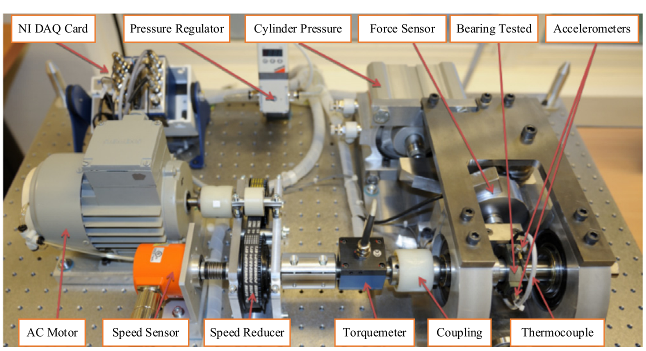


Fig. 9. PRONOSTIA experimental platform.

on CNN is firstly presented before reviewing its applications, which aims to provide the understanding and preparation for the researchers, engineers and even beginners who intend to apply convolutional networks for fault diagnosis.

3.1. The basic convolutional neural network

Take one-dimensional mechanical signal as an example, a simple CNN is displayed in Fig. 10, which contains one input layer, multiple convolution-pool layers, several fully-connected layers and one output layer. Moreover, two popular operations, including batch normalization and dropout, are also embedded in this architecture, which aim to improve the model performance. In following subsections, each operation will be introduced respectively.

3.1.1. Convolution

The convolution is a pivotal step for feature extraction in CNN, which endows two attractive advantages: 1) Sparse connection. Each convolutional kernel is connected with the local patch of previous features and can efficiently reduce the number of parameters and speed up training; 2) Weight sharing. The weights of each kernel remain unchanged when it slides on feature maps, which further reduces parameters. In general, multiple kernels are contained in one convolutional layer, which aims to learn comprehensive feature representations. Mathematically, let $\mathbf{x} \in \mathbb{R}^d$ be the *d*-dimensional mechanical data, the *j*-th feature map can be described as:

$$\mathbf{c}^j = \mathbf{x} * \mathbf{w}^j + \mathbf{b}^j \tag{1}$$

where $\mathbf{w}^j \in \mathbb{R}^h$ represents the *j*-th filter, it is used to code the input x and generate the *j*-th feature map $\mathbf{c}^j = [c_1^j, c_2^j, \ldots, c_{d-h+1}^j]$; b^j denotes the bias term. In addition, stride and padding respectively control the convolution density and output size. The intuitive schematic of convolution is displayed in Fig. 11.

In addition to vanilla convolution, some advanced convolutions have also been developed in recent years for a variety of purposes. To expand the receptive field without loss of resolution or coverage, dilated convolution was developed in [37]. As shown in Fig. 12, a general 3×3 convolutional kernel is shown in Fig. 12 (a) and a 2-dilated 3×3 kernel and 4-dilated 3×3 kernel are displayed in Fig. 12 (b) and (c). It is worth noting that although the receptive field is enlarged, the number of parameters is identical.

To cut down the amount of calculation, separable convolution [38] has been creatively developed, which includes spatially separable convolution and depthwise separable convolution. From Fig. 13 (a) to (b), it can be seen that spatially separable convolution transforms original 3×3 convolution kernel to a 3×1 and a 1×3 kernels. The depthwise separable convolution is carried out by a depthwise convolution and a pointwise convolution as shown in Fig. 14. This convolution may be promising to develop real-time diagnosis by reducing computation.

Furthermore, there are also other awesome convolutions, such as group convolution [18] and attention-convolution [39]. To put it simply, group convolution can not only improve the training efficiency of the model, but also achieve more impressive performance than common convolution since sparse relationship among filters leads to the regularization effects for model. Attention-convolution aims to obtain better capability of feature learning by enhancing fault-related features and ignoring irrelevant features.

In addition to the innovation of convolutional mode, kernel size is also a key factor for affecting feature learning. It may not be appropriate for all convolutional layers to be set to the kernels with fixed small-size, such as 3×3, especially in one-dimensional network based diagnostic framework. Through reviewing and summarizing relevant literature, the larger kernel in initial layer is likely to mine more comprehensive mechanical fault information since large size has better characterization ability of frequency domain [40–42]. Then, small kernels are followed for achieving deeper

Table 9
The summary of different datasets.

Name	Monitoring object	Data type	Classification	Prediction	Transfer diagnosis
CWRU	Motor bearing	Multi-vibration	√	×	√
PHM 09	Gearbox, bearing, shaft	Multi-vibration	√	×	√
Paderborn	Bearing	Current-vibration	√	×	√
IMS	Bearing	Vibration	√	√	×
C-MAPSS	Turbofan engine	21 sensor data	×	√	√
PHM 10	Milling cutter	Current-vibration-AE	×	√	×
FEMTO	Bearing	Temperature-vibration	×	√	√

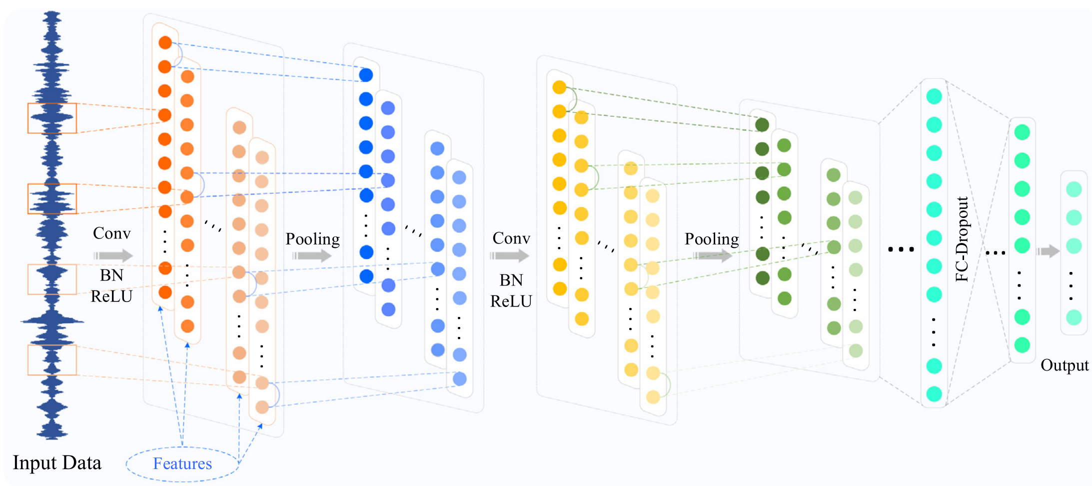


Fig. 10. An example of convolutional neural network. Conv: Convolution; BN: Batch Normalization; ReLU: Rectified Linear Unit activation function; FC: Fully-connected layer.

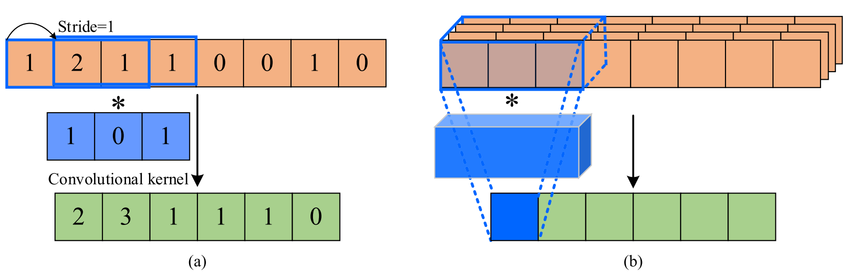


Fig. 11. Schematic of one-dimensional convolution. (a) Single-channel input; (b) Multi-channel input.

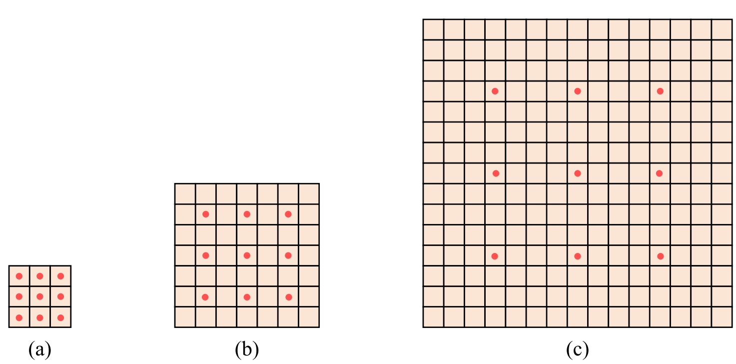


Fig. 12. Schematic of dilated convolution. (a) 3 × 3 kernel with a receptive field of 3 × 3; (b) 2-dilated 3 × 3 kernel with a receptive field of 7 × 7; (c) 4-dilated 3 × 3 kernel with a receptive field of 15 × 15.

network architecture and finer feature learning. Therefore, empirical evidence gives the rule of thumb that the kernel size can be designed from large to small in constructing deep convolutional networks for better fault feature extraction.

3.1.2. Activation function

After the convolutional operation, the non-linear activation function, such as sigmoid function, tanh function, and rectified linear unit (ReLU) [43], is usually employed to express complex fea-

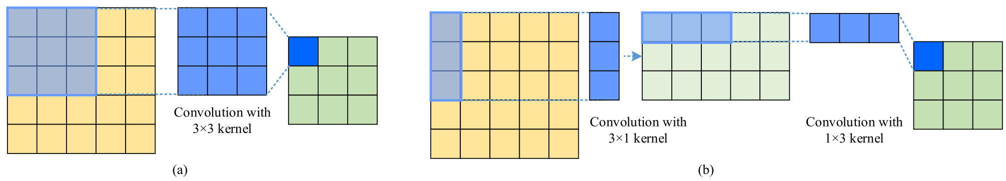


Fig. 13. Schematic of spatially separable convolution. (a) General convolution; (b) spatially separable convolution.

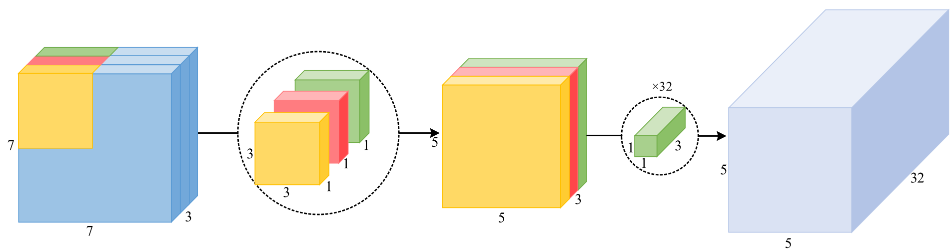


Fig. 14. Schematic of depthwise separable convolution.

tures. Compared to sigmoid and tanh functions, ReLU is widely used in CNN since it can not only speed up learning, but also alleviate the issue of gradient vanishing. However, a potential disadvantage of ReLU unit is that it has zero gradient whenever the unit is not active. This may cause units that do not active initially never active as the gradient-based optimization will not adjust their weights. To address the shortcoming of ReLU and more advanced activation functions have been proposed as shown in Fig. 15, such as leaky ReLU (LReLU) [44], parametric ReLU (PReLU) [45] and exponential linear unit (ELU) [46].

3.1.3. Pooling

The pooling operation (i.e. down-sampling) aims to obviate redundancy and enable features more robust after convolution. The common pooling operations include the max pooling and average pooling while the difference between them is whether to take the maximum or average value in the pooling region [47]. Take the max pooling as an example, the mathematical description is given as follows:

$$po_k^j = \max\{c^j_{k:k+r-1}\} \tag{2}$$

where c^j represents input and r is the pool size; po_k^j denotes the maximum value in the corresponding pooling region.

3.1.4. Batch normalization

The batch normalization (BN) [48] has become a popular technique to alleviate internal covariance shift and promote network training. Mathematically, given the d-dimensional feature $\mathbf{x} = \{x^{(1)}, \ldots, x^{(d)}\}$, the operation of BN is described as follows:

$$\tilde{x}^{(k)} = \frac{x^{(k)} - \mathbb{E}[x^{(k)}]}{\sqrt{\mathrm{Var}[x^{(k)}]}},\ h^{(k)} = \gamma^{(k)}\tilde{x}^{(k)} + \beta^{(k)} \tag{3}$$

where $x^{(k)}$ and $h^{(k)}$ represent the k-th activation of input and output, respectively; $\mathbb{E}[\cdot]$ and $\mathrm{Var}[\cdot]$ denote the expectation and variance; $\gamma^{(k)}$ and $\beta^{(k)}$ stand for the parameters to be learned.

3.1.5. Dropout

Dropout [49] is a technique that prevents overfitting and provides a way of approximately combining different networks. The key operation of dropout is to randomly drop neuron units (along with their connections) of the network during training as shown in Fig. 16. Specifically, a unit is present with probability$p$at training time, while the unit is always present and the weights are multiplied by$p$at test time.

3.1.6. Fully-connected layer

After stacking multiple convolution-pool modules, the fully-connected layers are usually employed to process features further. The mathematical calculation of the fully-connected layer is the same as the traditional perception, it can be described as:

$$\mathbf{fc}^l = \sigma(\mathbf{w}^l \cdot \mathbf{fc}^{l-1} + \mathbf{b}^l) \tag{4}$$

where $\mathbf{fc}^l$ represents the output features of l-th fully-connected layer; $\mathbf{w}^l$ and $\mathbf{b}^l$ stand for the connection weight and bias, respectively; $\sigma(\cdot)$ denotes the non-linear activation function.

3.1.7. Decision layer

After the feature extraction, the decision layer is usually followed to get the final results. There are usually two typical outputs in fault diagnosis problem, the one is label prediction and the other is single variable output, such as RUL prediction. Softmax function has become one of the most popular choices for the classification task owing to its effectiveness. Mathematically, Softmax operation can be described as

$$p(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{i=1}^{N_c} e^{z_i}} \tag{5}$$

where $p(\mathbf{z})_i$ represents the probability that $\mathbf{z}$ belongs to category i, $i \in N_c$.

3.1.8. Optimizer

After constructing convolutional network networks, gradient descent methods are usually used to train network to get optimal parameters within acceptable time. For example, the initial approaches include Batch Gradient Descent (BGD), Stochastic Gradient Descent (SGD), and Mini-Batch Gradient Descent (MBGD). In BGD, all training data are calculated to get the gradient in each update, which thus is able to converge to the global optimum of the convex plane or the local optimum of the non-convex plane. As this algorithm need calculate all training data for an update, it is pretty slow and impractical in realistic application. In contrast, SGD only utilizes one sample for each update and thus the time of each update in SGD is greatly less than BGD. However, SGD will cause the severe oscillation of objective function due to frequent

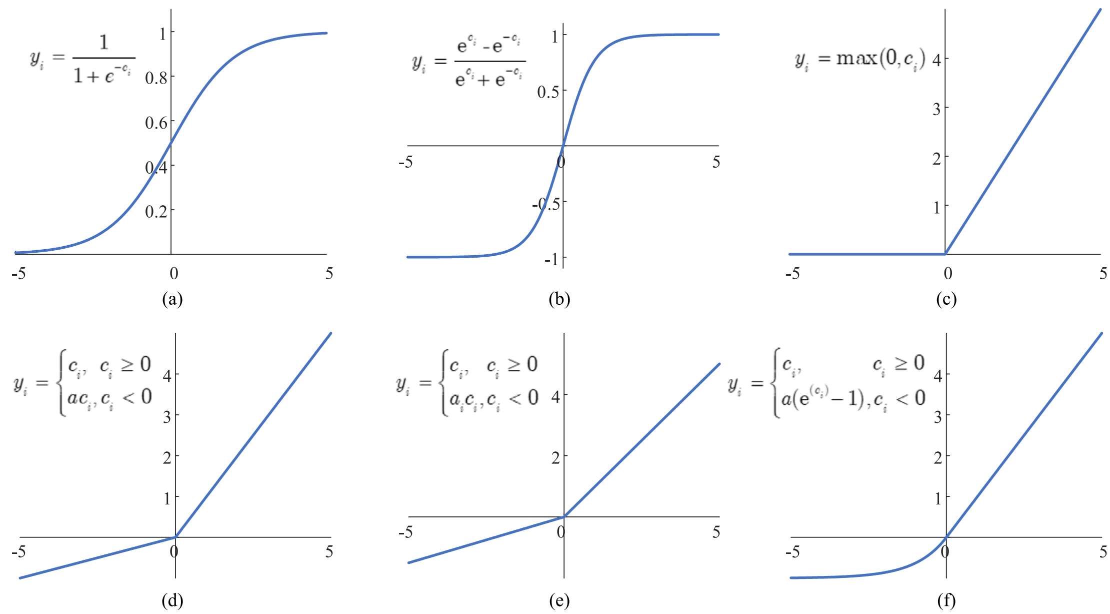


Fig. 15. Nonlinear activation functions, where $c_i$and y_i represent input and output features. (a) sigmoid; (b) tanh; (c) ReLU; (d) LReLU; (e) PReLU; (f) ELU.

