Mechanical Systems and Signal Processing 143 (2020) 106825

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery

Xiang Li [a,b,c], Xu Li [d], Hui Ma [e,b,*]

[a] College of Sciences, Northeastern University, Shenyang 110819, China
[b] Key Laboratory of Vibration and Control of Aero-Propulsion System Ministry of Education, Northeastern University, Shenyang 110819, China
[c] Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati 45221, USA
[d] State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, China
[e] School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Article history:
Received 13 April 2019
Received in revised form 20 January 2020
Accepted 16 March 2020
Available online 24 March 2020

Keywords:
Fault diagnosis
Deep learning
Unsupervised learning
Weakly supervised learning
Clustering

ABSTRACT

Despite the recent advances on intelligent data-driven machinery fault diagnostics, large amounts of high-quality supervised data are mostly required for model training. However, it is usually difficult and expensive to collect sufficient labeled data in real industries, and the difficulty in data preparation significantly hinders the application of the intelligent diagnostic methods. In order to address the data sparsity issue with insufficient labeled data, a deep learning-based fault diagnosis method is proposed in this study, exploring additional unsupervised data which are generally easy for collection. A three-stage training scheme is adopted, i.e. pre-training, representation clustering and enhanced supervised learning. The auto-encoder structure is used for feature extraction, and distance metric learning and k-means clustering method are integrated in the neural network architecture for unsupervised learning. Two rotating machinery datasets are used for validations. The proposed method not only achieves promising diagnostic performance on the semi-supervised learning tasks with few labeled data, but also is well suited for pure unsupervised learning problems. The experimental results suggest the proposed method offers a promising approach on exploiting unsupervised data for fault diagnostics.

1. Introduction

In modern industries, condition monitoring and health management of rotating machines are of great importance. Timely and accurate machinery fault diagnosis has always been highly demanded, which is able to reduce maintenance cost, increase operating safety and enhance machine reliability [1,2]. In the literature, the conventional fault diagnosis methods based on physical models and signal processing techniques have been well established [3–6]. However, with the rapid development of rotating machines especially in the high-end industries such as intelligent manufacturing, aero-space industries *etc.*, the conventional approaches are becoming less capable of precisely diagnosing machinery faults in complex mechanical systems. In the recent years, benefited from the advances in sensing technology and computing algorithm, intelligent data-driven fault diagnosis methods have been attracting increasing attention due to the great merits of high accuracy, fast response and low requirement for special expertise. Intuitive relationship between collected raw data and machine health condition can be built, that largely facilitates practical applications of the data-driven fault diagnosis methods in real industries [7,8].

* Corresponding author at: School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China.
E-mail address: huima@me.neu.edu.cn (H. Ma).

https://doi.org/10.1016/j.ymssp.2020.106825
0888-3270/

Recently, the neural network-based approaches have been successfully and widely developed, and promising diagnostic results have been obtained. Besides the basic multi-layer perceptron (MLP) architecture, the variants such as convolutional neural networks (CNN) [9–12] and recurrent neural networks (RNN) [13] have also been popularly used for machinery fault diagnostics. A two-stage learning approach for intelligent fault diagnosis was proposed by Lei et al. [14], where sparse filtering and a two-layer neural network are adopted for feature extraction, and softmax regression is used for machine health condition classification. Sparse auto-encoder architecture was proposed by Sun et al. [15] in unsupervised deep neural network for induction motor fault diagnosis. Partial corruption of the input data is added to enhance the robustness of the learned feature representation. Lu et al. [16] utilized a stacked denoising auto-encoder for machinery fault identifications with signals containing environmental noise and working condition fluctuations. The recent advances in deep learning-based fault diagnostics have demonstrated the effectiveness of the emerging data-driven algorithm on processing high-dimensional machinery signals, and the conventional diagnostic tasks with sufficient supervised training data have been well addressed [17,18].

Despite the promising diagnostic performance of the intelligent methods reported in the literature, one of the main limitations in the current studies lies in the requirement of the availability of sufficient supervised data for training. In the real industrial scenarios, considering the system uncertainty and labor cost, it is usually difficult and expensive to accurately collect data under certain machine health conditions, especially in the severe faulty states. Therefore, the labeled high-quality machinery data are scarce in most cases. Consequently, by exploring the limited supervised data, overfitting phenomenon can easily occur using the latest data-driven fault diagnosis methods with strong model capabilities, and the model performance on the testing data significantly deteriorates. As shown in Fig. 1, the data sparsity issue remarkably affects the generalization performance of the intelligent diagnostic models, and thus hinders the application in real industries.

In order to address the data sparsity problem, different data augmentation methods have been proposed, aiming to expand the limited supervised dataset by creating additional samples based on the real data. For instance, the synthetic over-sampling approaches have been popular in data augmentations, where new samples can be generated through interpolations of the real samples [19,20]. Transformations of the available data have also been used for data augmentation,

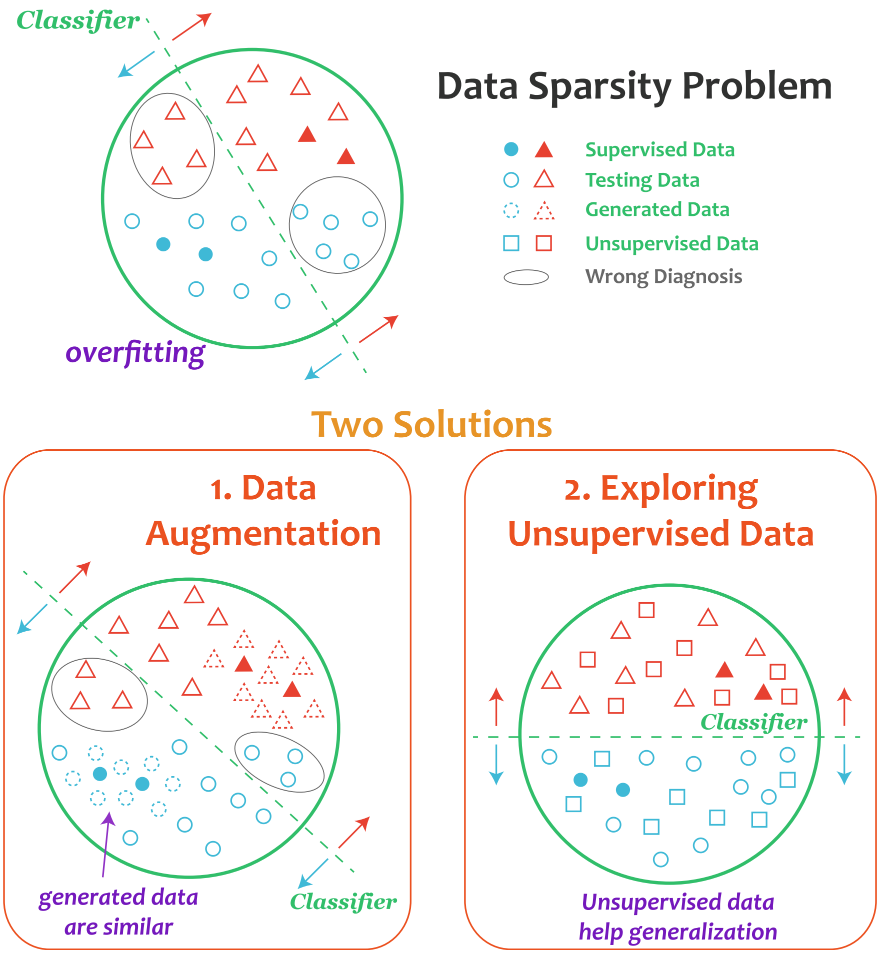

Fig. 1. Data sparsity problem in fault diagnostics, and two methods to address this issue, i.e. data augmentation and exploring unsupervised data.

resulting in additional data variants to assist model training. Li et al. [21] proposed a data augmentation method using different signal processing techniques such as additional noise, time stretching *etc*. The results show the deep learning-based fault diagnosis methods can largely benefit from the expanded dataset with more generated labeled instances. With respect to the data unbalance problem which is similar with data sparsity, Zhang et al. [22] proposed an improved synthetic over-sampling approach called weighted minority over-sampling for data augmentation of the minority classes. A deep auto-encoder and a decision tree were employed afterwards for fault diagnostics. Bunkhumpornpat et al. proposed the Safe-Level-SMOTE method [23], where each minority class instance is assigned a safe level, and the popular synthetic minority oversampling technique (SMOTE) is implemented to generate new data accordingly. While the training dataset can be explicitly enlarged using data augmentations, the new samples are generally similar with the real samples, lacking sufficient diversity of data, and limited improvement in model generalization can be usually achieved [21,24–26].

In fault diagnostic tasks, it is noted that despite the difficulty in collection of labeled data, unsupervised data without precise machine health condition information can be easily measured in general [27,28]. Instead of the artificially generated samples from data augmentation, the unsupervised data are collected under real operating conditions, thus holding stronger validity and diversity. The data-driven fault diagnosis methods are expected to benefit from the exploitation of unsupervised data, and stronger model generalization ability can be achieved. The data-driven tasks with explorations of both supervised and unsupervised data are also known as semi-supervised learning problems [29–31].

The semi-supervised learning task in fault diagnosis has been receiving growing research interests in the past years [32]. Razavi-Far et al. [33] developed a semi-supervised deep learning scheme for diagnosing multiple defects, which contains two modules of information fusion and decision making. Human interaction during the training can be minimized in the learning framework, and the diagnostic efficiency can be improved. A hybrid data-driven scheme for diagnosing faults was proposed in [34], where the training data are mostly unsupervised, and the raw vibration signals are progressively transformed into efficient features for decision making. An unsupervised learning method called general normalized sparse filtering was proposed by Zhang et al. [35], where small amount of training samples are used for feature extraction and fault diagnosis. The effectiveness was validated through experiments on rolling bearings and planetary gears. Liu et al. [27] proposed a categorical adversarial auto-encoder for unsupervised fault diagnosis of rolling bearings. The auto-encoder is optimized using adversarial training scheme, and a prior distribution on the latent space is imposed for better clustering of samples.

In this paper, a deep learning-based fault diagnosis method is proposed for rotating machines with availabilities of few supervised data and sufficient unsupervised data. An auto-encoder architecture is firstly used to learn high-level features, and a deep representation clustering method is adopted for unsupervised learning. Distance metric learning and k-means clustering approach are integrated in the proposed framework. Pseudo labels are attached on the unsupervised data, which are further utilized for training the enhanced fault diagnosis model. Experiments on two rotating machinery datasets are implemented to validate the proposed method, and promising results are obtained in different semi-supervised and unsupervised learning tasks.

The proposed method is able to achieve strong generalization ability against variations of machine operating conditions, and the results are competitive with the existing related studies. When the training and testing data are subject to different distributions, e.g. data are collected under different rotating speeds, transfer learning techniques have been popularly used [36–39]. Specifically, domain adaptation methods have been much preferred in cross-domain fault diagnosis studies [40,41], which aim to extract domain-invariant features in the presence of domain shift phenomenon [38,42–44]. Lu et al. [45] minimized the maximum mean discrepancy (MMD) between domains and adopted a weight term in the network optimization objective in the cross-domain fault diagnosis problems.

This paper proposes a deep learning-based method for rotating machinery fault diagnostics on semi-supervised learning problems, which can be also extended to unsupervised learning tasks [46]. Different from most existing studies, the raw frequency-domain information is used as model input, rather than the hand-crafted features, that largely facilitates potential applications. The proposed method can achieve promising diagnostic performance with weakly supervised data, relaxing the requirement for large amounts of high-quality labeled samples by most data-driven approaches. Strong robustness against variations of operating conditions can be also obtained through unsupervised learning, performing better exploitation of the unlabeled data in real industrial scenarios.

The remainder of this paper starts with the proposed fault diagnosis method in Section 2. Experiments are carried out for validation and investigation in Section 3. We close the paper with conclusions in Section 4.

2. Proposed method

2.1. Problem formulation

The semi-supervised learning problem in fault diagnostics is investigated in this study, where it is assumed that limited supervised data and sufficient unsupervised data are available. Let $\mathcal{D}_{sp} = \{(\mathbf{x}_i^{sp}, y_i^{sp})\}_{i=1}^{n_{sp}}$ denote the small set of the supervised data, where $\mathbf{x}_i^{sp} \in \mathbb{R}^{N_{input}}$ represents the sample data of N_{input} dimensions, y_i^{sp} is the corresponding machine health condition label, and n_{sp} is the number of the samples. Correspondingly, $\mathcal{D}_{un} = \{\mathbf{x}_i^{un}\}_{i=1}^{n_{un}}$ denotes the set of the unsupervised data, where $\mathbf{x}_i^{un} \in \mathbb{R}^{\mathbb{N}}_{\mathbf{input}}$ is the sample data, and n_{un} is the number of the unlabeled samples.

This paper aims to build a fault diagnostic classifier f based on $\mathcal{D}_{sp}$ and $\mathcal{D}_{un}$, which holds strong generalization ability on the testing dataset $\mathcal{D}_{test} = \left\{\left(\mathbf{x}_i^{test}, y_i^{test}\right)\right\}_{i=1}^{n_{test}}$ where $\mathbf{x}_i^{test}$ denotes the sample data, y_i^{test} is the corresponding label and n_{test} is the number of the testing samples. In order to evaluate the robustness of the proposed method against variations of operating conditions, the data spaces in $\mathcal{D}_{sp}, \mathcal{D}_{un}$ and $\mathcal{D}_{test}$ can be different with each other, while the label spaces are usually the same in the fault diagnosis problems, i.e. the machine health conditions are identical in different datasets.

2.2. Overview

Fig. 2 shows the overview of the proposed method. First, the raw machinery vibration acceleration data are collected by sensors. By using fast Fourier transformation (FFT), the frequency-domain information is used as the sample data. In general, three stages are used in the proposed fault diagnosis scheme, i.e. pre-training, representation clustering and enhanced supervised learning. In the first pre-training stage, high-level representations of the supervised and unsupervised data are learned in an unsupervised approach, and the auto-encoder neural network is used. Next, the k-means clustering method is integrated in the framework, taking the high-level data representations as inputs. Through optimizations of the features using distance metric learning, good clustering effects can be obtained, and pseudo labels can be assigned to the clustered unlabeled samples. In this way, the labeled dataset can be significantly enlarged, that can be used to train a discriminative model afterwards.

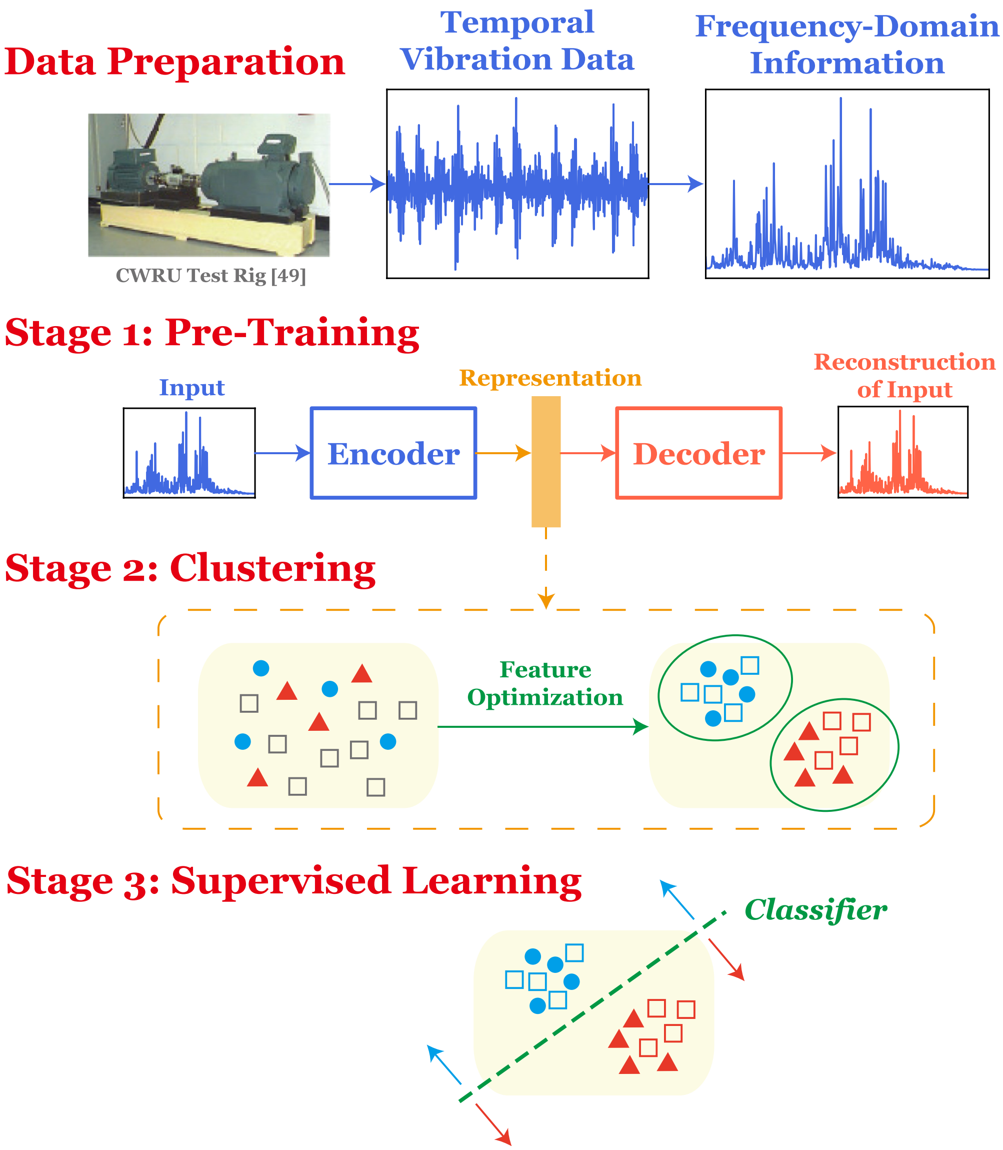

Fig. 2. Overview of the proposed fault diagnosis method.