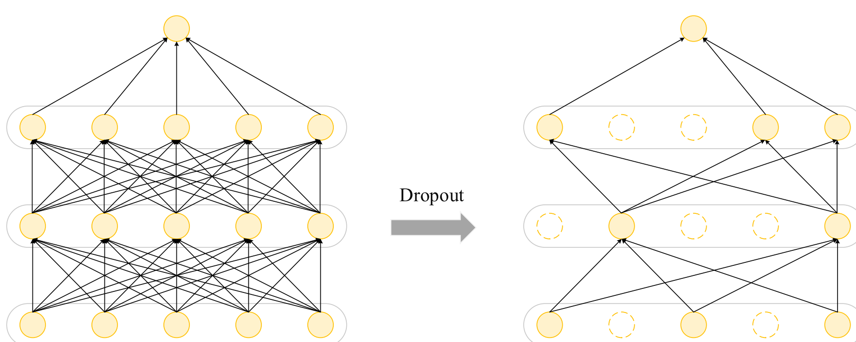


Fig. 16. The example of dropout operation.

updates, while this oscillation might help jump to a better local minimum. MBGD is proposed by integrating the characteristics of BGD and SGD, which uses a small batch of samples for each update, so that it can perform more efficient than BGD and more stable than SGD.

Although BGD has been widely used to train convolutional neural networks, however, existing disadvantages still cannot be ignored, especially with the evolution of deep convolutional neural networks [50]. To name a few, 1) Choosing a proper learning rate can be difficult. A small learning rate will lead to painfully slow convergence while large one will cause loss function oscillate or even deviate at the minimum; 2) The same learning rate to the updates of all parameter. If the features have different frequencies, we might not want to update all of them to the same extent.

To deal with the first problem, the Momentum algorithm [51] is presented to prevent oscillations in one dimension and using the exponentially weighted average of the gradient to achieve faster convergence. However, there is a big jump after the cumulative gradient due to blindly following gradient descent. To address this issue, Nesterov Accelerate Gradient algorithm gives a predictability that makes the Momentum slow down before the slope becomes positive.

In response of second challenge, another family of adaptive optimization algorithms is developed, including Adagrad [52], Adadelta [53], Root Mean Square prop (RMSprop) and Adaptive Moment Estimation (Adam) [54]. Specifically, Adagrad is proposed to adapt the learning rate to parameters, performing smaller updates for frequent feature-related parameters, and performing larger updates for infrequent ones. Nevertheless, the main weakness of Adagrad is that the learning rate will eventually become infinitely small. Therefore, Adadelta is designed to address the monotonically decreasing learning rate. Instead of accumulating all squared gradients, it sets a fixed size window to restrict the number of accumulated squared gradients. At the same time, the sum of gradients is recursively defined as the decaying average of all previous squared gradients, rather than inefficiently storing the previous squared gradients. Similar to Adadelta, RMSprop is another algorithm to solve the problem of the radically diminishing learning rate of Adagrad.

Adam algorithm is designed by integrating the Momentum and the RMSprop. Not only does Adam store the exponential decay average of the past square gradients like the RMSprop, but it also maintains the exponential decay average of the past gradients like the Momentum. Practice has proved that Adam algorithm works well on many problems and is applicable to many different CNN structures.

Although various optimizers have been developed, there is currently no unified guideline for choosing the best optimizer. Through existing literature review, rules of thumb can be summarized as reference. 1) Momentum and Adam algorithms are the two most popular methods in community of machine fault diagnosis and can be prioritized; 2) The performance of optimizers is closely related to data distribution, it is recommendable to try different optimizers when solving practical personalized diagnostic issues; 3) If there is large oscillation or divergence during training, reducing the learning rate may be a good choice.

3.2. Variants and extension of convolutional network

In addition to the original convolutional neural network, the improved variants have also been developed and applied to the field of fault diagnosis for more excellent performance. Therefore, several common variants will be introduced briefly in this subsection, including residual network, densely connected convolutional network, and generative adversarial convolutional network.

3.2.1. Residual network

It is common that blindly stacking deeper convolutional layers in the regular CNN will lead to the performance degradation or gradient vanishing/explosion. To address these problems, an improve convolutional model, named residual network (ResNet) [55], has been proposed and become the typical representative in deep networks owing to noticeable improvements. The ResNet is generally composed of multiple residual learning blocks and each contains the convolutional layers, BN layers and activation layers. A simple example is shown in Fig. 17. From this figure, it can be seen that the output of the residual block is calculated as $\mathbf{y} = f(\mathbf{x}, \mathbf{w}) + \mathbf{x}$, where f denotes the residual mapping to be learned; $\mathbf{w}$ represents the parameters. The operation $f + \mathbf{x}$ is carried out by a shortcut connection of element-wise addition. Note that a projection by shortcut connection is performed to match the input and output when they have different dimensions.

3.2.2. Densely connected convolutional network

In addition to ResNet, another popular deeper structure named densely connected convolutional network (DenseNet) [56] has also attracted increasing attentions owing to its excellent performance. The DenseNet is composed multiple dense connection block, in which the current layer receives the feature maps from all previous layers. A simple example is shown in Fig. 18. More specifically, given the input $\mathbf{x}_0$, the feature calculation of the l-th layer can be described as $\mathbf{x}_l = F_l([\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{l-1}])$, where $[\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{l-1}]$denotes the concatenation of the feature maps generated in layer 0,. . ., l-1; F_l represents a composite function of three consecutive operations, i.e. BN, ReLU, and convolution.

3.2.3. Generative adversarial convolutional network

Generative adversarial convolutional network (GAN) [57] has become an important research hotspot with promising performance on data generation. The GAN usually contains two models, i.e. a generative model G and a discriminative model D, which are pitted against each other to find a Nash equilibrium. As shown in Fig. 19, the G is trained to learn the distribution of real data and generate samples from noise and the D is trained to distinguish the real samples and pseudo samples by enabling the high output probability of real samples and the low probability of generative samples. In other words, the optimization of adversarial learning is a minimax game as follows:

$$\min_G \max_D L(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \tag{6}$$

where $P_{\mathrm{data}}(\mathbf{x})$ represents true data distribution; $P_{\mathbf{z}}(\mathbf{z})$ is the prior noise distribution $\mathbf{z}$.

3.3. Epilog

In this section, the concept on basic CNN and several variants is introduced, which aims to help readers better understand the work mechanism of convolutional neural network. Unfortunately, there is no specific guidelines for architecture selection and design, thus researchers need to design and optimize networks following own task requirements. Furthermore, most advanced network architectures are developed based on image data features, however, engineers are strongly encouraged to explore novel and practical convolutional architectures that fits the characteristic of industrial mechanical data, which is promising for more excellent fault diagnosis.

4. Applications

In this section, existing literature applying CNN to machine fault diagnosis is systematically reviewed and summarized, which covers published journal and conference papers in recent three years. In particular, we introduce the related literature according to the following three aspects: fault classification, health prediction, and transfer diagnosis.

4.1. Applications in fault classification

Fault classification is the earliest and most widely studied subfield in CNFD inspired directly by image classification. In this section, the applications in fault classification are systematically reviewed. To make the narrative more organized, we elaborate these studies depend on the structure characteristics of convolutional network and categorize them from three aspects, i.e. two-dimensional (2D) convolutional network based classification, one-dimensional (1D) convolutional network based classification, and fault classification based on convolutional network variants.

4.1.1. 2D convolutional network based classification

In the beginning, the convolutional networks used for machine fault diagnosis are the original 2D structure by imitating the image processing. Since mechanical signal is a 1D time series in almost all cases, the main idea is to convert the 1D data into the 2D form in this case. Therefore, we firstly refine various signal conversion methods and then summarize the related applications.

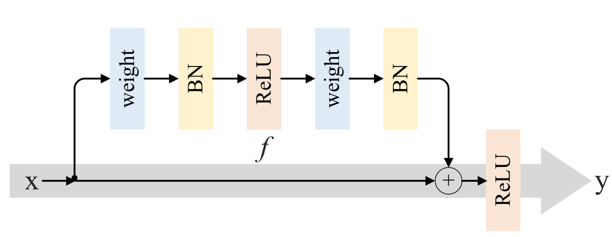


Fig. 17. Residual learning connection, where weight represents convolution; BN stands for batch normalization; and ReLU is rectified linear unit.

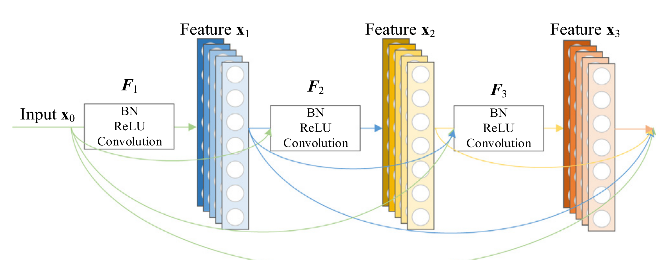


Fig. 18. Dense connection block.

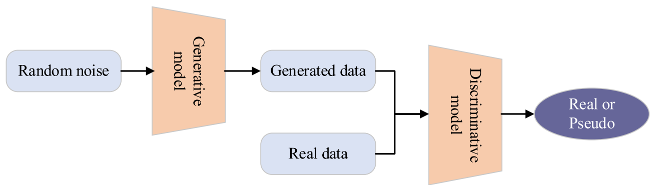


Fig. 19. Framework of GAN.

- Data matrix transformation

Data matrix transformation refers to that researchers directly arranged raw mechanical data into a 2D format as the model input. A simple illustration is shown in Fig. 20. Guo et al. [58] transformed vibration data into matrix 2D and then proposed a hierarchical CNN with adaptive learning rate for fault pattern recognition and fault size evaluation. Wang et al. [59] arranged the raw vibration signal into 2D input and introduced an adaptive CNN for bearing fault diagnosis, in which the particle swarm optimization method was added to determine the main parameters of CNN. Shao et al. [60] proposed a bearing fault diagnosis method based on the deep convolutional belief network and the compressed sensing technique. Similarly, they [61] also utilized the auto-encoder to compress data for fault diagnosis of electric locomotive bearing. Wang et al. [62] constructed mechanical data into Hankel matrix and proposed a CNN based hidden Markov model for bearing fault classification. Gong et al. [63] firstly integrated the temporal and spatial multichannel raw signals to construct the model input. After that, the 2D convolutional network was designed for feature learning and then SVM was used for fault classification. Jing et al. [64] presented an adaptive multi-sensor data fusion based CNN method for planetary gearbox fault diagnosis, which aimed to optimize a combination of different fusion levels to satisfy the requirements of different diagnosis tasks. Chen et al. [65] fused horizontal and vertical direction vibration data into 2D matrix and presented a deep CNN for health state identification of planetary gearboxes. Han et al. [66] presented a diagnostic framework of complex systems by combining the spatiotemporal pattern network with convolutional network, in which the former was used for spatiotemporal feature learning and the latter was used for condition classification. Yang et al. [67] used the hierarchical symbolic analysis to process original signal and then built a three-layer convolutional network for fault diagnosis of rotating machinery. Liu et al. [68] presented a dislocated time series CNN for fault classification, in which a dislocated layer was introduced to constructed 2D input data. Yang et al. [69] transformed multi-source vibration signals to construct the 2D matrix and then proposed a CNN based method for fault diagnosis of reciprocating compressor.

- Image transformation

Image transformation means that researchers try to convert 1D mechanical signals into the image in pixel format shown in Fig. 21. Xia et al. [70] presented a fault diagnosis method for rotating machinery based on multiple sensors fusion and CNN. In this method, raw vibration signals from sensors of different locations were aligned into 2D images as the input. Hoang et al. [71] converted raw vibration signals into vibration images, then a simple two-layer convolutional model was constructed for rolling bearing fault classification. Zhang et al. [72] proposed an equitable sliding stride segmentation approach to expanse data volume. Next a hybrid model based on the convolutional network and bi-gate recurrent unit was constructed for feature learning and classification. Hoang et al. [73] converted the motor current signals into gray images and presented a decision level fusion based CNN for bearing fault identification. Wang et al. [74] used multi-sensor data fusion to construct image data and then designed a four-layer convolutional network for fault classification. Hu et al. [75] used the compressed sensing technology to reduce data size and transform data into image pixel. After that, an improved multi-scale convolutional network was constructed for fault recognition of machinery. Wang et al. [76] converted multi-sensor vibration signals into RGB color images to refine features and enlarge the differences between different types of fault signals. Then, an improved LeNet-5 was designed for fault diagnosis. In [77], raw mechanical signals were transformed into a square matrix through non-overlapping cutting and normalization. Then a modified LeNet-5 was designed for feature learning and fault classification.

- Time or frequency domain transformation

Another type of conversion is to use the statistics of time or frequency domain as input information of convolutional network. Chen et al. [78] calculated statistical measures of vibration signals from the time and frequency domains as the model input and then applied one-layer convolutional network for fault identification of bearings and gears. Janssens et al. [79] utilized the discrete Fourier transform to process the accelerometer signals and presented a simple convolutional network for bearing condition recognition. Bhadane et al. [80] used the statistical features extracted from vibration data as the model input and developed a 2D CNN for bearing fault classification. Lu et al. [81] proposed a convolutional network based health state classification method for rolling bearing. In this method, the time and frequency domain features of vibration data were extracted to build the input matrix. Li et al. [82] used the root mean square maps from the spectrum of two vibration data as the input and presented a CNN with an improved Dempster-Shafer evidence theory for bearing fault diagnosis. Tra et al. [83] utilized the spectral energy maps of the acoustic

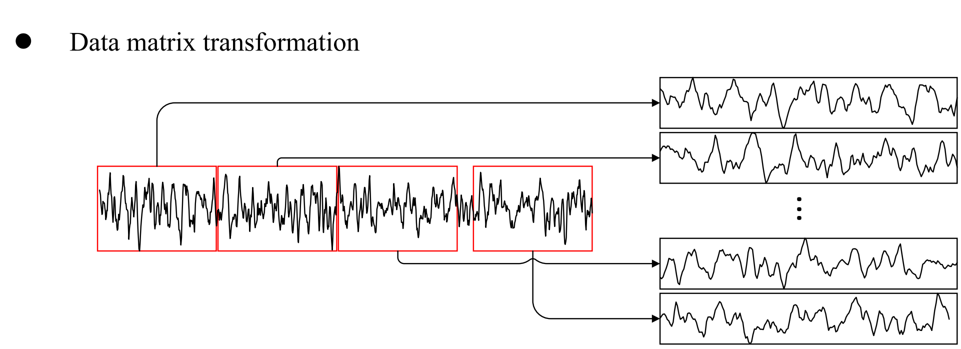


Fig. 20. Simple illustration of data matrix transformation.

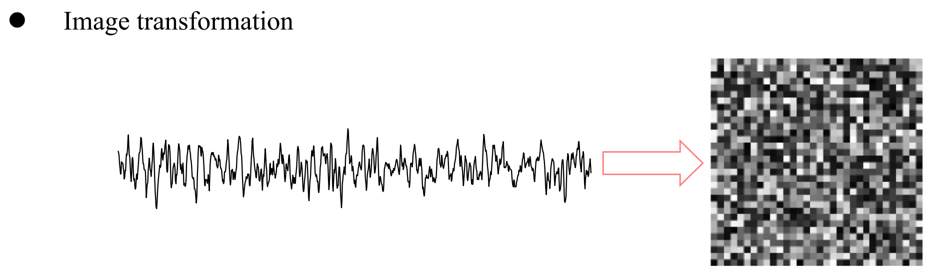


Fig. 21. Simple illustration from mechanical data to image.

emission signals as the model input. Then a CNN with the stochastic diagonal Levenberg-Marquardt algorithm was proposed for incipient bearing fault diagnosis under variable operating speeds. Prosvirin et al. [84] transformed 1D acoustic emission signals into 2D kurtogram images and then utilized the CNN for feature extraction and bearing fault classification. Tian et al. [85] integrated features from time and frequency domains as the input and developed a deep CNN with an immunity algorithm for rolling bearing fault diagnosis. Tra et al. [86] used the energy distribution maps of acoustic emission (AE) spectra to train a convolutional network for fault diagnosis under variable speed conditions. Li et al. [87] constructed feature images from multi central frequencies as well as vibration frequency spectrum and then used CNN to process these images for gear fault identification. Yao et al. [88] integrated features of time and frequency domains from multi-channel acoustic signals as input data and established a convolutional network for gear fault diagnosis. Kien et al. [89] visualized the spectrums of vibration signals as grayscale images and then proposed a deep convolutional network to process these images for crack detection of gears.

- Wavelet transform

Wavelet transform (WT) preprocessing method is to convert mechanical time series into 2D time–frequency (TF) representation as the network input. Ding et al. [90] used the wavelet packet energy image as the input of deep convolutional network and presented an energy-fluctuated multiscale feature mining approach for spindle bearing fault diagnosis. Gao et al. [91] used the complex Morlet wavelet to acquire the 2D TF maps from vibration signals. Then CNN was designed for rolling bearing fault diagnosis. Guo et al. [92] employed the continuous WT to decompose vibration signals into scalogram according to the rotating speed. Then a Pythagorean spatial pyramid pooling based convolutional network was presented for bearing fault diagnosis. Xu et al. [93] utilized WT to convert vibration signals into 2D grayscale images. Then LeNet-5 was built to learn multi-level features and the random forest classifiers were used for bearing fault classification. Islam et al. [94] employed discrete wavelet packet transform to process the AE signals and proposed a convolutional diagnostic model for bearing fault classification. Sun et al. [95] used dual-tree complex WT to acquire the multiscale features to train the CNN for gear fault recognition. Cabrera et al. [96] proposed a convolutional diagnostic network pre-trained by the stacked convolutional auto-encoder for fault severity assessment of helical gearbox, in which the TF features acquired by WT were used as the model input. Han et al. [97] proposed a dynamic ensemble CNN for gear fault diagnosis. In their method, wavelet packet transform was employed to construct multi-level wavelet coefficients matrices for representing the nonstationary vibration signals. Then multiple paralleled CNNs with shared parameters and a dynamic ensemble layer were designed for feature extraction and fault classification. Grezmak et al. [98] presented an explainable deep CNN with layer-wise relevance propagation for gearbox fault diagnosis, in which the TF images from continuous WT were used as the model input. Liang et al. [99] adopted WT to extract TF information from vibration signals to train a CNN for compound fault diagnosis of gearbox. Guo et al. [100] used continuous WT to convert the original mechanical signals as the input of the convolutional network for rotor fault diagnosis. Shao et al. [101] converted the vibration and current signals into the time–frequency representation by WT. Then a deep CNN was designed to predict induction motor conditions. Hsueh et al. [102] used empirical WT to process the current signals into grayscale images and trained a convolutional network for induction motor fault classification. Chen et al. [103] employed continuous WT to process raw vibration signals. Then a convolutional model with a square pooling architecture was designed for feature extraction. Finally, extreme learning machine was used for fault classification. Cao et al. [104] firstly adopted dual tree wavelet to decompose the machine spindle vibration signals. Then the reconstructed sub-signal sequences from different scales and their Hilbert envelope demodulation spectra were stacked to train CNN for tool wear state identification.

- Short Time Fourier Transform

Analogous to wavelet transform, Short Time Fourier Transform (STFT), as another common TF analysis approach, has also been used for data preprocessing in CNFD. Verstraete et al. [105] utilized STFT to generate image representations of raw vibration signals and then constructed a CNN for bearing classification. Pandhare et al. [106] employed the time–frequency features obtained by STFT to train convolutional network model for bearing fault diagnosis. Xin et al. [107] used STFT to calculate TF features and then employed sparse auto-encoder and convolutional network for feature extraction. Finally, Softmax classifier was used to obtain the classification results. Yu et al. [108] transformed the phonetic signals into spectrograms using STFT and then used VGG 16 for wind turbine fault diagnosis. In [109], Wen el al. presented a snapshot ensemble convolutional network for fault diagnosis of pump and bearing. This method can find the proper range of learning rate when facing a new dataset. In [110], Wang et al. proposed a CNN based motor fault diagnosis method, in which STFT was used to pretreat raw signals to acquire TF maps.

- Other preprocessing technologies

Li et al. [111] used S-transform to process original data into the TF coefficients matrix. Then CNN was designed for feature learning and fault classification. Wen et al. [112] proposed a convolutional network based two-level hierarchical diagnosis network, in which S-transform was used to preprocess data. Jeong et al. [113] employed shaft orbit shape images as monitoring information to train the CNN for fault diagnosis. Waziralilah et al. [114] used the Gabor transform to process the raw vibration signal and then presented a CNN for bearing fault diagnosis. Zhao et al. [115] used

Hilbert transform and synchrosqueezing transform to calculate the TF representations of the vibration signals. Then these features were used to train a convolutional model for bearing fault classification. Janssens et al. [116] employed VGG to process infrared thermal video for condition detection of the machine. Jia et al. [117] presented a convolutional network based fault detection model by processing infrared thermography images. Li et al. [118] proposed a rotating machinery condition monitoring method, in which the CNN was designed to process the infrared thermal images for feature extraction and fault classification. Chen et al. [119] developed a CNN based degradation state identification approach for planetary gear, in which the singular spectrum of raw data was used to train the model. Wang et al. [120] employed singular value decomposition based on the phase space reconstruction to analyze the bearing vibration signal and then utilized CNN to process the obtained features for bearing fault diagnosis. Zhu et al. [121] proposed a symmetrized dot pattern to transform vibration signals into 2D images, then these images were used to train a convolutional network for fault diagnosis. Li et al. [122] used K-singular value decomposition to enhance the resolution of TF features obtained by Wigner-Ville Distribution and then built a CNN for planetary gearbox fault classification. Udmale et al. [123] utilized the kurtogram of raw signals to train CNN for bearing fault diagnosis. In [124], Senanayaka proposed a gearbox fault diagnosis method based on multiple classifiers and data fusion. Specifically, the vibration spectrum was used as the input of multilayer perceptron while the features from STFT and CWT were used as the input of CNN. Finally, the naïve Bayes combiner was employed to integrate the results of two classifiers.

In this subsection, the applications of 2D CNN for fault classification were reviewed systematically. A clear and intuitive summary is displayed in Table 10, which aims to help reader search these studies quickly depending on the signal transform approach or analysis object.

Table 10
Summary of the applications of 2D CNN for fault classification.

Transform method	References	Object
Data matrix	[58 59 60 61 62 63]	Bearing
	[64 65]/[66]/[67]	Gear/bearing, wind turbine/bearing, pump
	[68]/[69]	Motor/compressor
Images	[70 71 72 73]	Bearing
	[74 75 76]/[77]	Bearing, gear/bearing, pump
Time or frequency domain transform	[79 80 81 82 83 84 85 86]/[78]/[87 88 89]	Bearing/bearing, gear/gear
Wavelet transform	[90 91 92 93 94]	Bearing
	[95 96 97 98]/[99]	Gear/gear, rotor, bearing
	[100 101]/[102]/[103]/[104]	Rotor/motor/bearing, gear/tool
Short time Fourier transform	[105 106]/[107 108]	Bearing/bearing, gear
	[109]/[110]	Bearing, pump/motor
Other	[111 112 114 115 116 117 120 123]/[118 119 122]	Bearing/gear
	[113]/[121]/[124]	Rotor/bearing, rotor/bearing, rotor, gear

4.1.2. 1D convolutional network based classification

In addition to 2D convolutional network based fault classification, a more direct strategy is to construct 1D convolutional diagnostic model to process original time-series data. In this subsection, the applications of 1D convolutional network for fault classification are reviewed according to raw sensor data types, such as vibration and AE data.

- Vibration data

Vibration data has been the most common source of information in machine fault diagnosis due to its legibility and intuitiveness. Eren [125] used raw vibration signal as the input to train a 1D CNN for IMS bearing fault detection. Pan et al. [126] combined the CNN with the long short term memory (LSTM) network and proposed an improved bearing fault diagnosis method. Inspired by the second generation wavelet transform, Pan et al. [127] improved the convolutional network and proposed a LiftingNet to process raw mechanical data for fault classification. Qian et al. [128] constructed an adaptive overlapping CNN for bearing fault diagnosis, in which the raw vibration signals were used to train model. Jia et al. [129] proposed a normalized CNN for imbalanced fault classification of machinery. In this method, a neuron activation maximization algorithm was presented to help understand the feature learning process of the network. Eren et al. [130] developed a compact adaptive 1D CNN for real-time bearing fault diagnosis. Ma et al. [131] integrated the residual convolutional network, deep belief network as well as deep auto-encoder and proposed an ensemble deep learning method for fault diagnosis of rotor bearing system. Wang et al. [132] proposed a multi-scale learning network with the 1D and 2D convolution channels to learn the local correlation of adjacent and nonadjacent intervals in vibration signals for bearing fault diagnosis. Huang et al. [133] added a multi-scale cascade layer at the front of conventional CNN and proposed a convolutional network with multi-scale information for bearing fault diagnosis. Qiao et al. [134] used raw vibration signals as the input and proposed an adaptive weighted multiscale CNN for bearing fault diagnosis under variable operating conditions. Abdeljaber et al. [135] utilized the compact CNN to present an online condition monitoring method for fault detection and severity identification of bearings.

Huang et al. [136] utilized raw vibration signals as the model input and proposed a deep decoupling CNN for intelligent compound fault diagnosis. Liu et al. [137] combined the denoising convolutional autoencoder with CNN to develop an anti-noise fault diagnosis method. Han et al. [138] proposed an enhanced convolutional network with enlarged receptive fields for planetary gearbox fault diagnosis. Considering the inherent multiscale characteristics of vibration signals, Jiang et al. [139] developed a multi-scale CNN for wind turbine gearbox fault diagnosis. In [140], Sun et al. firstly used back-propagation based neural network to learn the local filters. Then these local filters were used to build the feed-forward CNN for feature learning. Finally, the learned features were fed into SVM classifier for induction motor fault classification. Yuan et al. [141] applied multi-sourced heterogeneous monitoring data as the input and presented a multi-mode CNN based method for rotor system fault diagnosis. Afrasiabi et al. [142] proposed an accelerated CNN for bearing fault diagnosis of induction motors, in which the pruning connection and weight sharing technique were used to compress model without loss of accuracy. Chen et al. [143] utilized 1D CNN to learn features from raw vibration signals and then fed these features into a bidirectional LSTM network for wear state identification of tool.

In addition to directly using the raw vibration data, some scholars adopted the features extracted from vibration data to train 1D CNN for fault diagnosis. Xie et al. [144] used the CNN to learn features from frequency spectrum of vibration signals and then integrated these features with energy entropy of empirical mode decomposition and time domain features for final fault classification. Sadoughi et al. [145] used spectral kurtosis and envelope spectrum analysis to process raw mechanical data and proposed a physics-based CNN for fault diagnosis of rotating machinery. To maintain the diagnosis performance in the noisy environment and different working loads, Zhang et al. [40] developed a CNN with training interference based diagnosis method. In [146], Dong et al. used 1D CNN and 2D CNN to respectively extract features from the frequency spectrum and STFT spectrum of vibration signals, then these features were jointly used for rolling bearing degradation monitoring. Jing et al. [147] developed a CNN based method to extract features from the frequency spectrum of vibration signal for gearbox fault diagnosis. In [148], Ma et al. used coefficients of wavelet packet decomposition as the model input and proposed a lighted CNN for bearing fault diagnosis.

- Other data

Different from vibration data, other mechanical signals are also used in 1D convolutional network based classification, including current, AE, and build-in encoder data. For instance, Ince et al. [149] directly used raw current signals to train 1D CNN for real-time motor condition monitoring, the results showed the effectiveness and superiority of their approach. Besides, they proposed a real-time broken rotor bar fault detection model based on the shallow 1D CNN [150]. Khan et al. [151] used the motor current as the input and developed an analytical model for inter-turn fault diagnosis by combining the 1D CNN and LSTM network. Kao et al. [152] applied a 1D CNN to the fault diagnosis of magnet synchronous motor over a wide speed range by using current data as the model input. In [153], the motor vibration signal and the stator current signal were firstly segmented by analysis windows of varying lengths for the joint representation. Then, the CNN and LSTM network were designed to automatically learn discriminative features and achieve motor fault diagnosis. In addition to current data, AE data analysis is also a common monitoring manner for fault diagnosis. Li et al. [154] used convolutional network and gate recurrent unit (GRU) to reprehensively extract features from AE and vibration data. Then the learned features were concatenated for gear pitting fault diagnosis. In [155], Appana et al. utilized the CNN to process the envelope spectrums of AE signals to achieve bearing fault diagnosis under varying rotating speeds. In light of the drawbacks of external sensors, Jiao et al. [16] proposed a build-in encoder information based CNN for intelligent fault diagnosis. In this method, a multivariate encoder signal was presented based on information fusion to capture comprehensive mechanical health information, then a 1D CNN was designed for adaptive feature learning and condition classification.

According to the types of sensor data, the comprehensive review on the applications of 1D convolutional network is presented in this subsection. Depending on above literature, a concise generalization is displayed in Table 11.

4.1.3. Classification based on convolutional network variants

As introduced in Section 3.2, many variants of CNN have also been studied and applied to the field of fault diagnosis. Thus in this subsection, we will review these publications according to different network variants.

- Applications of ResNet in fault classification

Zhao et al. [156] employed a series of wavelet packet coefficients as the model input and proposed a deep residual network for planetary gearbox fault diagnosis under serious noise environment. The comparison results showed higher accuracies than other deep learning approaches. Furthermore, they [157] combined WT with ResNet and proposed the multiple wavelet coefficients fusion based deep residual network for planetary gearbox fault diagnosis, which aimed to learn more easily-distinguished features from the input data. Li et al. [158] proposed a deep residual learning network for fault diagnosis, in which the data augmentation techniques were presented to artificially create additional valid samples for model training. The results showed that their method can achieve high diagnosis accuracy with small original training dataset. Zhang et al. [159] used raw vibration signals as the model input to train a deep ResNet for bearing fault diagnosis, the results showed the superiority to traditional CNN model. Peng et al. [160] presented a deeper 1D CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. Ma et al. [161] proposed a lightweight deep residual convolutional network based on depthwise separable convolutions for bearing fault diagnosis, in which model parameters can be reduced while achieving high classification accuracy. Zhuang et al. [162] proposed a stacked residual dilated CNN for bearing fault diagnosis by combining the dilated convolution, the input gate structure of LSTM and the residual network. Su et al. [163] utilized raw time sequences as the input and presented a residual-squeeze net for fault diagnosis of high-speed train bogie. Ma et al. [164] proposed a fault diagnosis method of planetary gearbox under nonstationary running conditions using ResNet with demodulated time-frequency features. Considering the non-stationary conditions of machine, Liu et al. [165] proposed a multi-scale kernel based residual convolutional network for motor fault diagnosis. The results showed the superiority compared with state-of-the-art methods. From above review, it can be seen that ResNet diagnostic model is promising for more comprehensive feature extraction and higher diagnosis accuracy in modern industry, especially in complex mechanical equipment or industrial environment.

- Applications of GAN in fault classification

Due to the excellent data generation characteristics, GANs have been gradually applied to the field of fault diagnosis, especially for the diagnostic scenario with imbalanced data sets. Cao et al. [166] firstly transformed the time-domain signals into image data. Then a GAN was designed for rolling bearing fault classification. The results illustrated the potential of GAN on the fault diagnosis with small samples. Xie et al. [167] developed a GAN to generate the samples of minority classes for bearing fault diagnosis, which aimed to address the issue of data imbalance. Shao et al. [168] proposed an auxiliary classifier GAN based diagnostic framework to generate synthesized data and achieve induction motor fault diagnosis. Afrasiabi et al. [169] combined GAN with temporal CNN and proposed a wind turbine fault diagnosis method, in which the former was used as the feature extractor and the latter was used as the fault classifier. Li et al. [170] proposed an enhanced GAN for fault diagnosis of rotating machinery with imbalanced data. In their method, a 2D convolutional network was used to build the generator and discriminator, which aims to produce small samples to balance the dataset. Suh et al. [171] employed the nested scatter plot method to transform raw vibration signals into 2D images, then a 2D CNN was designed for bearing fault classification. In addition, a GAN was embedded in this framework to generate fault images for the data imbalance issue. To address the issue of lacking labeled fault data, Guo et al. [172] proposed a multi-label 1D GAN

Table 11
Summary of the applications of 1D CNN for fault classification.