2.3. Stage 1: Pre-training

First, the proposed neural network model is pre-trained to obtain the high-level representations of data using auto-encoder structure. In general, two modules are adopted in the auto-encoder, i.e. encoder and decoder. For an input sample data $\mathbf{x}$, an encoding function f_{en} with parameters $\boldsymbol{\theta}_{en}$ is used for extraction of features $\mathbf{x}_f = f_{en}(\mathbf{x})$. The decoder module with decoding function f_{de} aims to reconstruct the original data $\mathbf{x}$ based on the extracted features $\mathbf{x}_f$, and the parameters of f_{de} are denoted as $\boldsymbol{\theta}_{de}$. In summary, the auto-encoder model f_{ae} can be formulated as,

$$\begin{aligned} \hat{\mathbf{x}} &= f_{ae}(\mathbf{x}) = f_{de}(f_{en}(\mathbf{x})), \\ \mathbf{x}_f &= f_{en}(\mathbf{x}), \\ \hat{\mathbf{x}} &= f_{de}(\mathbf{x}_f), \end{aligned} \tag{1}$$

where $\hat{\mathbf{x}}$ denotes the output of auto-encoder.

The neural network architecture of the auto-encoder is presented in Fig. 3. In this study, the convolutional neural networks (CNNs) [18,47] are mostly adopted for data processing.

Specifically, three convolutional layers are used in the encoder module, whose filter numbers are 32, 16 and 8, respectively. The dropout technique is used after each layer for regularization. After the flatten layer, a fully-connected layer with N_{rep} neurons is adopted. The structure of the decoder is similar with that of the encoder, and three convolutional layers with 32, 16 and 8 filters are used. A fully-connected layer with N_{input} neurons are adopted after flattening, representing the reconstructed data sample. The model parameters $\boldsymbol{\theta}_{en}$ and $\boldsymbol{\theta}_{de}$ can be optimized through minimizing the reconstruction error L_{ae},

$$\begin{aligned} L_{ae} &= \frac{1}{N_{input}} \sum_{i=1}^{n_{pre}} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_1, \\ \hat{\boldsymbol{\theta}}_{en}, \hat{\boldsymbol{\theta}}_{de} &= \underset{\boldsymbol{\theta}_{en}, \boldsymbol{\theta}_{de}}{\arg\min} \; L_{ae}(\boldsymbol{\theta}_{en}, \boldsymbol{\theta}_{de}) \end{aligned} \tag{2}$$

where n_{pre} denotes the number of the samples in the pre-training stage, generally including both the supervised and unsupervised training data. $\hat{\boldsymbol{\theta}}_{en}$ and $\hat{\boldsymbol{\theta}}_{de}$ denote the optimal values of $\boldsymbol{\theta}_{en}$ and $\boldsymbol{\theta}_{de}$ respectively. The L_1 distance error of reconstruction is considered to enhance model robustness against noise. The optimization problem can be readily solved using the stochastic gradient descent (SGD) algorithm, and the parameters can be optimized as,

$$\begin{aligned} \boldsymbol{\theta}_{en} &\leftarrow \boldsymbol{\theta}_{en} - \delta \frac{\partial L_{ae}}{\partial \boldsymbol{\theta}_{en}}, \\ \boldsymbol{\theta}_{de} &\leftarrow \boldsymbol{\theta}_{de} - \delta \frac{\partial L_{ae}}{\partial \boldsymbol{\theta}_{de}}, \end{aligned} \tag{3}$$

where δ denotes the learning rate. In this way, the learned features $\mathbf{x}_f$ are thus the high-level representations of $\mathbf{x}$.

2.4. Stage 2: Representation clustering

After network pre-training, a deep representation clustering method is proposed for semi-supervised and unsupervised learning. The k-means clustering approach is integrated in the deep learning framework, and the distance metric learning is used for feature optimization. The k-means approach takes a set of samples as inputs, and clusters into several distinct groups based on a geometric criterion [48]. In this study, the extracted features by the encoder are fed into the k-means method, and each group is represented by a centroid $\{\mu_k\}_{k=1}^{N_{cluster}}$ where $N_{cluster}$ denotes the number of the groups. Recall the supervised data $\mathcal{D}_{sp} = \left\{\left(\mathbf{x}_i^{sp}, y_i^{sp}\right)\right\}_{i=1}^{n_{sp}}$ and unsupervised data $\mathcal{D}_{un} = \left\{\mathbf{x}_i^{un}\right\}_{i=1}^{n_{un}}$ are available for training. Based on the pre-trained encoder, the high-level representations of $\mathbf{x}_i^{sp}$ can be obtained as $f_{en}(\mathbf{x}_i^{sp})$, which are used to determine the initial positions of the centroids in k-means clustering.

Let N_{class} denote the number of the machine health conditions. It should be noted that in semi-supervised learning, N_{class} is known to the users and $N_{cluster}$ can be simply set equal to N_{class} for corresponding clustering. Furthermore, the proposed method is also well suited for unsupervised learning where no machine condition information is available, and N_{class} is thus assumed to be unknown to the users. In unsupervised learning, $N_{cluster}$ and N_{class} can be different from each other. In the following, it is assumed $N_{cluster} = N_{class}$ for simplicity.

At this stage, an iteration scheme in optimization is proposed to obtain the pseudo labels of the unsupervised data. In each iteration, which is denoted as an episode in this study, (1) the centroids are firstly initialized, (2) the k-means clustering approach is implemented, and (3) the distance metric learning method is adopted afterwards for updating the parameters of the feature extractor, i.e. the encoder.

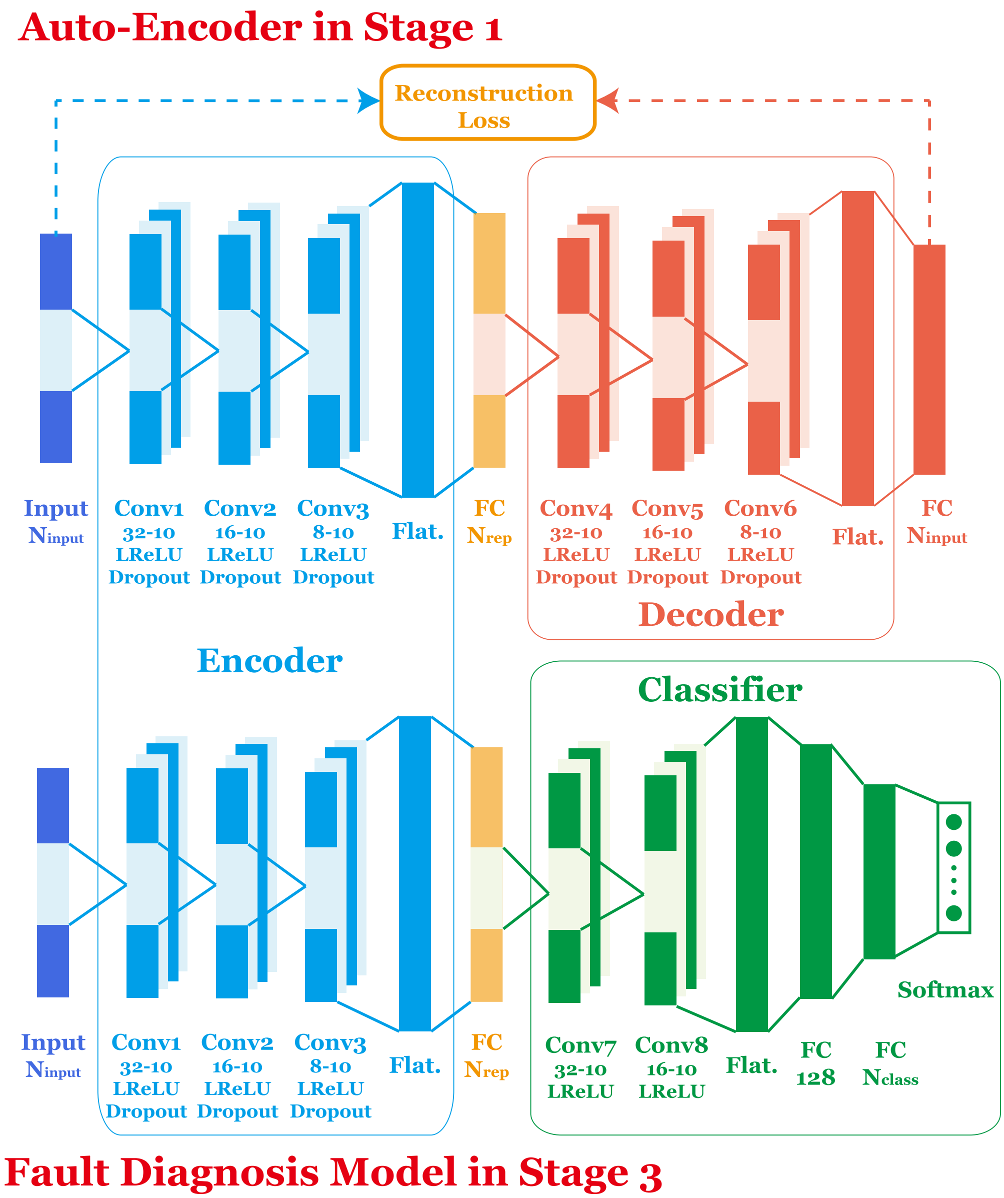

Fig. 3. Architectures of the auto-encoder in Stage 1 and the fault diagnosis model in Stage 3.