Signal Type	References	Object
Vibration data	[125 126 127 128 129 130 131 132 133 134 135]	Bearing
	[136 137 138]	Bearing, gear
	[139]/[140 141 142]/[143]	Wind turbine gearbox/motor/tool
	[144 145 40 146 148]/[147]	Bearing/gear
Other data	[149 150 151 153]/[152 155]/[154 16]	Motor/bearing/gear

for fault diagnosis, in which the auxiliary classifier GAN was used to generate real damage data and then the generated data and real data are both used to train fault classifier. The experimental results showed that the proposed method can improve diagnosing accuracy from 95% to 98% when model was trained with the generated data.

- Applications on other variants in fault classification

Jiao et al. [10] employed the built-in encoder and external vibration signals as the input in parallel and presented a deep coupled dense convolutional network based intelligent fault diagnosis. The presented method can promote information passing and mitigate gradient vanishing/exploration issues, the results also verified the superiority than traditional convolutional network. Li et al. [173] presented an improved inception network to process raw vibration signals for gear pitting fault diagnosis. In [174], Chen et al. proposed a deep inception net with atrous convolution to bridge the gap between artificial and real damage for bearing fault diagnosis. Zhu et al. [175] employed STFT to convert signals into 2D graphs as the input and proposed a capsule network with the inception block and a regression branch for bearing fault diagnosis. Chen et al. [176] proposed a deep capsule network with stochastic delta rule for rolling bearing fault diagnosis, in which raw vibration signals were used as the model input.

4.1.4. Summary

This subsection reviews the classification applications in machine fault diagnosis using convolutional neural networks. In comparison to traditional neural networks, appealing characteristics, i.e. sparse local connection, weight sharing and down sampling, make CNN stand out in feature learning. In addition, advanced tricks, such as innovative convolution architectures, normalization, regularization, optimization strategies and so on, make CNN continue to produce surprising diagnostic results. To name a few, the diagnostic models with shortcut connections, i.e. ResNet and DenseNet, are able to alleviate the issue of gradient vanishing/-exploding while promote information passing. Depthwise separable convolutional network can significantly reduce the number of parameters while maintaining performance, thus it is promising to realize the "small, light, and fast" real-time diagnosis. Furthermore, according to above literature, advanced signal processing methods also promote the performance of CNN by pre-processing the input data for revealing more fault information. As a result, the combination of internal advantages and external assistance allows CNN to exhibit impressive fault diagnosis performance.

Despite these methods have achieved certain success, some more practical problems still cannot be ignored. For instance, the success of convolutional neural networks is based on the large-scale datasets with a tremendous number of labeled samples. However, in many practical situations, a large number of labeled samples are inaccessible, especially for fault data in highly sophisticated equipment. Furthermore, above most approaches assume that the distributions of training data and test data are same, however, this assumption is not hold in real industry. Consequently, it is necessary to solve these realistic problems and develop advanced convolutional diagnostic approaches for the promising employment in modern intelligent industry.

4.2. Applications in health prediction

Different from the fault classification, the purpose of health prediction is to track the degraded state of machinery, even if no apparent failure occurs. This branch is vitally important in the field of machine fault diagnosis, which allows maintenance personnel make early judgments and decisions to avoid losses and injuries. Therefore, the applications in health prediction are reviewed and summarized according to the application object in this section.

4.2.1. Health prediction of bearing

Rolling bearings play an important role and are widely used in modern machinery. The deterioration or failure of bearings will lead to machine breakdown and even disaster. Therefore, numerous studies have been conducted to assess and predict the health condition of bearings. Yoo et al. [177] used continuous WT to obtain the TF images for the health indicator (HI) construction. Then a CNN was designed to process these images for bearing RUL prediction. Belmiloud et al. [178] used wavelet packet decomposition to extract features as the model input and then presented a deep CNN based method for adaptive HI construction. Hinchi et al. [179] proposed a bearing RUL estimation method, in which the convolutional layer and LSTM layer were integrated to learn features from raw sensor data. Guo et al. [180] proposed a method for HI construction, in which the trend burr was considered and the results showed that their proposed method was more effective than other methods in terms of tradability, monotonicity and scale similarity. Ren et al. [181] presented a spectrum-principal-energy-vector algorithm to obtain the eigenvector raw signals, then a CNN was trained for bearing RUL prediction. She et al. [182] proposed a multi-channel CNN with exponentially decaying learning rate to construct wear indicator and evaluate the health of rolling bearing, in which the original multi-channel signals were used as the input. Mao et al. [183] employed CNN to learn features from the marginal spectrum of Hilbert-Huang transform. After that, a LSTM network was constructed for RUL prediction of bearings. Li et al. [184] used STFT to process raw vibration signals and obtain the TF domain information. Then a deep CNN was built to extract multi-scale features for RUL estimation. Zhu et al. [185] used WT to acquire the TF representation and then trained a multi-scale CNN to learn global and local features for RUL estimation. The results showed enhanced performance in prediction accuracy compared to tradition data-driven methods. Wang et al. [186] converted 1D signals into 2D images to train CNN for RUL prediction, in which the maximum correlation entropy with regular terms was employed as the loss function for better performance compared to the mean square error. Zhang et al. [187] proposed a deep multilayer perceptron convolutional network for HI construction, in which the outlier region correction method was introduced to detect and remove outliers while enhance the interpretability of HI. Yang et al. [188] utilized raw mechanical signals to trained a double-CNN model for RUL prediction, in which the first CNN was used to identify the incipient fault point and the second CNN model was applied for RUL prediction. Considering the prediction uncertainty, Peng et al. [189] introduced a Bayesian multi-scale convolutional network based method for bearing health prognostic, which shows the more accurate performance than point estimates. Yao et al. [190] proposed a bearing RUL estimation method by combining empirical model decomposition with ensemble CNNs, in which the former can reveal the nonstationary property of degradation data and help CNN to get a more accurate prediction. Wang et al. [191] integrated the CNN and LSTM network to process time-series data and calculate an unsupervised *H*-statistic for bearing performance degradation assessment. Liu et al. [192] presented a joint-loss CNN for bearing fault recognition and RUL prediction in parallel. The results showed the proposed method can capture common features between different relative tasks and improve the generalization capability. Wang et al. [193] proposed a deep separable convolutional network for RUL prediction of machinery, in which the data from different sensors were used to train a separable convolutional building block with a residual connection for feature learning. Furthermore, they proposed a recurrent convolutional network for RUL prediction [194], in which recurrent convolutional layers were designed to model the temporal dependencies and variational inference was utilized to quantify the uncertainty of prediction results.

4.2.2. Health prediction of turbofan engine

Benefitting from the public C-MAPSS dataset, many researchers validated the proposed prognosis approaches on the turbofan engine. Babu et al. [195] constructed 2D data matrix from multivariate time series to train a CNN with two-convolution layers and two-fully connected layers for RUL estimation. Li et al. [196] adopted the time window approach to process the multi-variate temporal data and then developed a deep CNN for feature extraction and RUL estimation. Wen et al. [197] presented a deep residual CNN for RUL estimation, in which the *k*-fold ensemble method was adopted to enhance the prediction preformation. Li et al. [198] presented a directed acyclic graph network for RUL prediction by combining CNN and LSTM network. The comparative results showed that the proposed method had better prediction accuracy. Al-Dulaimi et al. [199] proposed a hybrid deep network framework for RUL estimation, in which the LSTM network and CNN were arranged in parallel for feature learning and then a multilayer fully connected network was designed for feature fusion and decision making. Ruiz-Tagle Palazuelos et al. [200] introduced a capsule neural network for degradation estimation of turbofan engine and the results showed the superiority than traditional CNN based methods. Kong et al. [201] adopted the polynomial regression to construct HI and then designed a hybrid deep model based on CNN and LSTM network for RUL prediction.

4.2.3. Other health prediction application

In addition to the bearing and turbofan engine, some scholars also developed convolutional network based prediction approaches to other applications, such as CNC machine. Zhao et al. [202] proposed a convolutional bi-directional long short term memory (Bi-LSTM) network for CNC machining tool health monitoring. This method combined the advantages of CNN and LSTM network to obtain more accurate prediction. Qiao et al. [203] proposed a hybrid deep learning framework for gearbox fault diagnosis and tool wear prediction, in which the multiple convolutional layers and LSTM layers were firstly designed to extract local spatiotemporal features and then a holistic convolution-LSTM layer was designed to extract holistic spatiotemporal features. Aghazadeh et al. [204] employed CNN to establish a deep learning algorithm for tool wear estimation, in which WT and spectral subtraction algorithms were designed to intensify the effect of tool wear and reduce the effect of cutting parameters. Huang et al. [205] utilized features from time-domain, frequency domain and TF domain of multi-sensor signals as health information and proposed a deep CNN based method for tool wear prediction. In [206], Fu et al. combined the CNN and LSTM network to establish the logical relationship of observed variables for condition monitoring of wind turbine gearbox bearing. Kong et al. [207] presented a health monitoring method of wind turbines based on SCADA data. In their approach, the CNN and GRU were integrated to learn spatial and temporal features, and then the exponential weighted moving average control chart was designed for condition recognition. Luo et al. [208] employed dual-tree complex wavelet to obtain multiscale characteristics as the input. Then an enhanced convolutional LSTM network was designed for damage monitoring of the automotive suspension component. Li et al. [209] developed a scalable degradation assessment approach for bandsaw machine by proposing a dual-phase modeling method. In this approach, a physics informed model was firstly established to generate the HI to monitor wear condition using the vibration and acoustic signals. Then a deep CNN based surrogate model was designed to replace the physics informed model by using alternative low-cost sensor data.

4.2.4. Summary

In this section, applications of convolutional networks in health prediction are systemically reviewed. The summarization reveals that many researchers have successfully deployed convolutional networks based prediction approaches to machine health prediction, which are able to address weak generality, flexibility and intelligence of previous physical and mathematical models. Nevertheless, more existing issues require to be pointed further. It can be seen that most existing methods mainly provide the point estimate rather than a prognostics distribution as the final result, which may be rigid and inflexible in realistic industry. More specifically, that is, current most methods are unable to quantify the prediction uncertainty. However, uncertainty is critical for both health prognostics and subsequent decision making, which is especially important for equipment with high safety requirements. Therefore, the exploration on uncertainty quantification is promising in future prediction research. In addition, it should be noticed that lifetime data are difficult to obtain in practical industry. In other words, there is usually no sufficient data to train a complete life prediction model, hence how to build models using experimental or simulation data and then expand them to practical industrial applications should be paid to more attention.

4.3. Applications in transfer diagnosis

Although CNN on fault classification and health prediction of machinery have acquired certain achievements, an assumption that training data and test data have the same data distribution is necessary for most of the above methods. In practical industrial scenarios, the data distribution differences are inevitable due to natural wear of equipment, changes in operating conditions, interference from environment and human, and so on. Consequently, the performance of above most models will be seriously degraded when the data distributions between training set (source domain) and test set (target domain) are different. An immediate solution is to retrain or build new model, however, a large number of labeled data are necessary in this case. In many task scenarios, sufficient labeled instances are either difficult to collect, or their labeling cost is prohibitive. Therefore, it is quite necessary to explore how to apply the previously models established on the related domain to the new diagnostic scenarios. In light of these issues, transfer learning or domain adaptation technologies have been introduced to machine fault diagnosis, especially its combination with the deep convolutional networks. In this section, a summary on the applications of convolutional network in transfer diagnosis will be introduced in detail. Before starting the literature review, three common tricks are firstly introduced, including parameters transfer, moment matching strategies, and adversarial domain adaptation.

4.3.1. Parameter transfer

Parameter transfer is also called pre-train model based transfer, which means that the partial parameters of network trained in the source domain are fixed and transferred while remaining parameters will be fine-tuned using labeled data in the target domain. Intuitive understanding is shown in Fig. 22, in which the idea comes from the fact that the features of previous layers are general and transferable while the features of the last few layers are task-specific. Therefore, the model can be used to new target domain by using the few labeled target data to fine-tune parameters of task-specific layers.

Some researchers have applied this technology to achieve transfer diagnosis of machinery in recent years. Cao et al. [210]

proposed a deep CNN based transfer learning approach for gearbox fault diagnosis. The first part of their method was constructed by a part of a pre-trained network and the second part was the fully connected layers retrained by gear data. Hasan et al. [211] used the frequency spectrum as the input and presented a parameter transfer based 1D convolutional network for bearing fault diagnosis under variable working conditions. Hemmer et al. [212] employed a CNN pre-trained by ImageNet dataset to learn features from WT images of vibration and AE signals. Then a sparse autoencoder-based SVM was designed to process these features for bearing fault classification. Zhong et al. [213] proposed a transfer learning framework for gas turbine fault diagnosis, in which the CNN trained on large-scale annotated normal dataset was transferred to fault diagnosis task with limited fault data for feature learning and then the SVM was designed as the classifier for fault classification. Wen et al. [214] converted the raw time-domain signals to RGB images to fine-tune a pre-trained ResNet-50 model for fault diagnosis. Furthermore, they utilized the negative correlation learning to retrain several fully-connected layers and the Softmax classifier of the pre-trained ResNet-50 for fault classification [215]. Han et al. [216] presented a transfer learning framework for fault diagnosis of unseen machine conditions, in which the CNN trained on large datasets was transferred to new tasks with proper fine-tuning. In addition, they designed three transfer learning strategies to investigate the feature transferability in the different network levels. Shao et al. [217] employed WT to convert raw signals into images and developed a deep transfer framework for machine fault diagnosis, in which the labeled TF images were used to fine-tune the higher layers of a convolutional network pre-trained by ImageNet dataset. Ma et al. [218] introduced a frequency slice wavelet transform method to process the raw vibration signals into 2D TF images. Then these images were used to dine-tune a pre-trained AlexNet model for bearing fault diagnosis. Hasan et al. [219] used the acoustic spectral images of AE signals to reflect mechanical health state and proposed a CNN based parameter transfer learning approach for bearing fault diagnosis under variable speed conditions. Chen et al. [220] proposed a parameter transfer based method for fault diagnosis of rotary machinery, in which a wide kernel 1D CNN was designed for learning transferable features.

In this subsection, applications on parameters transfer based fault diagnosis are reviewed. Although this transfer strategy is easy to understand and operate, a distressing issue still exists that labeled data in the target domain is necessary. Therefore, these transfer learning approaches will encounter unexpected obstacles due to the unavailability of labeled data in practical industrial applications.

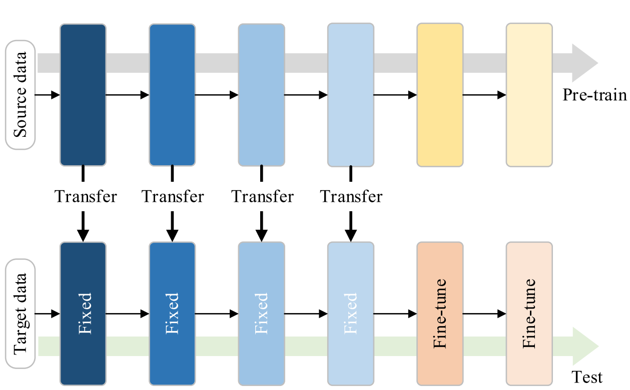


Fig. 22. Illustration of parameter transfer.

4.3.2. Discrepancy measure based transfer diagnosis

Discrepancy measure based method is generally achieved by minimizing a certain distance between hidden activations of convolutional network in different domains, in which the key is to explore the efficient discrepancy metric function. The common metrics include Maximum Mean Discrepancy (MMD) and correlation alignment. In this subsection, the definition of MMD and correlation alignment is firstly introduced to help understand this transfer learning manner, then the applications on discrepancy measure based transfer diagnosis are systemically reviewed.