(1) Initialization

To initialize the positions of the centroids,

$$\begin{aligned}
&\mu_k = \bar{\mathbf{x}}_f^{sp,k}, \\
&\mathbf{x}_f^{sp,k} = f_{en}(\mathbf{x}_i^{sp} | y_i^{sp} = k), \\
&k = 1, 2, \ldots, N_{cluster},
\end{aligned} \tag{4}$$

where $\mathbf{x}_f^{sp,k}$ denotes the high-level representation set of the labeled samples belonging to the k-th class, and $\bar{\mathbf{x}}_f^{sp,k}$ is the mean of the vectors in $\mathbf{x}_f^{sp,k}$. Eq. (4) suggests the centroids are initialized as the means of the supervised data in each class. In unsupervised learning tasks, the centroids can be randomly initialized instead.

(2) K-means clustering

Next, the k-means clustering method is implemented on the extracted features of the unsupervised data $f_{en}(\mathbf{x}_i^{un})$, which are clustered into $N_{cluster}$ groups. The labels of the updated centroids are in accordance with the initialized labels of the supervised data despite the fluctuations in optimization, and the assignments of the unlabeled samples to different centroids are thus used as the pseudo labels.

(3) Distance metric learning

Afterwards, the distance metric learning is used for feature optimization. Specifically, regularizations are applied on the learned features of the encoder, where the distances of intra-class variations are minimized and the distances of inter-class variations are maximized. By using the L_1 distance, the intra-class metric L_{intra} measuring the clustering compactness can be defined as

$$L_{intra} = \frac{1}{N_{cluster}N_{rep}} \sum_{k=1}^{N_{cluster}} \frac{1}{n_k^{un}} \sum_{i=1}^{n_k^{un}} \|f_{en}(\mathbf{x}_i^{un}|\hat{y}_i^{un} = k) - \mu_k\|_1, \tag{5}$$

where $\hat{y}_i^{un}$ denotes the pseudo label of the unsupervised data $\mathbf{x}_i^{un}$, and n_k^{un} is the number of the samples assigned to the k-th group. The inter-class metric L_{inter} measuring the separability can be defined using the centroid vectors as,

$$L_{inter} = \frac{1}{N_{rep}} \sum_{i=1}^{N_{cluster}-1} \sum_{j=i+1}^{N_{cluster}} \left\|\mu_i - \mu_j\right\|_1. \tag{6}$$

While learning the distance metric, the reconstruction loss of the auto-encoder structure is also considered to avoid information loss. In summary, the optimization objective at this stage $L_{cluster}$ can be formulated as,

$$L_{cluster} = \alpha_{ae}L_{ae} + \alpha_{intra}L_{intra} - \alpha_{inter}L_{inter}, \tag{7}$$

where $\alpha_{ae} > 0, \alpha_{intra} > 0$ and $\alpha_{inter} > 0$ denote the penalty coefficients for L_{ae}, L_{intra} and L_{inter} respectively. The encoder module is optimized to achieve,

$$\hat{\boldsymbol{\theta}}_{en}, \hat{\boldsymbol{\theta}}_{de} = \arg\min_{\boldsymbol{\theta}_{en},\boldsymbol{\theta}_{de}} L_{cluster}(\theta_{en}, \boldsymbol{\theta}_{de}). \tag{8}$$

The parameters can be updated as,

$$\begin{aligned} \boldsymbol{\theta}_{en} &\leftarrow \boldsymbol{\theta}_{en} - \delta(\alpha_{ae}\frac{\partial L_{ae}}{\partial \boldsymbol{\theta}_{en}} + \alpha_{intra}\frac{\partial L_{intra}}{\partial \boldsymbol{\theta}_{en}} - \alpha_{inter}\frac{\partial L_{inter}}{\partial \boldsymbol{\theta}_{en}}), \\ \boldsymbol{\theta}_{de} &\leftarrow \boldsymbol{\theta}_{de} - \delta\alpha_{ae}\frac{\partial L_{ae}}{\partial \boldsymbol{\theta}_{de}}. \end{aligned} \tag{9}$$

The three steps are sequentially implemented in each episode, and multiple episodes are carried out at this stage. In addition, it is noted that neural networks have high possibility of data overfitting [21]. Therefore, in the proposed method, if the unsupervised data sample is incorrectly assigned to a class centroid using k-means, its pseudo label can be wrong throughout distance metric learning.

In order to address this issue, a noise augmentation technique is proposed, where additional noise is added to the unsupervised data in each training episode. Data variants can be obtained in each optimization iteration, which alleviate the underlying overfitting risk. Specifically, additional Gaussian noises are considered, and the noisy data are generated based on different signal-to-noise ratio (SNR),

$$\mathrm{SNR(dB)} = 10\log_{10}(P_{signal}/P_{noise}), \tag{10}$$

where P_{signal} and P_{noise} denote the powers of the original signal and the additional Gaussian noise, respectively. S_{aug} represents the signal-to-noise ratio used in the noise augmentation technique in this study. Let n_{epi} denote the number of the episodes, and n_{epo2} represents the number of the training epochs in each episode. The algorithm in Stage 2 is summarized in Algorithm 1.

Algorithm 1: Representation clustering in semi-supervised learning

1: Input: Supervised data $\mathcal{D}_{sp} = \{(\mathbf{x}_i^{sp}, y_i^{sp})\}_{i=1}^{n_{sp}}$, unsupervised data $\mathcal{D}_{un} = \{\mathbf{x}_i^{un}\}_{i=1}^{n_{un}}$
2: Output: Pseudo labels $\{\hat{y}_i^{un}\}_{i=1}^{n_{un}}$ of unsupervised data
3: **for** episode $= 1, 2, \ldots, n_{epi}$ **do**
4: Initialize centroids $\{\mu_k\}_{k=1}^{N_{cluster}}$ based on $\{f_{en}(\mathbf{x}_i^{sp})\}_{i=1}^{n_{sp}}$ with Eq. (4)
5: Add noise on $\{\mathbf{x}_i^{un}\}_{i=1}^{n_{un}}$
6: Implement k-means clustering to obtain $\{\hat{y}_i^{un}\}_{i=1}^{n_{un}}$
7: **for** epoch $= 1, 2, \ldots, n_{epo2}$ **do**
8: Update $\boldsymbol{\theta}_{en}$ and $\boldsymbol{\theta}_{de}$ using Eq. (9)
9: **end for**
10: **end for**

2.5. Stage 3: Enhanced supervised learning

Through feature optimization in Stage 2, the unsupervised samples can be well clustered in the corresponding groups of the supervised data, and by attaching the pseudo labels, the labeled dataset can be significantly enlarged. Based on the extracted features of the encoder, a discriminative model for fault diagnosis can be trained using the expanded dataset. Concretely, a classifier module with parameters $\boldsymbol{\theta}_c$ is used as shown in Fig. 3, taking the high-level representations $\mathbf{x}_f$ as inputs. Two convolutional layers with 32 and 16 filters are used, followed by a flatten layer. Two fully-connected layers are adopted, with 128 and N_{class} neurons, respectively. The softmax function is finally used for machine health condition classification [14], which estimates the probabilities of the samples belonging to different labels. The supervised learning scheme is adopted in this stage, and the cross-entropy loss function L_s for classification is used as,

$$L_s = -\frac{1}{n_{sp}} \sum_{i=1}^{n_{sp}} \sum_{k=1}^{N_{cluster}} 1\{y_i^{sp} = k\} \log \frac{e^{x_{c,i,k}^{sp}}}{\sum_{j=1}^{N_{cluster}} e^{x_{c,i,j}^{sp}}} - \frac{1}{n_{un}} \sum_{i=1}^{n_{un}} \sum_{k=1}^{N_{cluster}} 1\{\hat{y}_i^{un} = k\} \log \frac{e^{x_{c,i,k}^{un}}}{\sum_{j=1}^{N_{cluster}} e^{x_{c,i,j}^{un}}}, \tag{11}$$

where $x_{c,i,k}^{sp}$ and $x_{c,i,k}^{un}$ denote the k-th element of the output vector in the classifier module, taking the i-th sample in the supervised and unsupervised data as inputs, respectively. The parameters of the classifier are optimized to minimize L_s, while the encoder is fixed to preserve the learned features,

$$\hat{\boldsymbol{\theta}}_c = \arg \min_{\boldsymbol{\theta}_c} \; L_s(\hat{\boldsymbol{\theta}}_{en}, \boldsymbol{\theta}_c), \tag{12}$$

where $\hat{\boldsymbol{\theta}}_c$ denotes the optimal values of $\boldsymbol{\theta}_c$. The parameters can be updated as,

$$\boldsymbol{\theta}_c \leftarrow \boldsymbol{\theta}_c - \delta \frac{\partial L_s}{\partial \boldsymbol{\theta}_c}. \tag{13}$$

3. Experimental study

3.1. Datasets

3.1.1. CWRU dataset

The publicly available CWRU rolling bearing dataset which is provided by the Bearing Data Center of Case Western Reserve University [49], is used to validate the effectiveness of the proposed method. The dataset is composed of multivariate vibration signals generated by a bearing test-rig, as presented in Fig. 2. The bearing data are collected by acceleration transducers with sampling rates of 12 kHz. Four load conditions are considered, i.e. 0, 1, 2 and 3 hp, and four rotating speeds of 1797, 1772, 1750 and 1730 rpm are implemented correspondingly. The vibration data used in this study were collected from the drive end of the motor on four different health conditions: (1) normal condition (H); (2) outer race fault (OF); (3) inner race fault (IF); and (4) ball fault (BF). The three kinds of faults are artificially created by electro-discharge machining with diameters of 7, 14 and 21 mils respectively. In summary, this dataset includes 10 bearing health states under the four operating conditions.

3.1.2. Bogie dataset

In order to further validate the proposed method, a more realistic and practical experimental setup of high-speed multi-unit train bogie bearing is built, and the test rig is presented in Fig. 4. The accelerometer is mounted on the load module to collect the vibration acceleration signals with sampling frequency of 5 kHz, and the sampling time is 10 s in different operating conditions. Three bearing rotating speeds of 1950, 1770 and 1590 rpm are implemented, the load conditions of which are denoted as S1, S2 and S3, respectively. Similar with the CWRU dataset, three kinds of bearing faults are considered, i.e. outer race fault (OF), roller fault (RF) and inner race fault (IF), which are artificially created by electro-discharge machining and shown in Fig. 4. Three fault severities are also implemented, i.e. incipient, medium and severe faults.

The detailed information of the two datasets is presented in Table 1.

3.2. Compared approaches

In this study, different approaches are implemented for comparisons, which share similar experimental settings and parameters with the proposed method.

(1) Basic

The Basic method follows the conventional supervised learning paradigm, where only the cross-entropy classification loss of the limited supervised data is considered, and the unsupervised data are not used for training. Considering the small amount of the labeled data used in the data-driven approach, overfitting has large opportunity to occur, and this approach thus provides the baseline performance of the traditional data-driven fault diagnosis methods.

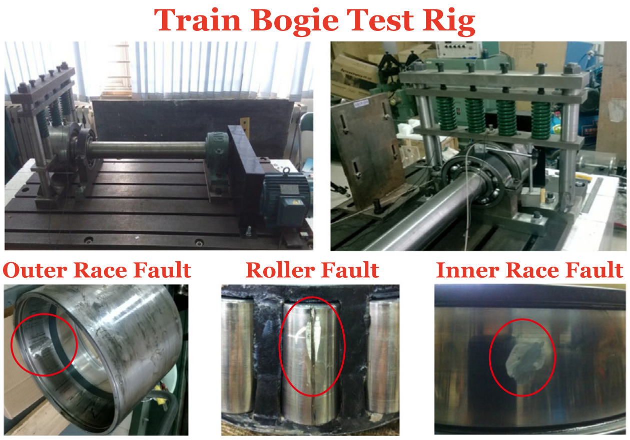

Fig. 4. Test rig and bearing faults in the Bogie dataset [39].

(2) NoAE
In order to examine the benefits of applying the auto-encoder scheme, the NoAE method is used where the unsupervised learning with the auto-encoder is removed from the proposed method in both Stage 1 and 2. Specifically, the reconstruction loss L_{ae} in Eq. (2) is not considered in optimization. This approach evaluates the benefits of the representation learning in the deep learning framework.

(3) ClusterOnly
In the proposed fault diagnosis framework, it is noted that the clustering effects obtained in Stage 2 can be readily used for diagnosing faults of the testing data. Therefore, the ClusterOnly method is implemented where the testing samples are directly fed into the encoder for feature extraction, and the label of the clustering centroid with the shortest L_1 distance with the testing sample in the feature space is considered the diagnosis results. The Stage 3 is not implemented in this approach. This method shows the effects of the proposed enhanced supervised learning step on the model performance in different tasks.

(4) NoNoise
The proposed noise augmentation method is able to enhance model generalization and robustness, alleviating model overfitting during training. The NoNoise method is implemented where the noise augmentation technique is not used in the proposed fault diagnosis framework for comparisons. The other stages are the same with the proposed method during training.

(5) DataAug
Data augmentation with additional noise is an alternative approach to address the data sparsity issue. In this study, additional Gaussian noise with SNR of 20, 8 and 4 is added on the original temporal vibration acceleration data to create fake samples. The enlarged supervised training dataset is used to train the fault diagnosis model afterwards, and this method is denoted as DataAug for comparisons. This approach offers the results of typical data augmentation methods on the data sparsity problems.

3.3. Numerical experiments

In this study, different fault diagnosis tasks are investigated on the two rotating machinery datasets, where variations of operating conditions, different preparations of the unsupervised data *etc.* are considered. The detailed descriptions of the tasks are presented in Tables 2 and 3. The dimension of the samples of machinery temporal vibration acceleration data is assumed to be 1024 in this study, and the network input dimension is thus $N_{input} = 512$ through FFT. In each concerned oper-

Table 1
Information of the CWRU and Bogie datasets. Inc., Med. and Sev. denote incipient, medium and severe, respectively.

Dataset	Descriptions										
	Condition Label	H	I-1	I-2	I-3	B-1	B-2	B-3	O-1	O-2	O-3
CWRU	Fault Location	N/A	IF	IF	IF	BF	BF	BF	OF	OF	OF
	Fault Size (mil)	0	7	14	21	7	14	21	7	14	21
Bogie	Fault Location	N/A	IF	IF	IF	RF	RF	RF	OF	OF	OF
	Fault Severity	N/A	Inc.	Med.	Sev.	Inc.	Med.	Sev.	Inc.	Med.	Sev.

Table 2
Data descriptions of the fault diagnosis tasks in the CWRU dataset (load).

Task	Supervised Data	Unsupervised Data	Testing Data
C1	0	0	0
C2	1	0, 2, 3	0, 2, 3
C3	1	1	0, 2, 3
C4	1	0, 1, 2, 3	0, 2, 3
C5	2	0, 1, 3	0, 1, 3
C6	2	2	0, 1, 3
C7	2	0, 1, 2, 3	0, 1, 3
C8	3	0, 1, 2	0, 1, 2
C9	3	3	0, 1, 2
C10	3	0, 1, 2, 3	0, 1, 2

Table 3
Data descriptions of the fault diagnosis tasks in the Bogie dataset (load).

Task	Supervised Data	Unsupervised Data	Testing Data
B1	S1	S1	S1
B2	S2	S2	S2
B3	S3	S3	S3

ating condition, with respect to each machine health state, N_{sp} labeled samples are available as the supervised data, N_{un} unlabeled samples are used as the unsupervised data, and N_{test} samples are tested. By default, $N_{test} = 100$ is adopted in this study.

In the model training process, the *Xavier* normal initializer is adopted for network parameter initializations. The back-propagation (BP) algorithm is applied for the updates of all the parameters, and the Adam optimization method [50] is adopted. In each training epoch, a mini-batch of samples are evaluated to obtain the gradients in optimiziation, and the batch size is 64 by default. The network architecture and parameters are mostly determined from validations in the CWRU task, where the supervised, unsupervised and testing data are from load 3 with $N_{sp} = 1$ and $N_{un} = 100$. The reported experimental results are generally averaged by 10 trials, and the default values of the related parameters are presented in Table 4.

3.3.1. Performance with weakly supervised data

First, the fault diagnosis problems with weakly supervised data are investigated. The experimental results of the ten fault diagnosis tasks C1-C10 on the CWRU dataset are presented in Table 5. Since the CWRU dataset with clean vibration signals is relatively easy for diagnostics, the extreme experimental setting is used in this study to examine the performance of the proposed method, where only one labeled sample in each machine health condition is used as supervised information, i.e. $N_{sp} = 1$. 100 unsupervised samples in each class are adopted for unsupervised learning, i.e. $N_{un} = 100$, and the noise augmentation technique with $S_{aug} = 20$ is used.

It can be observed that despite the close to 100% testing accuracies obtained by different methods in the task C1 where the training and testing data are collected from the same operating condition, the diagnostic performance significantly deteriorates in the other cross-domain tasks where the training and testing data are from different distributions. The results show the cross-domain fault diagnosis problems are still challenging especially with weakly supervised data.

Under the extreme setting with only 1 labeled sample in each condition, the proposed method generally achieves the highest testing accuracies in different cases. The Basic method with supervised learning alone mostly fails to diagnose faults, and the testing accuracies are only around 70%. The NoAE method without the unsupervised learning using auto-encoder obtains larger standard deviations, that shows projecting data into the same sub-space improves the model stability. Noticeable increases in the testing accuracies are observed by the proposed method compared with the ClusterOnly and NoNoise approaches. That suggests the proposed method outperforms the direct fault diagnosis approach based on the clustering

Table 4
Parameters used in this paper.

Parameter	Value	Parameter	Value
Epochs at Stage 1	4000	α_{ae}	1
n_{epi} Episodes at Stage 2	50	α_{intra}	1
n_{epo2} Epochs in each episode at Stage 2	200	α_{inter}	1
Epochs at Stage 3	4000	N_{rep}	128
S_{aug}	20	N_{input}	512
Learning rate at Stage 1	5e−5	Batch Size	64
Learning rate at Stage 2	1e−5	Dropout Rate	0.5
Learning rate at Stage 3	5e−5		

Table 5
Average testing accuracies of the fault diagnosis tasks with weakly supervised data in the CWRU dataset. The numbers in the brackets denote the standard deviations (%).