Maximum mean discrepancy [221,222] measures the distribution divergences by the mean embedding of two distributions in the reproducing kernel Hilbert space $\mathcal{H}$. Specifically, given the source domain $\mathcal{D}_s = \{\mathbf{x}_i^s\}$ and target domain $\mathcal{D}_t = \{\mathbf{x}_j^t\}$, which are drawn from distributions P and Q, respectively. The MMD between two domains can be calculated as:

$$\mathcal{L}_{MMD} = \sup_{\phi\in\mathcal{H}} (\mathbb{E}_P[\phi(\mathbf{x}^s)] - \mathbb{E}_Q[\phi(\mathbf{x}^t)]) \tag{7}$$

where ϕ represents the feature map; $P = Q$ if and only if $\mathcal{L}_{MMD} = 0$. In practical application, the MMD is calculated as the empirical estimation based on the kernel mean embedding:

$$\begin{aligned}\hat{\mathcal{L}}_{MMD} &= \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\phi(\mathbf{x}_i^s) - \frac{1}{n_t}\sum_{i=1}^{n_t}\phi(\mathbf{x}_i^t) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s}k(\mathbf{x}_i^s,\mathbf{x}_j^s) + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}k(\mathbf{x}_i^t,\mathbf{x}_j^t) - \frac{2}{n_s n_t}\sum_{i=1}^{n_s} \\ &\quad \times \sum_{j=1}^{n_t}k(\mathbf{x}_i^s,\mathbf{x}_j^t)\end{aligned} \tag{8}$$

where $k(\cdot,\cdot)$ represents the characteristic kernel; n_s and n_t are the number of source samples and target samples.

Different from MMD metric, correlation alignment [223] is defined based on the second-order statistics. Mathematically, given the source feature matrix $D_s \in \mathbb{R}^{n_s\times d}$ and target feature matrix $D_t \in \mathbb{R}^{n_t\times d}$, where the row represents the number of sample and the column denotes the feature dimension. The covariance matrices of two feature matrices can be calculated as:

$$\begin{aligned} C_s &= \tfrac{1}{n_s-1}(D_s^{\mathrm{T}}D_s - \tfrac{1}{n_s}(\mathbf{1}^{\mathrm{T}}D_s)^{\mathrm{T}}(\mathbf{1}^{\mathrm{T}}D_s)) \\ C_t &= \tfrac{1}{n_t-1}(D_t^{\mathrm{T}}D_t - \tfrac{1}{n_t}(\mathbf{1}^{\mathrm{T}}D_t)^{\mathrm{T}}(\mathbf{1}^{\mathrm{T}}D_t)) \end{aligned} \tag{9}$$

where C_s and C_t stand for the covariance matrices of two domains, respectively. $\mathbf{1}$ is a column vector with all elements equal to 1; $\cdot^T$stands for the transposition. Based on two covariance matrices, the correlation alignment is defined as follows:

$$\mathcal{L}_c = \frac{1}{4d^2}\| C_s - C_t \|_F^2 \tag{10}$$

where $\|\cdot\|_F^2$ represents the squared matrix Frobenius norm. It is worth noting that as MMD is usually calculated with the help of kernel tricks, it will produce higher computational complexity than correlation alignment.

In the community of fault diagnosis, Zhang et al. [224] proposed a CNN for cross-domain fault diagnosis under varying working conditions, in which the MMD was used to minimize the domain difference. Li et al. [225] presented a convolutional network model for rolling bearing fault diagnosis under noisy and changing working condition. In their approach, a feature clustering method was introduced to minimize the difference of intra-class and maximize the difference of inter-class. Meanwhile, the MMD was adopted to reduce the domain difference. Furthermore, they proposed a multi-layer domain adaptation approach for bearing fault diagnosis [226], in which the multi-kernel maximum mean discrepancy was employed as the metric function to reduced distribution differences between different domains. In [227], Xiao et al. presented a domain adaptation method for motor fault diagnosis, in which CNN was adopted to extract multi-level features from raw vibration signals and the MMD was incorporated to reduce the feature distribution differences of multiple layers. Yang et al. [228] combined the CNN with MMD to introduce a transfer learning network for fault diagnosis from laboratory bearings to locomotive bearings. Han et al. [229] extended the marginal distribution adaptation to the joint distribution adaptation and proposed a deep transfer network for fault diagnosis. Xu et al. [230] proposed a convolutional transfer discrimination network for unbalanced fault diagnosis under variable rotational speeds, in which the MMD was used to reduce the distribution differences of high-dimensional features. Zhu et al. [231] converted raw vibration data into gray pixel images as the network input and proposed a deep transfer learning approach based on multi-Gaussian kernels MMD for rolling bearing fault diagnosis under different operating conditions. To address the cross-domain diagnosis problem with insufficient target samples, Li et al. [232] used the MMD based generative convolutional networks to generate fake target fault samples and proposed a domain adaptation approach for bearing fault diagnosis. In [233], a renewable fusion method was proposed to address the fault diagnosis scenarios with variable speed conditions and unbalanced samples, in which the second order statistics alignment was used to reduce feature distribution differences and the contrastive loss function was employed to promote the similar features learning.

4.3.3. Adversarial learning based transfer diagnosis

In addition to above two moment matching algorithms, another transfer learning strategy, named the adversarial domain adaptation [234], is attracting increasing attentions. Unlike the discrepancy measure based method, adversarial domain adaptation constructs a two-player minimax game to learn transferable features. Specifically, as shown in Fig. 23, the adversarial domain adaptation network is usually composed of a feature extractor F, a label classifier C and a domain discriminator D, in which the discriminator is trained to distinguish whether the features are from the source domain or the target domain while the feature extractor tries to fool the discriminator. In the training process, meanwhile, the classifier is trained to minimize the classification error of source data. Give the source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ with n_s samples and target domain $\mathcal{D}_t = \{(\mathbf{x}_i^t)\}_{i=1}^{n_t}$ with n_t samples. The overall objective of the adversarial domain adaptation network is described as follows:

$$\mathcal{L}_{ada} = \frac{1}{n_s}\sum_{\mathbf{x}_i \in \mathcal{D}_s} \mathcal{L}_y(C(F(\mathbf{x}_i)), y_i) - \frac{\lambda}{n_s + n_t} \times \sum_{\mathbf{x}_i \in (\mathcal{D}_s \cup \mathcal{D}_t)} \mathcal{L}_d(D(F(\mathbf{x}_i)), d_i) \tag{11}$$

where $\mathcal{L}_y$ represents the classification loss function; $\mathcal{L}_d$ denotes the domain identification loss function; y_i and d_i stand for the category label and domain label, respectively; λ is the trade-off parameter.

Inspired by the adversarial domain adaptation, Han et al. [41] introduced a deep adversarial convolutional network for machine fault diagnosis and the results showed that the proposed method was superior to the conventional convolutional networks. Guo et al. [235] integrated the moment matching with adversarial learning strategies to develop a deep convolutional transfer learning network for fault diagnosis of machines, in which the training and test dataset were acquired from different machines. Jiao et al. [236] developed a residual joint adaptation adversarial network for cross-domain intelligent diagnosis, in which the joint maximum mean discrepancy was employed to align joint feature distributions while the domain discriminator is deployed for auxiliary adversarial adaptation. Zhang et al. [237] proposed a Wasserstein distance guided multi-adversarial convolutional network for fault diagnosis under different operating conditions. The experimental results showed the improved performance than MMD based methods. To improve the Wasserstein distance-based adversarial approach, Wang et al. [238] presented an adversarial domain adaptation network based on the triplet loss for bearing fault diagnosis and the results showed the better performance. Xie et al. [239] proposed a transfer learning approach for fault diagnosis using the cycle-consistent GAN, in which the GAN was designed to generate new sample for unknown conditions to pre-train a classifier. From the perspective of decision boundaries, Jiao et al. [240] developed an unsupervised adversarial adaptation network to achieve cross-domain fault diagnosis. In this method, two task classifiers were constructed without the domain discriminator. Furthermore, they presented a novel weighed domain adaptation network to address more realistic problem [241], i.e. partial transfer diagnosis, in which the source domain and target domain have different label spaces.

In addition to above three popular transfer diagnosis methods, several ingenious technologies for transfer diagnosis are also introduced. For example, Zhang et al. [42] presented a CNN for cross-domain diagnosis, in which the wide convolutional kernels and adaptive batch normalization were adopted to achieve the domain adaptation ability. Duan et al. [242] proposed an auxiliary model based domain adaptation method for reciprocating compressor diagnosis under different operating conditions, in which a pre-trained CNN was used for feature learning and a marginalized stacked auto-encoder was used to eliminate data distribution difference. Zhang et al. [243] proposed an instance-based ensemble deep transfer learning network for degradation recognition of ball screw. In this method, the source instances were firstly filtered out iteratively, then multiple auto-encoders with different activations were trained for feature extraction, finally, the SVM was used for degradation recognition. Xiao et al. [244] proposed a transfer learning diagnostic model by integrating the modified TrAdaBoost algorithm and convolutional neural network, which is able to address small sample size problem in machinery fault diagnosis.

4.3.4. Summary

In this section, applications in transfer diagnosis are methodically reviewed. Through summarizing related literature, some observations can be found. Although the parameter transfer based technology is intuitive and easy-to-operate, the annotated data of the target domain are necessary. Therefore, it may cause these methods to be limited in realistic industry due to the difficulty of data labelling. In contrast, discrepancy measure and adversarial learning approaches do not require labeled samples of target

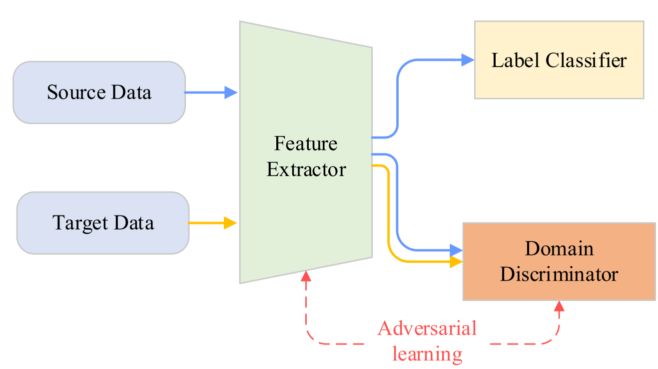


Fig. 23. Illustration of adversarial domain adaptation.

domain and thus can tackle the unsupervised domain adaptation diagnostic problems. In particular, owing to the artful design and outstanding performance, the adversarial domain adaptation is attracting more and more attention and is promising in the future fault diagnosis research. To conclude, a concise and convenient summary is listed in Table 12. The first column represents different transfer strategies and the third column denotes specific methodology for achieving transfer diagnosis. The last column stands for the application scenarios, where "Image → machine" represents that the pre-trained model is from the field of image processing; "operating conditions" and "machines" denotes the transfer between different conditions or different machines, respectively.

4.4. Epilog

In previous sections, the published literature on CNFD has been systemically reviewed. In summary, a pie chart displayed in Fig. 24 is given to describe the overall research profile. It can be found from Fig. 24 that more than 65% of publications are focused on the fault classification task. In addition, most of transfer diagnosis approaches are also oriented to the health condition classification. In contrast, the number of literature on health prediction accounts for only 17.6% of the total. This phenomenon may be because the implementation of fault classification is easier and intuitive while the health prediction usually requires additional assistance, such as the HI construction and health stage division etc. Nevertheless, it is worth noting that the fault classification mainly focuses on various failure modes and the machinery is already unable to work and produce normally when these failures occur. Prior to these obvious faults, mechanical equipment usually undergoes a degradation process, in which the foreseeable action should be taken instead of waiting for the eventual failure. Therefore, health prediction, i.e. degradation monitoring or RUL prediction, should be paid more attentions in future research.

Next, some diagnostic results for various public databases are further summarized for convenient reference and comparison. Table 13 summarizes some representative results in classification applications on public datasets. In this table, the second column describes the main network architecture, in which only convolution modules are listed, specifically, 3-2D Conv represents that the model contains three 2D convolutional layers. The third and fourth columns list the optimization algorithms and main characteristics of these methods. After that, the last two columns give the dataset and average results, in which "(Categories)" represents the number of categories selected when using this dataset.

Table 14 shows some representative results in health prediction application on public datasets, in which the "(Cases)" indicates the number of cases constructed when using this dataset. Furthermore, different from the classification application, there are various evaluation indicators in health prediction application, such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), normalized root mean square error (NRMSE), and so on. Therefore, it is necessary to ensure the consistency of evaluation indicators during reference and comparison.

Several representative results in transfer diagnosis are listed in Table 15, where the fifth column indicates the number of transfer scenarios constructed when using this dataset. The result in last column represents the average result of various scenarios. Furthermore, it should be pointed out that the related literature on this field is huge and abundant, meanwhile the new research is also emerging. Therefore, it is inevitable that some papers are missing from the current review. In addition, some non-English publications are also not considered in this work because of the limitation of language proficiency.

Table 12
Summary of applications in transfer diagnosis.

Transfer strategy	Reference	Methodology	Scenarios
Parameter transfer	[210 212 214 215 217 218]	Pre-train by image data	Image → machine
	[211 213 216 219]	Pre-train by mechanical data	operating conditions
	[220]		operating conditions and machines
Discrepancy measure	[224 225 226 227 229 230 231 232]	MMD	operating conditions
	[228]		machines
	[233]	Correlation alignment	operating conditions
Adversarial learning	[41]	Adversarial discriminator	operating conditions
	[235 236]	discriminator and MMD	Machines/operating conditions
	[237 238]	Adversarial Wasserstein	operating conditions
	[239]	GAN	operating conditions
	[240 241]	Classifier discrepancy	operating conditions
Other	[42 242 243 244]	/	operating conditions

5. Conclusions

(1) Almost all models mentioned in the above literature are trained and tested in the experimental or simulated scenario, thus these models may be unsuitable to be directly applied to realistic industry due to data differences. In addition, the training data and test data even come from the same experiment in some research, which thus can produce the excellent results owing to the data similarity and make researchers blindly believe in the algorithm ability. Consequently, it is of significance to improve model generalization and study transfer learning algorithms by utilizing the reasonable experimental data and realistic industrial data.

(2) It is known that the proper parameter setting plays a decisive role in the performance of deep convolutional network. Reviewing the above literature, the design and selection of network parameters (including architecture and hyperparameters) are mainly determined by authors' subjectivity while a specific standard or rule for selecting appropriate parameters has not been formed. Although a set of parameters cannot be ideally applied to various tasks, the study on parameter selection trick or the relation between parameters and mechanical signal characteristics is still promising and significant.

(3) In some applications, especially 2D convolutional networks, the data transform or signal processing technology is necessary. Thus it will increase the complexity of the overall framework and reduce the efficiency as well as the level of intelligence. On the contrary, the diagnostic models based on raw signals can avoid the requirements for domain knowledge while construct the end-to-end diagnostic framework. But the noise and interference existed in raw data may disturb the model convergence and even lead the model astray. Therefore, it is suggested to objectively view the merit and demerit between the deep convolutional networks and advanced signal processing algorithms, and organically integrate them for better performance.

6. Prospects

Despite CNNs have achieved great advancements in machine fault diagnosis, there still several aspects need to be further explored and investigated. Therefore, in this section, we will share

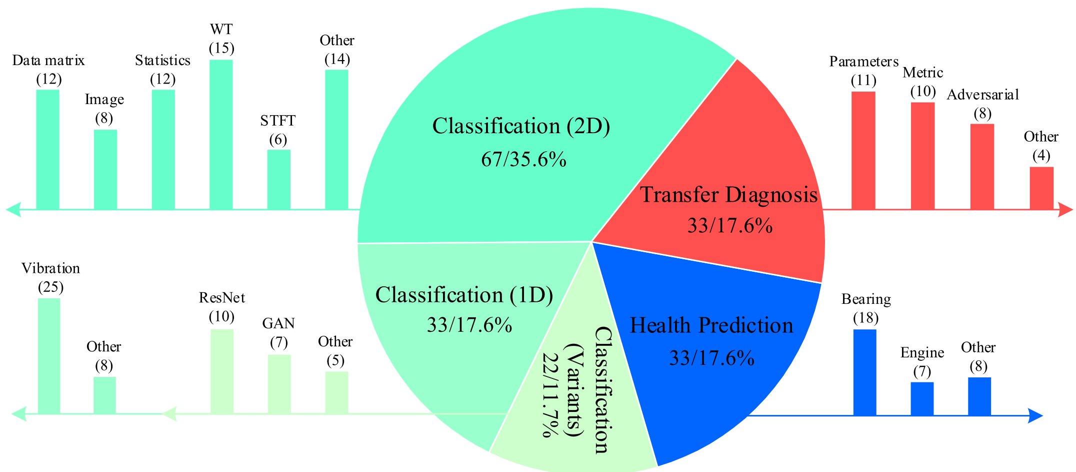


Fig. 24. Pie chart of various CNFD applications.

Table 13
Results of classification applications on public datasets.