Task	Basic	NoAE	ClusterOnly	NoNoise	DataAug	Proposed
C1	99.8 (0.1)	99.9 (0.0)	98.4 (0.5)	99.9 (0.0)	99.9 (0.0)	99.9 (0.0)
C2	72.4 (4.6)	82.3 (6.3)	81.2 (3.2)	84.5 (3.3)	82.5 (2.3)	87.9 (3.1)
C3	72.4 (4.6)	74.6 (4.6)	80.2 (2.1)	81.3 (1.4)	82.5 (2.3)	87.3 (1.5)
C4	72.4 (4.6)	83.2 (5.8)	84.5 (2.7)	84.4 (2.5)	82.5 (2.3)	88.6 (2.0)
C5	70.5 (3.0)	87.2 (3.7)	86.5 (2.4)	88.6 (2.1)	78.6 (3.6)	90.6 (3.1)
C6	70.5 (3.0)	85.1 (3.0)	83.5 (3.6)	82.3 (5.3)	78.6 (3.6)	88.5 (1.7)
C7	70.5 (3.0)	87.5 (4.5)	87.4 (3.7)	88.7 (3.1)	78.6 (3.6)	92.5 (2.6)
C8	68.4 (4.8)	83.4 (5.1)	87.2 (1.8)	83.6 (1.8)	81.4 (2.7)	87.4 (3.6)
C9	68.4 (4.8)	79.4 (5.3)	86.4 (3.7)	82.4 (3.3)	81.4 (2.7)	85.2 (4.0)
C10	68.4 (4.8)	83.6 (4.1)	90.5 (2.1)	85.3 (2.0)	81.4 (2.7)	91.6 (2.4)

effect in Stage 2, and the proposed noise augmentation technique is able to further enhance the model generalization ability with respect to different operating scenarios. The DataAug method is a popular approach for weakly supervised learning problem through data augmentation, and its performance is less competitive compared with the proposed method.

Furthermore, it is noted that in the cross-domain fault diagnosis tasks, when the unsupervised data and testing data are from the same distribution, it is relatively easy to learn the generalized diagnostic knowledge, and the testing accuracies are generally higher. For instance, the performances of different methods in the task C5 are generally better than those in the task C6, and the testing accuracies in the task C7 are basically the highest where the supervised data under load 2 are used. That indicates the exploration of the unsupervised data significantly benefits the development of the fault diagnosis model, and better generalization ability can be achieved if more information of machine operating conditions is covered in the unsupervised data.

In order to show the detailed experimental results in different tasks, Fig. 5 shows the confusion matrices of the proposed method on the CWRU dataset. Specifically, the tasks C5 and C10 are illustrated, and the predictions and the ground truths of different classes are presented. It is noted that generally, promising testing accuracies can be obtained in different health conditions. Due to the general weak features in the ball fault conditions, the class B-1 achieves relatively lower diagnosis accuracies. However, higher than 75% testing accuracies can be still obtained. That shows the proposed method can well address the data sparsity issue in the CWRU dataset.

The experimental results on the Bogie dataset are presented in Table 6. Since the Bogie experiment is more practical and realistic, significant noises are observed in the collected data, and it is more challenging for fault diagnosis. Therefore, the diagnostic tasks on the same operating conditions are considered in this study. For the semi-supervised learning, only one labeled sample in each machine health condition is assumed to be available as the weakly supervised information., i.e. $N_{sp} = 1$. 100 unsupervised samples in each class are adopted for unsupervised learning, i.e. $N_{un} = 100$, and the noise augmentation technique with $S_{aug} = 20$ is used.

Similar with the results on the CWRU dataset, the proposed method still achieves the highest testing accuracies in different scenarios. Over 90% diagnostic accuracies are obtained using only one labeled sample for each health state, that shows the effectiveness of the proposed method in the realistic industrial scenario. The baseline method using conventional supervised learning achieves only around 60% accuracies. The NoAE, ClusterOnly, NoNoise and DataAug methods are also not competitive with the proposed method in different tasks.

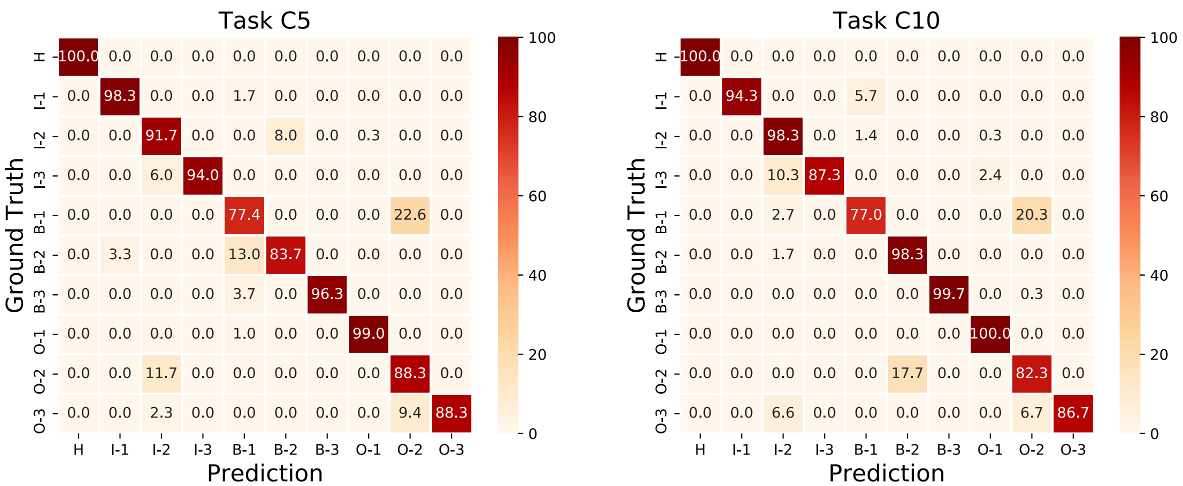

Fig. 5. The confusion matrices in tasks C5 and C10 with the proposed method. The numerical testing accuracies are presented (%).

Table 6
Average testing accuracies of the fault diagnosis tasks with weakly supervised data in the Bogie dataset. The numbers in the brackets denote the standard deviations (%).

Task	Basic	NoAE	ClusterOnly	NoNoise	DataAug	Proposed
B1	63.2 (5.2)	89.3 (4.6)	92.6 (2.1)	86.4 (3.6)	82.4 (5.6)	96.3 (1.7)
B2	58.5 (6.2)	88.5 (3.7)	87.4 (1.8)	85.4 (3.3)	85.8 (4.0)	91.4 (2.5)
B3	61.4 (4.8)	85.4 (5.3)	91.5 (2.5)	84.4 (2.8)	89.4 (3.5)	93.5 (2.9)

Fig. 6 shows the confusion matrices of the proposed method on the Bogie dataset. The tasks B1 and B3 are investigated, and the predictions and the ground truths of different classes are presented. Similar with the CWRU dataset, promising class-level diagnosis accuracies are still obtained on the Bogie dataset, and higher than 80% testing accuracies can be achieved.

In summary, through comparisons with different approaches, the effectiveness and superiority of the proposed method are validated in different fault diagnosis tasks in this section. Strong generalization ability with respect to variations of operating conditions can be achieved using the proposed semi-supervised learning method, and high testing accuracies are obtained with weakly supervised data, i.e. one labeled sample for each class.

3.3.2. Performance with unsupervised data

The proposed method can be also extended to address the fault diagnosis problems with purely unsupervised data. In order to examine the effectiveness of the proposed method in such cases, the standard unsupervised evaluation metric is used in this study [51,52]. It is assumed that the number of the clusters is equal to that of the actual machine health conditions, i.e. $N_{cluster} = N_{class}$. The unsupervised clustering accuracy ACC can be defined as,

$$ACC = \max_{map} \frac{\sum_{i=1}^{n_{test}} \mathbf{1}\{y_i^{test} = map(c_i^{assign})\}}{n_{test}}, \tag{14}$$

where c_i^{assign} denotes the cluster assignment of the i-th sample by the proposed method, and $map()$ ranges over all the possible one-to-one mappings between the clusters and the real class labels. Generally, Eq. (14) obtains the best corresponding class-level prediction accuracy on the testing data.

The experimental results on the CWRU and Bogie fault diagnosis tasks are presented in Table 7, where no supervised data is used and the settings are similar with those in the previous section. Specifically, the testing data are directly processed for clustering. Since no supervised learning is implemented, the Stage 3 is not considered, and the proposed method presents the clustering performance in the Stage 2. It can be observed from the results that the proposed method is able to achieve promising diagnostic performances. Over 95% accuracies are obtained in different scenarios on the CWRU tasks with unsupervised learning. Noticeable improvements are observed compared with the NoAE and NoNoise methods. Similar results are also achieved on the Bogie tasks. The effectiveness of the proposed method on addressing the unsupervised learning fault diagnosis tasks is thus validated.

3.3.3. Model performance analysis

In this section, the diagnostic performance of the proposed method is extensively investigated in different scenarios.