Reference	Architecture	Optimizer	Characteristic	Dataset (Categories)	Result
[58]	3-2D Conv	SGD	Identify the fault type and size in two stages	CWRU (4)	97.90% (type) 98.10% (size)
[66]	3-2D Conv	SGD	Training and test data have different fault severities	CWRU (4)	92.40%
[70]	2-2D Conv	SGD	Multi-sensor information fusion	CWRU (9)	99.41%
[72]	2-2D Conv	Adam	Convolution with Bi-GRU	CWRU (10)	99.92%
[127]	LiftingNet	SGD	Inspired by CNN with second generation WT	CWRU (4)	99.63%
[40]	6-1D Conv	SGD	Adapt to noisy environment and load change	CWRU (10)	96.10%
[124]	2-2D Conv	SGD	Multiple classifiers for decision layer fusion	PHM09 (8/6)	96.07%
[147]	1-1D Conv	SGD	Analyze different inputs and network architectures	PHM09 (6)	99.33%
[73]	4-2D Conv	Adam	Decision-level information fusion	Paderborn (3)	97.27%
[132]	1D + 2D Conv	/	Multiscale learning and depthwise convolution	Paderborn (3)	98.58%
[174]	Inception net	Adam	Dilated convolution, diagnosis from artificial to real damages	Paderborn (3)	94.50%
[161]	Residual CNN	/	Depthwise Separable Convolutions and noise resistance	IMS (4)	97.75%

Table 14
Results of health prediction on public datasets.

Reference	Architecture	Characteristic	Dataset (Cases)	Evaluation and result
[192]	6@1D Conv	Joint loss	IMS (2)	MSE 0.0172
[193]	7@1D Conv	Separable convolution	C-MAPSS (4)	RMSE 16.17
[196]	5@1D Conv	Time window for sample preparation	C-MAPSS (4)	RMSE 17.73
[197]	1D Residual Conv	K-fold ensemble	C-MAPSS (4)	RMSE 17.50
[199]	CNN with LSTM	Multimodal feature learning in parallel	C-MAPSS (4)	RMSE 14.66
[202]	CNN with Bi-LSTM	Hybrid model for feature learning	PHM10 (3)	MAE 7.23 RMSE 9.23
[205]	3@2D Conv	Multi-domain feature fusion	PHM10 (3)	MAE 0.57 RMSE 0.80
[180]	2@1D Conv	Health indicator construction by considering trend burr	FEMTO (17)	Trendability 0.897; Monotonicity 0.406; Scale similarity 0.904
[184]	3@2D Conv	Multi-scale feature learning	FEMTO (17)	MAE 19.66; RMSE 23.62
[185]	3@2D Conv	Multi-scale feature learning	FEMTO (5)	MAE 1091.8; NRMSE 0.3514
[189]	3@2D Conv	Bayesian prognostics uncertainty	FEMTO (1)	MAE 19.8 RMSE 24.8
[194]	4@1D Recurrent Conv	Recurrent convolution with uncertainty quantification	FEMTO (4)	Cumulative relative accuracy 0.7942

some prospects with the readers, researchers and engineers who aim to promote the development of this field.

(1) More theoretical investigation is necessary to reveal the “black box” issue of CNFD approaches.
Although CNNs have been widely applied to fault diagnosis, in-depth theoretical research is still very rare. For example, the relation between the weights and the mechanical features has not been reasonably explained. In addition, the “black box” issue will make the companies or factories doubt the capabilities of these methods and refuse to apply them to realistic scenarios. Therefore, it is urgent to lift the veil of CNFD methods whether in academia or industry.

(2) How to identify unseen damage types or fault conditions?
The literature on the applications of CNFD generally only focuses on identifying the faults existed in the training set. It means that the model can be used only when the category of the test dataset is included in the training set. However, the construction of all-encompassing training set is expensive or even impossible. Moreover, some strange faults will inevitably occur in real scenarios with the changing of equipment itself and working environment. As a result, it

Table 15
Results of transfer diagnosis on public datasets.

Reference	Architecture	Optimizer	Dataset (Categories)	Scenarios	Result
[224]	5@1D Conv	Adam	CWRU (10)	6	99.60%
			PHM 09 (6)	12	82.62%
[240]	4@1D Conv	Adam	CWRU (10)	9	99.76%
[229]	5@1D Conv	SGD	CWRU (3)	2	98.20%
[233]	5@1D Conv	SGD	CWRU (10)	12	98.86%
[231]	2@2D Conv	SGD	CWRU (4)	12	94.79%
[241]	4@1D Conv	Adam	CWRU (10 to 4)	12	99.96%
[42]	5@1D Conv	Adam	CWRU (10)	6	95.90%
[41]	5@1D Conv	SGD	PHM09 (8)	1	99.40%
[238]	4@1D Conv	Adam	CWRU (10)	12	98.48%
			Paderborn (3)	6	99.46%
[215]	ResNet-50	SGD	Paderborn (3)	1	98.73%
[235]	6@1D Conv	SGD	CWRU-IMS	2	89.80%

is still an open question to explore the models which could distinguish the unseen damages or faults.

(3) There is a requirement to detect early damage from the point of quantitative analysis.
Most of CNFD applications only focus on how to identify the different health categories, however, there is no obvious fault types in early degradation stage. In particular, it is unreasonable to carry out diagnosis until the occurrence of large or significant failures in high-precise and vital industrial applications. Moreover, the decision should be performed according to the level of damage by the qualitative analysis. Considering this issue, therefore, the effort to explore the quantitative analysis of early weak damages should be encouraged.

(4) How to train the model using non-stationary data for diagnosis or prognosis under variable operating condition?
In previous variable operating conditions studies, the network models are usually trained by the data of smooth operation or the time-invariant features extracted by signal processing technologies. As a result, the former is only from one stationary condition to another and the later need certain expert knowledge. Moreover, collecting the stationary data is difficult and even impossible in the realistic continuously non-stationary operating environment. Consequently, how to use the non-stationary data to train model and achieve the reliable diagnosis is also an urgent problem.

(5) How to speed up the CNFD algorithm for the real-time diagnostic requirements?
It is known that a frequently mentioned drawback for the CNFD approaches is that they consume more time for training than classical shallow algorithms, thus this will lead to unsatisfactory in real-time and quick task requirement. Therefore, the exploration of the novel technology and trick is required to be paid more attentions to accelerate the CNFD algorithm in future research.

(6) How to establish the CNFD models for the requirement of equipment fleet?
According to the above literature, existing methods are mostly employed for the diagnosis and prognosis of a single machine. However, the cluster machine development is becoming an increasing trend in the rapid manufacturing and production era. Therefore, it is more significant to study the CNFD models with powerful generalization capability, so that to be freely applied to the other similar machines.

(7) How to utilize the opportunity of industrial big data to improve the performance of CNFD methods?
The sufficient data is the premise and foundation to achieve the excellent performance of deep networks. Reviewing the literature, the choice of data quantity heavily depends on the subjective factors or is limited by the experimental condition. As a result, many models are not optimal due to the simple or limited amount data. Therefore, how to seize the chance of industrial big data and utilize its characteristic, such as diversity and heterogeneity, to develop more robust and reliable models will be another promising topic in next research.

CRediT authorship contribution statement

Jinyang Jiao: Conceptualization, Methodology, Software, Writing - original draft. **Ming Zhao:** Writing - review & editing. **Jing Lin:** Supervision, Project administration, Funding acquisition. **Kaixuan Liang:** Writing - review & editing.

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.

Acknowledgments

This research was supported by National Natural Science Foundation of China (Grant Nos. 51421004, 91860205), the Defense Industrial Technology Development Program (Grant No. JCKY2018601C013), Fundamental Research Funds for the Central Universities (Grant No. xzy022019022).

References

[1] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, A.V. Vasilakos, A manufacturing big data solution for active preventive maintenance, IEEE Trans. Ind. Inform 13 (4) (2017) 2039–2047.
[2] Y. Lei, F. Jia, J. Lin, S. Xing, S.X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, IEEE Trans. Ind. Electron. 63 (5) (2016) 3137–3147.
[3] J. Jiao, M. Zhao, J. Lin, K. Liang, Hierarchical discriminating sparse coding for weak fault feature extraction of rolling bearings, Reliab. Eng. Syst. Saf. 184 (2019) 41–54.
[4] M. Zhao, X. Jia, A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery, Mech. Syst. Signal Pr. 94 (2017) 129–147.
[5] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: a review, Measurement 48 (2014) 292–305.
[6] M. Yu, D. Wang, M. Luo, Model-based prognosis for hybrid systems with mode-dependent degradation behaviors, IEEE Trans. Ind. Electron. 61 (1) (2014) 546–554.
[7] S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, W. Zhantao, Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing, Knowl.-Based Syst. (2019) 105022.
[8] K. Liang, M. Zhao, J. Lin, J. Jiao, An information-based K-singular-value decomposition method for rolling element bearing diagnosis, IAS Trans. 96 (2020) 444–456.

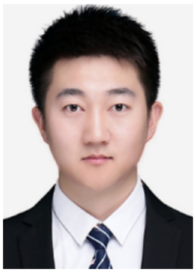
[9] Y. Liu, H. Gao, L. Guo, A. Qin, C. Cai, Z. You, A data-flow oriented deep ensemble learning method for real time surface defect inspection, IEEE Trans. Instrum. Meas. (2019).
[10] J. Jiao, M. Zhao, J. Lin, C. Ding, Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis, IEEE Trans. Ind. Electron. 66 (12) (2019) 9858–9867.
[11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[12] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[13] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Signal Pr. 72 (2016) 303–315.
[14] M. Gan, C. Wang, et al., Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, Mech. Syst. Signal Pr. 72 (2016) 92–104.
[15] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, J. Wang, Machine health monitoring using local feature-based gated recurrent unit networks, IEEE Trans. Ind. Electron. 65 (2) (2018) 1539–1548.
[16] J. Jiao, M. Zhao, J. Lin, J. Zhao, A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes, Knowl.-Based Syst. 160 (2018) 237–250.
[17] J. Zhang, S. Yi, G. Liang, G. Hongli, H. Xin, S. Hongliang, A new bearing fault diagnosis method based on modified convolutional neural networks, Chinese J. Aeronaut 33 (2) (2020) 439–447.
[18] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Proc. Adv. Neural Inf. Process. Syst (2012) 1097–1105.
[19] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: a review, Mech. Syst. Signal Pr. 108 (2018) 33–47.
[20] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Pr. 115 (2019) 213–237.
[21] D. Hoang, H. Kang, A survey on Deep Learning based bearing fault diagnosis, Neurocomputing 335 (2019) 327–335.
[22] M. Zhao, J. Lin, Health assessment of rotating machinery using a rotary encoder, IEEE Trans. Ind. Electron. 65 (3) (2018) 2548–2556.
[23] C. Ding, M. Zhao, J. Lin, J. Jiao, K. Liang, Sparsity-based algorithm for condition assessment of rotating machinery using internal encoder Data, IEEE Trans. Ind. Electron. 67 (9) (2020) 7982–7993.
[24] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: a systematic review from data acquisition to RUL prediction, Mech. Syst. Signal Pr. 104 (2018) 799–834.
[25] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, Mech. Syst. Signal Pr. 64 (2015) 100–131.
[26] https://www.phmsociety.org/competition/PHM/09.
[27] C. Lessmeier, J.K. Kimotho, D. Zimmer, W. Sextro. 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. In: Editor. Pub Place; 2016. pp. 5–8.
[28] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, J. Sound Vib. 289 (4–5) (2006) 1066–1090.
[29] A. Saxena, K. Goebel, D. Simon, N. Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In: Editor. Pub Place: IEEE; 2008. pp. 1–9.
[30] https://www.phmsociety.org/competition/phm/10.
[31] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In: Editor. Pub Place: IEEE Catalog Number: CPF12PHM-CDR; 2012. pp. 1–8.
[32] A. Soualhi, K. Medjaher, N. Zerhouni, Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression, IEEE Trans. Instrum. Meas. 64 (1) (2014) 52–62.
[33] R.K. Singleton, E.G. Strangas, S. Aviyente, Extended Kalman filtering for remaining-useful-life estimation of bearings, IEEE Trans. Ind. Electron. 62 (3) (2014) 1781–1790.
[34] http://www.mfpt.org/FaultData/FaultData.htm.
[35] https://drive.google.com/open id=1_ycmG46PARiykt82ShfnFfyQsaXv3_VK.
[36] https://figshare.com/articles/Gear_Fault_Data/6127874/1.
[37] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, In Proc. Int. Conf. Learn. Representations, 2016.
[38] F.C.C.O. Chollet. Xception: Deep learning with depthwise separable convolutions. In Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2017, pp. 1251–1258.
[39] H. Wang, Z. Liu, D. Peng, Y. Qin, Understanding and learning discriminant features based on multi-attention 1DCNN for wheelset bearing fault diagnosis, IEEE Trans. Ind. Inform. (2019).
[40] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Pr. 100 (2018) 439–453.
[41] 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.
[42] 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-Basel 17 (2) (2017) 425.
[43] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, Proc. Adv. Neural Inf. Process. Syst. (2010) 807–814.
[44] A.L. Maas, A.Y. Hannun, A.Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In Proc. Int. Conf. Mach. Learn., 2013.
[45] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, Proc. IEEE Int. Conf. Comput. Vis. (2015) 1026–1034.
[46] D.E. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint arXiv:1511.07289, 2015.
[47] K. Jarrett, K. Kavukcuoglu, M. Ranzato, Y. LeCun, What is the best multi-stage architecture for object recognition?, Proc. IEEE Int. Conf. Comput. Vis (2009) 2146–2153.
[48] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167, 2015.
[49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Machine Learn. Res. 15 (1) (2014) 1929–1958.
[50] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747, (2016).
[51] N. Qian, On the momentum term in gradient descent learning algorithms, Neural Networks 12 (1) (1999) 145–151.
[52] J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, J. Mach. Learn. Res. 12 (Jul) (2011) 2121–2159.
[53] M.D. Zeiler, Adadelta: an adaptive learning rate method, arXiv preprint arXiv:1212.5701, 2012.
[54] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
[55] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. In Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2016. pp. 770–778.
[56] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger. Densely connected convolutional networks. IN Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2017. pp. 4700–4708.
[57] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Proc. Adv. Neural Inf. Process. Syst. (2014) 2672–2680.
[58] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502.
[59] W. Fuan, J. Hongkai, S. Haidong, D. Wenjing, W. Shuaipeng, An adaptive deep convolutional neural network for rolling bearing fault diagnosis, Meas. Sci. Technol. 28 (9) (2017) 95005.
[60] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, S. Wu, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, Mech. Syst. Signal Pr. 100 (2018) 743–765.
[61] H. Shao, H. Jiang, H. Zhang, T. Liang, Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network, IEEE Trans. Ind. Electron. 65 (3) (2018) 2727–2736.
[62] S. Wang, J. Xiang, Y. Zhong, Y. Zhou, Convolutional neural network-based hidden Markov models for rolling element bearing fault identification, Knowl.-Based Syst. 144 (2018) 65–76.
[63] W. Gong, H. Chen, Z. Zhang, M. Zhang, R. Wang, C. Guan, Q. Wang, A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion, Sensors-Basel 19 (7) (2019) 1693.
[64] L. Jing, T. Wang, M. Zhao, P. Wang, An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox, Sensors-Basel 17 (2) (2017) 414.
[65] H. Chen, N. Hu, Z. Cheng, L. Zhang, Y. Zhang, A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes, Measurement 146 (2019) 268–278.
[66] T. Han, C. Liu, L. Wu, S. Sarkar, D. Jiang, An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems, Mech. Syst. Signal Pr. 117 (2019) 170–187.
[67] Y. Yang, H. Zheng, Y. Li, M. Xu, Y. Chen, A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network, IAS Trans. 91 (2019) 235–252.
[68] R. Liu, G. Meng, B. Yang, C. Sun, X. Chen, Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine, IEEE Trans. Ind. Inform 13 (3) (2016) 1310–1320.
[69] H. Yang, J. Zhang, L. Chen, H. Zhang, S. Liu, Fault diagnosis of reciprocating compressor based on convolutional neural networks with multisource raw vibration signals, Math. Probl. Eng. 2019 (2019).
[70] M. Xia, T. Li, L. Xu, L. Liu, C.W. de Silva, Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks, IEEE/ASME Trans. Mechatron. 23 (1) (2018) 101–110.
[71] D. Hoang, H. Kang, Rolling element bearing fault diagnosis using convolutional neural network and vibration image, Cogn. Syst. Res 53 (2019) 42–50.