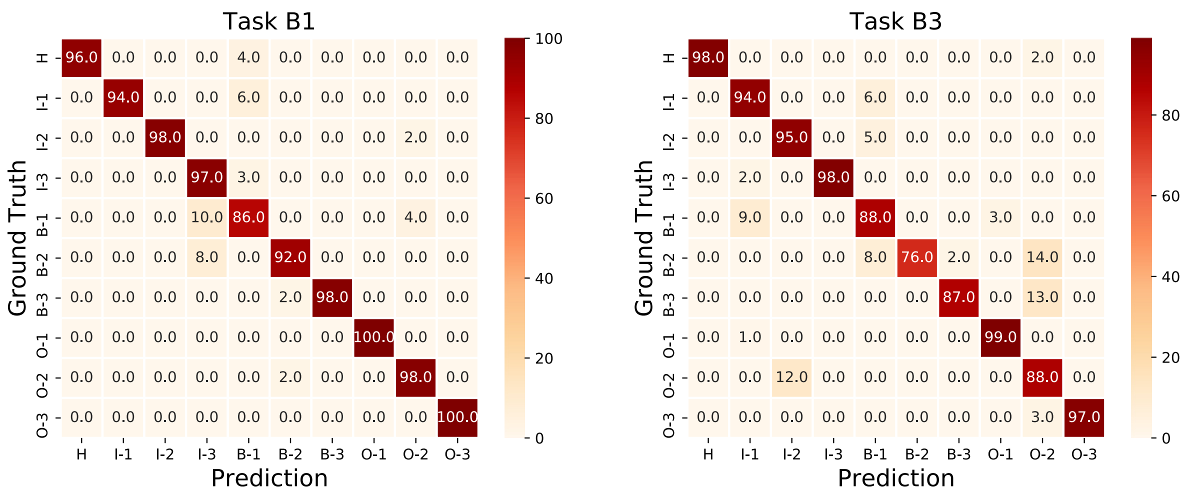

Fig. 6. The confusion matrices in tasks B1 and B3 with the proposed method. The numerical testing accuracies are presented (%).

Table 7
Descriptions and average testing accuracies of the unsupervised learning tasks in both the CWRU and Bogie datasets. The numbers in the brackets denote the standard deviations (%).

Task	Testing Data	NoAE	NoNoise	Proposed
C-Unsup1	load 1	95.7 (1.2)	95.4 (2.5)	97.4 (0.3)
C-Unsup2	load 2	95.5 (1.6)	90.4 (2.0)	95.7 (1.4)
C-Unsup3	load 3	93.7 (2.4)	93.5 (1.1)	97.2 (0.5)
C-Unsup4	load 4	95.3 (1.8)	91.7 (2.6)	96.5 (1.2)
B-Unsup1	load S1	82.5 (5.4)	81.4 (3.5)	90.6 (2.1)
B-Unsup2	load S2	80.1 (3.7)	77.3 (4.2)	87.4 (1.8)
B-Unsup3	load S3	85.3 (2.5)	83.6 (4.4)	91.5 (2.5)

First, the effects of the amount of the supervised data N_{sp} on the testing accuracies are evaluated in different tasks, and the results are presented in Fig. 7. It can be observed that on both the CWRU and Bogie datasets, larger N_{sp} clearly leads to higher testing accuracies in different cases. That indicates more labeled data remarkably improve the generalization ability of the proposed method, despite the fact that the proposed semi-supervised learning approach mostly focuses on the tasks with weakly supervised data. The results are also in accordance with the current understanding of deep learning, which is generally able to perform better with more high-quality labeled data.

Next, the influence of the unsupervised data are examined, and the relationship between the data amount N_{un} and the testing accuracy is presented in Fig. 8. In general, the proposed method explores the information of the unsupervised data for clustering. When N_{un} is larger, more data variants are used for model training, that enhances the model generalization ability. Correspondingly, higher testing accuracies are mostly obtained in different scenarios. The results suggest for the fault diagnostic tasks with weakly supervised data, it is feasible and promising to utilize more unsupervised data to improve the model performance. Considering the unsupervised data are mostly easy to collect in the industries, the proposed method is well suited for potential applications in real cases.

As one of the key hyper-parameters in the proposed architecture, the effects of the dimension of the high-level representation vector obtained by the feature extractor module are investigated. The testing results in different tasks are presented in Fig. 9. A general display pattern is observed that higher testing accuracies can be obtained with N_{rep} of the medium size. Small dimension such as 32 is generally not able to contain sufficient useful information of the data, and lower testing accuracies are obtained in different cases. When large dimension is used such as 512, significant drops on the model performance are observed. That indicates while the model capacity can be guaranteed using large vector dimensionality, the training difficulty of the deep neural network also remarkably increases. Therefore, a medium size of N_{rep} is suggested in practical implementations.

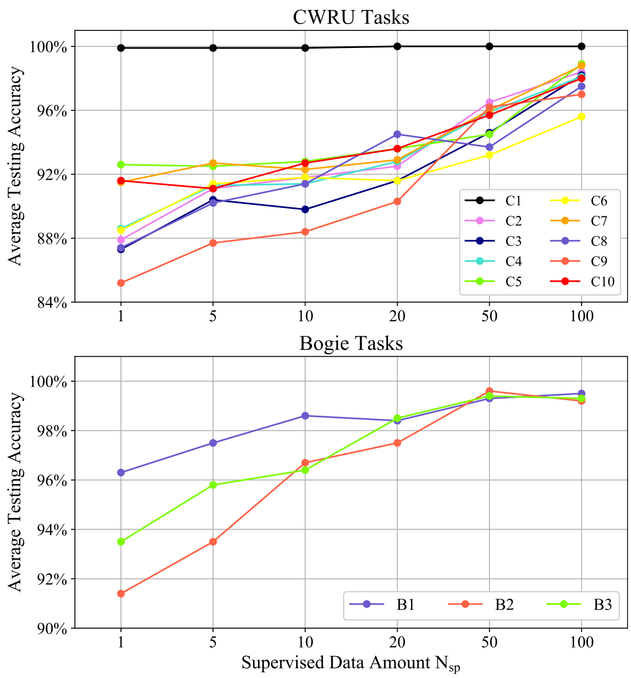

Fig. 7. Effects of the amount of supervised data on the model testing accuracies in different tasks.

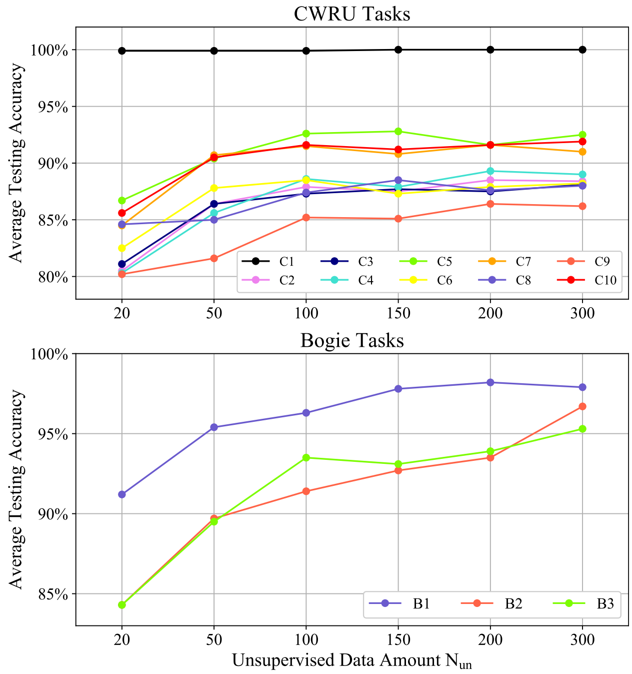

Fig. 8. Effects of the amount of unsupervised data on the model testing accuracies in different tasks.

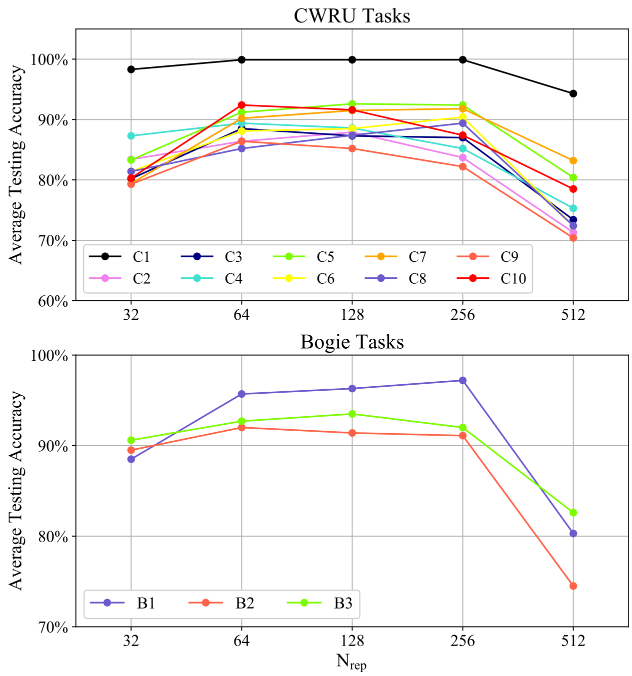

Fig. 9. Effects of the high-level representation dimensionality on the model testing accuracies in different tasks.

As mentioned in Section 2.4, using the proposed representation clustering method, overfitting has large opportunity to occur, that deteriorates the generalization ability of the model. The noise augmentation technique is proposed in this study to alleviate this issue. The effects of the adopted noise strength are investigated next, and the results are shown in Fig. 10. It is noted that compared with the plain method without noise augmentation, the proposed technique achieves significant improvements when weak additional noise is utilized. However, when stronger noise is adopted with SNR below 8, the model performance remarkably decays. Therefore, weak additional noise is suggested in the proposed fault diagnosis framework to avoid model overfitting.