[72] W. Zhang, D. Yang, H. Wang, X. Huang, M. Gidlund, CarNet: a dual correlation method for health perception of rotating machinery, IEEE Sens. J. 19 (16) (2019) 7095–7106.
[73] D.T. Hoang, H.J. Kang, A motor current signal based bearing fault diagnosis using deep learning and information fusion, IEEE Trans. Instrum. Meas. (2019).
[74] H. Wang, S. Li, L. Song, L. Cui, A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals, Comput. Ind. 105 (2019) 182–190.
[75] Z. Hu, Y. Wang, M. Ge, J. Liu, Data-driven fault diagnosis method based on compressed sensing and improved multi-scale network, IEEE Trans. Ind. Electron. 67 (4) (2019) 3216–3225.
[76] H. Wang, S. Li, L. Song, L. Cui, P. Wang, An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network, IEEE Trans. Instrum. Meas. (2019).
[77] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Ind. Electron. 65 (7) (2018) 5990–5998.
[78] Z. Chen, C. Li, R.E.V. Sanchez, Gearbox fault identification and classification with convolutional neural networks, Shock Vib. 2015 (2015).
[79] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, Convolutional neural network based fault detection for rotating machinery, J. Sound Vib. 377 (2016) 331–345.
[80] M. Bhadane, K.I. Ramachandran. Bearing fault identification and classification with convolutional neural network. In: Editor. Pub Place: IEEE; 2017. pp. 1–5.
[81] C. Lu, Z. Wang, B. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification, Adv. Eng. Inf. 32 (2017) 139–151.
[82] S. Li, G. Liu, X. Tang, J. Lu, J. Hu, An ensemble deep convolutional neural network model with improved DS evidence fusion for bearing fault diagnosis, Sensors-Basel 17 (8) (2017) 1729.
[83] V. Tra, J. Kim, S.A. Khan, J. Kim, Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal levenberg-marquardt algorithm, Sensors-Basel 17 (12) (2017) 2834.
[84] A. Prosvirin, J. Kim, J. Kim, Bearing fault diagnosis based on convolutional neural networks with kurtogram representation of acoustic emission signals, in: Advances in Computer Science and Ubiquitous Computing, Springer, 2017, pp. 21–26.
[85] Y. Tian, X. Liu, A deep adaptive learning method for rolling bearing fault diagnosis using immunity, Tsinghua Sci. Technol. 24 (6) (2019) 750–762.
[86] V. Tra, S.A. Khan, J. Kim, Diagnosis of bearing defects under variable speed conditions using energy distribution maps of acoustic emission spectra and convolutional neural networks, J. Acoust. Soc. Am. 144 (4) (2018) L322–L327.
[87] Y. Li, G. Cheng, Y. Pang, M. Kuai, Planetary gear fault diagnosis via feature image extraction based on multi central frequencies and vibration signal frequency Spectrum, Sensors-Basel 18 (6) (2018) 1735.
[88] Y. Yao, H. Wang, S. Li, Z. Liu, G. Gui, Y. Dan, J. Hu, End-to-end convolutional neural network model for gear fault diagnosis based on sound signals, Appl. Sci. 8 (9) (2018) 1584.
[89] B.H. Kien, D. Iba, Y. Ishii, Y. Tsutsui, N. Miura, T. Iizuka, A. Masuda, A. Sone, I. Moriwaki, Crack detection of plastic gears using a convolutional neural network pre-learned from images of meshing vibration data with transfer learning, Forsch. Ingenieurwes. 83 (3) (2019) 645–653.
[90] X. Ding, Q. He, Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis, IEEE Trans. Instrum. Meas. 66 (8) (2017) 1926–1935.
[91] D. Gao, Y. Zhu, X. Wang, K. Yan, J. Hong. A Fault Diagnosis Method of Rolling Bearing Based on Complex Morlet CWT and CNN. In: Editor. Pub Place: IEEE; 2018. pp. 1101–1105.
[92] S. Guo, T. Yang, W. Gao, C. Zhang, Y. Zhang, An intelligent fault diagnosis method for bearings with variable rotating speed based on Pythagorean spatial pyramid pooling CNN, Sensors-Basel 18 (11) (2018) 3857.
[93] G. Xu, M. Liu, Z. Jiang, D.S.O. Ffker, W. Shen, Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning, Sensors-Basel 19 (5) (2019) 1088.
[94] M.M. Islam, J. Kim, Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network, Comput. Ind. 106 (2019) 142–153.
[95] W. Sun, B. Yao, N. Zeng, B. Chen, Y. He, X. Cao, W. He, An intelligent gear fault diagnosis methodology using a complex wavelet enhanced convolutional neural network, Materials 10 (7) (2017) 790.
[96] D. Cabrera, F. Sancho, C. Li, M. Cerrada, R.E.V.S.A. Nchez, F. Pacheco, J.E.V. de Oliveira, Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation, Appl. Soft Comput. 58 (2017) 53–64.
[97] Y. Han, B. Tang, L. Deng, Multi-level wavelet packet fusion in dynamic ensemble convolutional neural network for fault diagnosis, Measurement 127 (2018) 246–255.
[98] J. Grezmak, P. Wang, C. Sun, R.X. Gao, Explainable convolutional neural network for gearbox fault diagnosis, Procedia CIRP 80 (2019) 476–481.
[99] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, Z. Zhang, Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform, Comput. Ind. 113 (2019) 103132.
[100] S. Guo, T. Yang, W. Gao, C. Zhang, A novel fault diagnosis method for rotating machinery based on a convolutional neural network, Sensors-Basel 18 (5) (2018) 1429.
[101] S. Shao, R. Yan, Y. Lu, P. Wang, R. Gao, DCNN-based multi-signal induction motor fault diagnosis, IEEE Trans. Instrum. Meas. (2019).
[102] Y. Hsueh, V.R. Ittangihal, W. Wu, H. Chang, C. Kuo, Fault diagnosis system for induction motors by CNN using empirical wavelet transform, Symmetry 11 (10) (2019) 1212.
[103] Z. Chen, K. Gryllias, W. Li, Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine, Mech. Syst. Signal Pr. 133 (2019) 106272.
[104] X. Cao, B. Chen, B. Yao, W. He, Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification, Comput. Ind. 106 (2019) 71–84.
[105] D. Verstraete, A.E.S. Ferrada, E.L.O.P. Droguett, V. Meruane, M. Modarres, Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings, Shock Vib. 2017 (2017).
[106] V. Pandhare, J. Singh, J. Lee, Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features, IEEE (2019) 320–326.
[107] Y. Xin, S. Li, C. Cheng, J. Wang, An intelligent fault diagnosis method of rotating machinery based on deep neural networks and time-frequency analysis, J. Vibroeng. 20 (6) (2018) 2321–2335.
[108] W. Yu, S. Huang, W. Xiao, Fault diagnosis based on an approach combining a spectrogram and a convolutional neural network with application to a wind turbine system, Energies 11 (10) (2018) 2561.
[109] L. Wen, L. Gao, X. Li, A new snapshot ensemble convolutional neural network for fault diagnosis, IEEE Access 7 (2019) 32037–32047.
[110] L. Wang, X. Zhao, J. Wu, Y. Xie, Y. Zhang, Motor fault diagnosis based on short-time Fourier transform and convolutional neural network, Chinese J. Mech. Eng. 30 (6) (2017) 1357–1368.
[111] G. Li, C. Deng, J. Wu, X. Xu, X. Shao, Y. Wang, Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and S-transform, Sensors-Basel 19 (12) (2019) 2750.
[112] L. Wen, X. Li, L. Gao, A new two-level hierarchical diagnosis network based on convolutional neural network, IEEE Trans. Instrum. Meas. 69 (2) (2020) 330–338.
[113] H. Jeong, S. Park, S. Woo, S. Lee, Rotating machinery diagnostics using deep learning on orbit plot images, Procedia Manuf. 5 (2016) 1107–1118.
[114] N.F. Waziralilah, A. Abu, M.H. Lim, L.K. Quen, A. Elfakarany, Bearing fault diagnosis employing Gabor and augmented architecture of convolutional neural network, J. Mech. Eng. Sci. 13 (3) (2019) 5689–5702.
[115] D. Zhao, T. Wang, F. Chu, Deep convolutional neural network based planet bearing fault classification, Comput. Ind. 107 (2019) 59–66.
[116] O. Janssens, R. Van de Walle, M. Loccufier, S. Van Hoecke, Deep learning for infrared thermal image based machine health monitoring, IEEE/ASME Trans. Mechatron. 23 (1) (2018) 151–159.
[117] Z. Jia, Z. Liu, C. Vong, M. Pecht, A rotating machinery fault diagnosis method based on feature learning of thermal images, IEEE Access 7 (2019) 12348–12359.
[118] Y. Li, J.X. Gu, D. Zhen, M. Xu, A. Ball, An evaluation of gearbox condition monitoring using infrared thermal images applied with convolutional neural networks, Sensors-Basel 19 (9) (2019) 2205.
[119] X. Chen, L. Peng, G. Cheng, C. Luo, Research on degradation state recognition of planetary gear based on multiscale information dimension of SSD and CNN, Complexity 2019 (2019).
[120] F. Wang, G. Deng, L. Ma, X. Liu, H. Li, Convolutional neural network based on spiral arrangement of features and its application in bearing fault diagnosis, IEEE Access 7 (2019) 64092–64100.
[121] X. Zhu, D. Hou, P. Zhou, Z. Han, Y. Yuan, W. Zhou, Q. Yin, Rotor fault diagnosis using a convolutional neural network with symmetrized dot pattern images, Measurement 138 (2019) 526–535.
[122] H. Li, Q. Zhang, X. Qin, Y. Sun, K-SVD based WVD enhancement algorithm for planetary gearbox fault diagnosis under a CNN framework, Meas. Sci. Technol. (2019).
[123] S.S. Udmale, S.S. Patil, V.M. Phalle, S.K. Singh, A bearing vibration data analysis based on spectral kurtosis and ConvNet, Soft. Comput. 23 (19) (2019) 9341–9359.
[124] J.S.L. Senanayaka, H. Van Khang, K.G. Robbersmyr, Multiple classifiers and data fusion for robust diagnosis of gearbox mixed faults, IEEE Trans. Ind. Inform 15 (8) (2019) 4569–4579.
[125] L. Eren, Bearing fault detection by one-dimensional convolutional neural networks, Math. Probl. Eng. 2017 (2017).
[126] H. Pan, X. He, S. Tang, F. Meng, An improved bearing fault diagnosis method using one-dimensional CNN and LSTM, Strojniski Vestnik/J. Mech. Eng. 64 (2018).
[127] J. Pan, Y. Zi, J. Chen, Z. Zhou, B. Wang, LiftingNet: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification, IEEE Trans. Ind. Electron. 65 (6) (2018) 4973–4982.
[128] W. Qian, S. Li, J. Wang, Z. An, X. Jiang, An intelligent fault diagnosis framework for raw vibration signals: adaptive overlapping convolutional neural network, Meas. Sci. Technol. 29 (9) (2018) 95009.
[129] F. Jia, Y. Lei, N. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, Mech. Syst. Signal Pr. 110 (2018) 349–367.
[130] L. Eren, T. Ince, S. Kiranyaz, A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier, J. Signal Process. Syst. 91 (2) (2019) 179–189.

[131] S. Ma, F. Chu, Ensemble deep learning-based fault diagnosis of rotor bearing systems, Comput. Ind. 105 (2019) 143–152.
[132] D. Wang, Q. Guo, Y. Song, S. Gao, Y. Li, Application of multiscale learning neural network based on CNN in bearing fault diagnosis, J. Signal Process. Syst. 91 (10) (2019) 1205–1217.
[133] W. Huang, J. Cheng, Y. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, Neurocomputing 359 (2019) 77–92.
[134] H. Qiao, T. Wang, P. Wang, L. Zhang, M. Xu, An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions, IEEE Access 7 (2019) 118954–118964.
[135] O. Abdeljaber, S. Sassi, O. Avci, S. Kiranyaz, A.A. Ibrahim, M. Gabbouj, Fault detection and severity identification of ball bearings by online condition monitoring, IEEE Trans. Ind. Electron. 66 (10) (2019) 8136–8147.
[136] R. Huang, Y. Liao, S. Zhang, W. Li, Deep decoupling convolutional neural network for intelligent compound fault diagnosis, IEEE Access 7 (2018) 1848–1858.
[137] X. Liu, Q. Zhou, J. Zhao, H. Shen, X. Xiong, Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network, Sensors-Basel 19 (4) (2019) 972.
[138] Y. Han, B. Tang, L. Deng, An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes, Comput. Ind. 107 (2019) 50–58.
[139] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Trans. Ind. Electron. 66 (4) (2019) 3196–3207.
[140] W. Sun, R. Zhao, R. Yan, S. Shao, X. Chen, Convolutional discriminative feature learning for induction motor fault diagnosis, IEEE Trans. Ind. Inform 13 (3) (2017) 1350–1359.
[141] Z. Yuan, L. Zhang, L. Duan, A novel fusion diagnosis method for rotor system fault based on deep learning and multi-sourced heterogeneous monitoring data, Meas. Sci. Technol. 29 (11) (2018) 115005.
[142] S. Afrasiabi, M. Afrasiabi, B. Parang, M. Mohammadi. Real-Time Bearing Fault Diagnosis of Induction Motors with Accelerated Deep Learning Approach. In: Editor. Pub Place: IEEE; 2019. pp. 155–159.
[143] Xie Chen, Yuan, Huang, Li, Research on a real-time monitoring method for the wear state of a tool based on a convolutional bidirectional LSTM model, Symmetry 11 (10) (2019) 1233.
[144] Y. Xie, T. Zhang, Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition, Shock Vib. 2017 (2017).
[145] M. Sadoughi, C. Hu, Physics-based convolutional neural network for fault diagnosis of rolling element bearings, IEEE Sens. J. 19 (11) (2019) 4181–4192.
[146] S. Dong, G. Wen, Z. Zhang, Y. Yuan, J. Luo, Rolling bearing incipient degradation monitoring and performance assessment based on signal component tracking, IEEE Access 7 (2019) 45983–45993.
[147] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1–10.
[148] S. Ma, W. Cai, W. Liu, Z. Shang, G. Liu, A lighted deep convolutional neural network based fault diagnosis of rotating machinery, Sensors-Basel 19 (10) (2019) 2381.
[149] T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-time motor fault detection by 1-D convolutional neural networks, IEEE Trans. Ind. Electron. 63 (11) (2016) 7067–7075.
[150] T. Ince, Real-time broken rotor bar fault detection and classification by shallow 1D convolutional neural networks, Electr. Eng. 101 (2019) 599–608.
[151] T. Khan, P. Alekhya, J. Seshadrinath. Incipient Inter-turn Fault Diagnosis in Induction motors using CNN and LSTM based Methods. In: Editor. Pub Place: IEEE; 2018. pp. 1–6.
[152] I. Kao, W. Wang, Y. Lai, J. Perng, Analysis of permanent magnet synchronous motor fault diagnosis based on learning, IEEE Trans. Instrum. Meas. 68 (2) (2019) 310–324.
[153] J. Wang, P. Fu, L. Zhang, R.X. Gao, R. Zhao, Multi-level information fusion for induction motor fault diagnosis, IEEE/ASME Trans. Mechatron. 24 (5) (2019) 2139–2150.
[154] X. Li, J. Li, Y. Qu, D. He, Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals, Appl. Sci. 9 (4) (2019) 768.
[155] D.K. Appana, A. Prosvirin, J. Kim, Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks, Soft. Comput. 22 (20) (2018) 6719–6729.
[156] M. Zhao, M. Kang, B. Tang, M. Pecht, Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes, IEEE Trans. Ind. Electron. 65 (5) (2018) 4290–4300.
[157] M. Zhao, M. Kang, B. Tang, M. Pecht, Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis, IEEE Trans. Ind. Electron. 66 (6) (2019) 4696–4706.
[158] X. Li, W. Zhang, Q. Ding, J. Sun, Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation, J. Intell. Manuf. 1–20 (2018).
[159] W. Zhang, X. Li, Q. Ding, Deep residual learning-based fault diagnosis method for rotating machinery, ISA T. 95 (2019) 295–305.
[160] D. Peng, Z. Liu, H. Wang, Y. Qin, L. Jia, A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains, IEEE Access 7 (2019) 10278–10293.
[161] S. Ma, W. Liu, W. Cai, Z. Shang, G. Liu, Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions, IEEE Access 7 (2019) 57023–57036.
[162] Z. Zhuang, H. Lv, J. Xu, Z. Huang, W. Qin, A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions, Appl. Sci. 9 (9) (2019) 1823.
[163] L. Su, L. Ma, N. Qin, D. Huang, A.H. Kemp, Fault diagnosis of high-speed train bogie by residual-squeeze net, IEEE Trans. Ind. Inform 15 (7) (2019) 3856–3863.
[164] S. Ma, F. Chu, Q. Han, Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions, Mech. Syst. Signal Pr. 127 (2019) 190–201.
[165] R. Liu, F. Wang, B. Yang, S.J. Qin, Multi-scale Kernel based residual convolutional neural network for motor fault diagnosis under non-stationary conditions, IEEE Trans. Ind. Inform 16 (6) (2020) 3797–3806.
[166] S. Cao, L. Wen, X. Li, L. Gao. Application of Generative Adversarial Networks for Intelligent Fault Diagnosis. In: Editor. Pub Place: IEEE; 2018. pp. 711–715.
[167] Y. Xie, T. Zhang. Imbalanced Learning for Fault Diagnosis Problem of Rotating Machinery Based on Generative Adversarial Networks. In: Editor. Pub Place; 2018. pp. 6017–6022.
[168] S. Shao, P. Wang, R. Yan, Generative adversarial networks for data augmentation in machine fault diagnosis, Comput. Ind. 106 (2019) 85–93.
[169] S. Afrasiabi, M. Afrasiabi, B. Parang, M. Mohammadi, M.M. Arefi, M. Rastegar, Wind turbine fault diagnosis with generative-temporal convolutional neural network, IEEE (2019) 1–5.
[170] Q. Li, L. Chen, C. Shen, B. Yang, Z. Zhu, Enhanced generative adversarial networks for fault diagnosis of rotating machinery with imbalanced data, Meas. Sci. Technol. 30 (11) (2019) 115005.
[171] S. Suh, H. Lee, J. Jo, P. Lukowicz, Y.O. Lee, Generative oversampling method for imbalanced data on bearing fault detection and diagnosis, Applied Sciences 9 (4) (2019) 746.
[172] Q. Guo, Y. Li, Y. Song, D. Wang, W. Chen, Intelligent fault diagnosis method based on full 1D convolutional generative adversarial network, IEEE Trans. Ind. Inform 16 (3) (2020) 2044–2053.
[173] X. Li, X. Li, Y. Qu, D. He, Gear pitting level diagnosis using vibration signals with an improved inception structure, Vibroengineering Procedia 20 (2018) 70–75.
[174] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, S. Liu, ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis, Neurocomputing 294 (2018) 61–71.
[175] Z. Zhu, G. Peng, Y. Chen, H. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing 323 (2019) 62–75.
[176] T. Chen, Z. Wang, X. Yang, K. Jiang, A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals, Measurement 148 (2019) 106857.
[177] Y. Yoo, J. Baek, A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network, Applied Sciences 8 (7) (2018) 1102.
[178] D. Belmiloud, T. Benkedjouh, M. Lachi, A. Laggoun, J.P. Dron, Deep convolutional neural networks for Bearings failure predictionand temperature correlation, Journal of Vibroengineering 20 (8) (2018) 2878–2891.
[179] A.Z. Hinchi, M. Tkiouat, Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network, Procedia Comput. Sci. 127 (2018) 123–132.
[180] L. Guo, Y. Lei, N. Li, T. Yan, N. Li, Machinery health indicator construction based on convolutional neural networks considering trend burr, Neurocomputing 292 (2018) 142–150.
[181] L. Ren, Y. Sun, H. Wang, L. Zhang, Prediction of bearing remaining useful life with deep convolution neural network, IEEE Access 6 (2018) 13041–13049.
[182] D. She, M. Jia, Wear indicator construction of rolling bearings based on multi-channel deep convolutional neural network with exponentially decaying learning rate, Measurement 135 (2019) 368–375.
[183] W. Mao, J. He, J. Tang, Y. Li, Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network, Adv. Mech. Eng. 10 (12) (2018) 754328416.
[184] X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction, Reliab. Eng. Syst. Saf. 182 (2019) 208–218.
[185] J. Zhu, N. Chen, W. Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, IEEE Trans. Ind. Electron. 66 (4) (2019) 3208–3216.
[186] Q. Wang, B. Zhao, H. Ma, J. Chang, G. Mao, A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion, J. Mech. Sci. Technol. (2019) 1–11.
[187] D. Zhang, E. Stewart, J. Ye, M. Entezami, C. Roberts, Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions, IEEE Trans. Instrum. Meas. (2019).

[188] B. Yang, R. Liu, E. Zio, Remaining useful life prediction based on a double-convolutional neural network architecture, IEEE Trans. Ind. Electron. 66 (12) (2019) 9521–9530.
[189] W. Peng, Z. Ye, N. Chen, Bayesian deep learning based health prognostics towards prognostics uncertainty, IEEE Trans. Ind. Electron. 67 (3) (2020) 2283–2293.
[190] Q. Yao, T. Yang, Z. Liu, Z. Zheng. Remaining Useful Life Estimation by Empirical Mode Decomposition and Ensemble Deep Convolution Neural Networks. In: Editor. Pub Place: IEEE; 2019. pp. 1–6.
[191] Z. Wang, H. Ma, H. Chen, B. Yan, X. Chu, Performance degradation assessment of rolling bearing based on convolutional neural network and deep long-short term memory network, Int. J. Prod. Res. 1–13 (2019).
[192] R. Liu, B. Yang, A.G. Hauptmann, Simultaneous bearing fault recognition and remaining useful life prediction using joint loss convolutional neural network, IEEE Trans. Ind. Inform 16 (1) (2020) 87–96.
[193] B. Wang, Y. Lei, N. Li, T. Yan, Deep separable convolutional network for remaining useful life prediction of machinery, Mech. Syst. Signal Pr. 134 (2019) 106330.
[194] B. Wang, Y. Lei, T. Yan, N. Li, L. Guo, Recurrent convolutional neural network: a new framework for remaining useful life prediction of machinery, Neurocomputing 379 (2020) 117–129.
[195] G.S. Babu, P. Zhao, X. Li. Deep convolutional neural network based regression approach for estimation of remaining useful life. In: Editor. Pub Place: Springer; 2016. pp. 214–228.
[196] X. Li, Q. Ding, J. Sun, Remaining useful life estimation in prognostics using deep convolution neural networks, Reliab. Eng. Syst. Saf. 172 (2018) 1–11.
[197] L. Wen, Y. Dong, L. Gao, A new ensemble residual convolutional neural network for remaining useful life estimation, Math. Biosci. Eng 16 (2019) 862–880.
[198] J. Li, X. Li, D. He, A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction, IEEE Access 7 (2019) 75464–75475.
[199] A. Al-Dulaimi, S. Zabihi, A. Asif, A. Mohammadi, A multimodal and hybrid deep neural network model for Remaining Useful Life estimation, Comput. Ind. 108 (2019) 186–196.
[200] A. Ruiz-Tagle Palazuelos, E.L. Droguett, R. Pascual, A novel deep capsule neural network for remaining useful life estimation, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, (2019) 1748006X-1986654X.
[201] Z. Kong, Y. Cui, Z. Xia, H. Lv, Convolution and long short-term memory hybrid deep neural networks for remaining useful life prognostics, Applied Sciences 9 (19) (2019) 4156.
[202] R. Zhao, R. Yan, J. Wang, K. Mao, Learning to monitor machine health with convolutional bi-directional LSTM networks, Sensors-Basel 17 (2) (2017) 273.
[203] H. Qiao, T. Wang, P. Wang, S. Qiao, L. Zhang, A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series, Sensors-Basel 18 (9) (2018) 2932.
[204] F. Aghazadeh, A. Tahan, M. Thomas, Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process, Int. J. Adv. Manuf. Technol. 98 (9–12) (2018) 3217–3227.
[205] Z. Huang, J. Zhu, J. Lei, X. Li, F. Tian, Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations, J. Intell. Manuf. 31 (2020) 953–966.
[206] J. Fu, J. Chu, P. Guo, Z. Chen, Condition monitoring of wind turbine gearbox bearing based on deep learning model, IEEE Access 7 (2019) 57078–57087.
[207] Z. Kong, B. Tang, L. Deng, W. Liu, Y. Han, Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units, Renew. Energy 146 (2020) 760–768.
[208] H. Luo, M. Huang, Z. Zhou, A dual-tree complex wavelet enhanced convolutional LSTM neural network for structural health monitoring of automotive suspension, Measurement 137 (2019) 14–27.
[209] P. Li, X. Jia, J. Feng, F. Zhu, M. Miller, L. Chen, J. Lee, A novel scalable method for machine degradation assessment using deep convolutional neural network, Measurement (2019), 107106.
[210] P. Cao, S. Zhang, J. Tang, Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning, IEEE Access 6 (2018) 26241–26253.
[211] M.J. Hasan, M. Sohaib, J. Kim. 1D CNN-Based Transfer Learning Model for Bearing Fault Diagnosis under Variable Working Conditions. In: Editor. Pub Place: Springer; 2018. pp. 13-23.
[212] M. Hemmer, H. Van Khang, K. Robbersmyr, T. Waag, T. Meyer, Fault classification of axial and radial roller bearings using transfer learning through a pretrained convolutional neural network, Designs 2 (4) (2018) 56.
[213] S. Zhong, S. Fu, L. Lin, A novel gas turbine fault diagnosis method based on transfer learning with CNN, Measurement 137 (2019) 435–453.
[214] L. Wen, X. Li, L. Gao, A transfer convolutional neural network for fault diagnosis based on ResNet-50, Neural Comput. Appl. 32 (2020) 6111–6124.
[215] L. Wen, L. Gao, Y. Dong, Z. Zhu, A negative correlation ensemble transfer learning method for fault diagnosis based on convolutional neural network, Math. Biosci. Eng. 16 (5) (2019) 3311–3330.
[216] T. Han, C. Liu, W. Yang, D. Jiang, Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions, ISA Trans. 93 (2019) 341–353.
[217] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inform 15 (4) (2019) 2446–2455.
[218] P. Ma, H. Zhang, W. Fan, C. Wang, G. Wen, X. Zhang, A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network, Meas. Sci. Technol. 30 (5) (2019) 55402.
[219] M.J. Hasan, M.M. Islam, J. Kim, Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions, Measurement 138 (2019) 620–631.
[220] Z. Chen, K. Gryllias, W. Li, Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network, IEEE Trans. Ind. Inform 16 (1) (2020) 339–349.
[221] K.M. Borgwardt, A. Gretton, M.J. Rasch, H. Kriegel, B. Sch, O. Lkopf, A.J. Smola, Integrating structured biological data by kernel maximum mean discrepancy, Bioinformatics 22 (14) (2006) e49–e57.
[222] X. Jia, M. Zhao, Y. Di, Q. Yang, J. Lee, Assessment of data suitability for machine prognosis using maximum mean discrepancy, IEEE Trans. Ind. Electron. 65 (7) (2018) 5872–5881.
[223] B. Sun, K. Saenko, Deep coral: correlation alignment for deep domain adaptation, Proc. Eur. Conf. Comput. Vision (2016) 443–450.
[224] B. Zhang, W. Li, X. Li, S. Ng, Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks, IEEE Access 6 (2018) 66367–66384.
[225] X. Li, W. Zhang, Q. Ding, A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning, Neurocomputing 310 (2018) 77–95.
[226] X. Li, W. Zhang, Q. Ding, J. Sun, Multi-Layer domain adaptation method for rolling bearing fault diagnosis, Signal Process. 157 (2019) 180–197.
[227] D. Xiao, Y. Huang, L. Zhao, C. Qin, H. Shi, C. Liu, Domain adaptive motor fault diagnosis using deep transfer learning, IEEE Access 7 (2019) 80937–80949.
[228] B. Yang, Y. Lei, F. Jia, S. Xing, An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings, Mech. Syst. Signal Pr. 122 (2019) 692–706.
[229] T. Han, C. Liu, W. Yang, D. Jiang, Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application, ISA Trans. (2019).
[230] K. Xu, S. Li, J. Wang, Z. An, W. Qian, H. Ma, A novel convolutional transfer feature discrimination network for imbalanced fault diagnosis under variable rotational speed, Meas. Sci. Technol. (2019).
[231] J. Zhu, N. Chen, C. Shen, A new deep transfer learning method for bearing fault diagnosis under different working conditions, IEEE Sens. J. (2019).
[232] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. 66 (7) (2019) 5525–5534.
[233] K. Xu, S. Li, X. Jiang, Z. An, J. Wang, T. Yu, A renewable fusion fault diagnosis network for the variable speed conditions under unbalanced samples, Neurocomputing 379 (2020) 12–29.
[234] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F.C.C.O. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Machine Learn. Res. 17 (1) (2016) 2030–2096.
[235] L. Guo, Y. Lei, S. Xing, T. Yan, N. Li, Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data, IEEE Trans. Ind. Electron. 66 (9) (2019) 7316–7325.
[236] J. Jiao, M. Zhao, J. Lin, K. Liang, Residual joint adaptation adversarial network for intelligent transfer fault diagnosis, Mech. Syst. Signal Pr. 145 (2020) 106962.
[237] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, B. Liang, A deep transfer model with wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions, IEEE Access 7 (2019) 65303–65318.
[238] X. Wang, F. Liu, Triplet loss guided adversarial domain adaptation for bearing fault diagnosis, Sensors-Basel 20 (1) (2020) 320.
[239] Y. Xie, T. Zhang, A transfer learning strategy for rotation machinery fault diagnosis based on cycle-consistent generative adversarial networks, IEEE (2018) 1309–1313.
[240] J. Jiao, M. Zhao, J. Lin, Unsupervised adversarial adaptation network for intelligent fault diagnosis, IEEE Trans. Ind. Electron. (2019).
[241] J. Jiao, M. Zhao, J. Lin, C. Ding, Classifier inconsistency based domain adaptation network for partial transfer intelligent diagnosis, IEEE Trans. Ind Inform. (2019).
[242] L. Duan, X. Wang, M. Xie, Z. Yuan, J. Wang, Auxiliary-model-based domain adaptation for reciprocating compressor diagnosis under variable conditions, J. Intell. Fuzzy Syst. 34 (6) (2018) 3595–3604.
[243] L. Zhang, L. Guo, H. Gao, D. Dong, G. Fu, X. Hong, Instance-based ensemble deep transfer learning network: A new intelligent degradation recognition method and its application on ball screw, Mech. Syst. Signal Pr. 140 (2020) 106681.

[244] D. Xiao, Y. Huang, C. Qin, Z. Liu, Y. Li, C. Liu, Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis, Proc. Instit. Mech. Eng., Part C (2019) 62159741.



Jinyang Jiao received the B.S. degree in vehicle engineering from China University of Petroleum, Qingdao, China, in 2015. He is currently working toward the Ph.D. degree in mechanical engineering in the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include deep learning, transfer learning, and intelligent fault diagnosis of machinery.

Ming Zhao received his B.S., M.S. and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2006, 2009 and 2013, respectively. He is currently an associate professor with School of Mechanical Engineering, Xi'an Jiaotong University. From August 2015 to August 2016, he was a visiting scholar at the National Science Foundation Center for Intelligent Maintenance Systems, University of Cincinnati, OH, USA. His research interests include non-stationary signal processing, health monitoring of rotating machinery.

Jing Lin received his B.S., M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1993, 1996 and 1999, respectively, all in mechanical engineering. He is currently a Professor with the Science & Technology on Reliability and Environmental Engineering Laboratory, School of Reliability and Systems Engineering, Beihang University, Beijing 100083, China. From July 2001 to August 2003, he was a Postdoctoral Fellow with the University of Alberta, Edmonton, AB, Canada, and a Research Associate with the University of Wisconsin–Milwaukee, Milwaukee, WI, USA. From September 2003 to December 2008, he was a Research Scientist with the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, under the sponsorship of the Hundred Talents Program. His current research directions are in mechanical system reliability, fault diagnosis, and wavelet analysis. Dr. Lin was a recipient of the National Science Fund for Distinguished Young Scholars in 2011.

Kaixuan Liang received the B.S. degree in Aircraft Manufacturing Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016. He is currently working toward the Ph.D. degree in mechanical engineering in the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include sparse presentation and fault feature enhancement for the monitoring of rotating machinery.