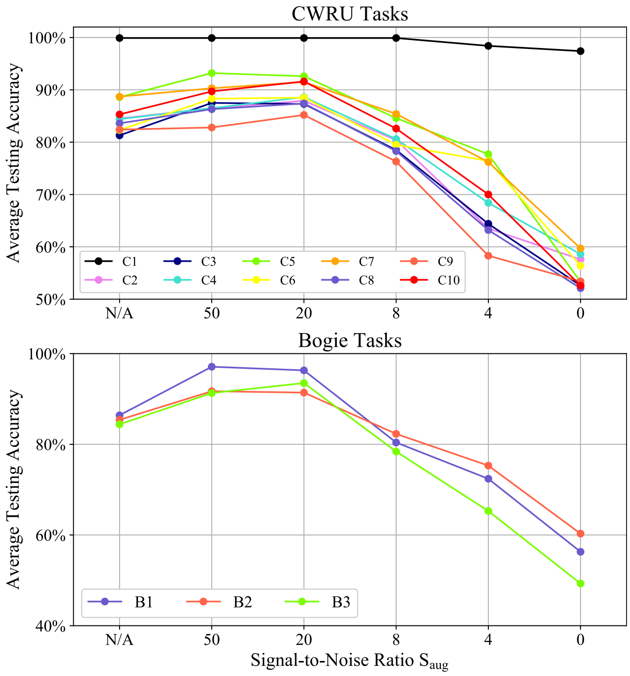

Fig. 10. Effects of the noise augmentation strength on the model testing accuracies in different tasks.

3.3.4. Visualizations

In this section, the proposed fault diagnosis method is quantitatively investigated through visualizations using the popular ‘‘t-SNE” technique, and the learned data representations by the feature extractor module are visualized. Fig. 11 shows the visualization results in the tasks C3 and C2, where the number of the clusters are equal to that of the real health conditions. It can be observed that for the task C3 where the supervised data and unsupervised data are from the same operating

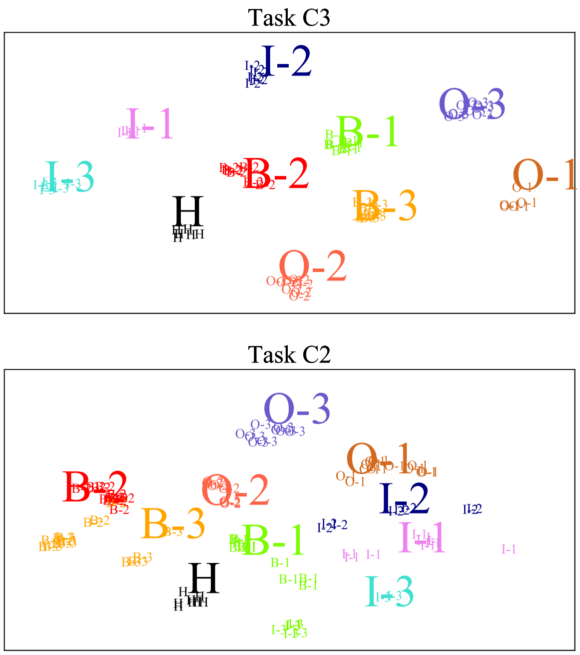

Fig. 11. Visualizations of the learned high-level representations in tasks C3 and C2. The small texts denote the real class labels of the unsupervised samples, and the large texts denote the labels of the cluster centroids. The colors represent the learned cluster assignments of the unsupervised samples to the corresponding centroids. A small portion of the samples are presented for better view.

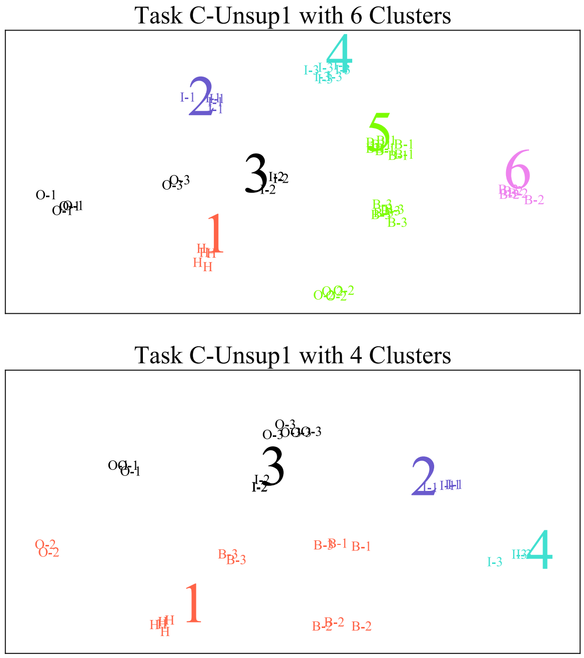

Fig. 12. Visualizations of the learned high-level representations in task C-Unsup1 with 6 and 4 clusters. The small texts denote the real class labels of the unsupervised samples, and the large texts denote the numbers of the cluster centroids. The colors represent the learned cluster assignments of the unsupervised samples to the corresponding centroids. A small portion of the samples are presented for better view.

condition, excellent clustering effect is obtained on the unsupervised data. The samples of the same classes are drawn closer to each other, and different classes are well separated. Nearly all the unsupervised samples are correctly assigned to the corresponding labels. For the task C2 where the supervised data and unsupervised data are from different operating conditions, a small amount of overlappings are observed for different classes. However, promising clustering phenomenon is still obtained on the unsupervised samples, despite the distribution discrepancy in the training data.

Moreover, the unsupervised learning cases are investigated next, where the numbers of clusters and real classes are not identical. Fig. 12 shows the visualization results in the task C-Unsup1, where 6 and 4 clusters are used respectively. It can be observed that despite fewer clusters are adopted, the samples of the same classes are still projected into the same regions, and different classes are separated. In some clusters, more than one class are covered. However, it seldomly occurs that the samples of the same condition are assigned to different clusters. That shows the extracted features by the proposed method preserve discriminative information with respect to different health conditions, which facilitates unsupervised clustering of the samples.

4. Conclusions

This paper proposes a deep representation clustering-based fault diagnosis method for rotating machines. A three-stage training scheme is adopted, i.e. pre-training, representation clustering and enhanced supervised learning. The auto-encoder structure is used to project data into a shared sub-space, and distance metric learning and k-means clustering approach are integrated in the proposed framework for unsupervised learning. Through explorations of the limited supervised data and sufficient unsupervised data, better discriminative features with respect to different machine health conditions can be captured, and the fault diagnosis classifier with strong generalization ability can be established. Based on the experimental results on two rotating machinery datasets, it is validated that the proposed method not only addresses the semi-supervised learning problem with limited labeled data, but also achieves promising results on the unsupervised learning tasks.

However, since the k-means clustering method is used in this study, the main limitation lies in that the number of clusters is supposed to be determined in advance. Further research works will be focused on the improved clustering approaches to relax the limitation, as well as validations with data in more industrial scenarios.

Acknowledgements

The material in this paper is based on work supported by grants (N180703018, N2005010, N180708009 and N170308028) from the Fundamental Research Funds for the Central Universities, and grant (11902202) from the National Natural Science Foundation of China.

References

[1] Y. Liu, Y.L. Zhao, J.T. Li, H. Ma, Q. Yang, X.X. Yan, Application of weighted contribution rate of nonlinear output frequency response functions to rotor rub-impact, Mech. Syst. Signal Process. 136 (2020) 106518.
[2] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, Z. Zhu, A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines, Mech. Syst. Signal Process. 116 (2019) 668–692.
[3] K. Yu, T.R. Lin, H. Ma, H. Li, J. Zeng, A combined polynomial chirplet transform and synchroextracting technique for analyzing nonstationary signals of rotating machinery, IEEE Trans. Instrum. Meas. (2019) 1.
[4] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, B. Wen, Review on dynamics of cracked gear systems, Eng. Fail. Anal. 55 (2015) 224–245.
[5] Z. Luo, J. Wang, R. Tang, D. Wang, Research on vibration performance of the nonlinear combined support-flexible rotor system, Nonlinear Dyn. 98 (1) (2019) 113–128.
[6] X. Jiang, C. Shen, J. Shi, Z. Zhu, Initial center frequency-guided VMD for fault diagnosis of rotating machines, J. Sound Vib. 435 (2018) 36–55.
[7] X. Li, W. Zhang, Q. Ding, X. Li, Diagnosing rotating machines with weakly supervised data using deep transfer learning, IEEE Trans. Industr. Inf. 16 (3) (2020) 1688–1697.
[8] H. Shi, L. Guo, S. Tan, X. Bai, Rolling bearing initial fault detection using long short-term memory recurrent network, IEEE Access 7 (2019) 171559–171569.
[9] X. Li, W. Zhang, Q. Ding, J.-Q. Sun, Multi-layer domain adaptation method for rolling bearing fault diagnosis, Signal Process. 157 (2019) 180–197.
[10] 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.
[11] W. Zhang, X. Li, Q. Ding, Deep residual learning-based fault diagnosis method for rotating machinery, ISA Trans. 95 (2019) 295–305.
[12] 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.
[13] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocomputing 240 (2017) 98–109.
[14] 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. Industr. Electron. 63 (5) (2016) 3137–3147.
[15] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, A sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement 89 (2016) 171–178.
[16] C. Lu, Z.Y. Wang, W.L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Process. 130 (2017) 377–388.
[17] 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 Process. 104 (2018) 799–834.
[18] W. Sun, R. Zhao, R. Yan, S. Shao, X. Chen, Convolutional discriminative feature learning for induction motor fault diagnosis, IEEE Trans. Industr. Inf. 13 (3) (2017) 1350–1359.
[19] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[20] 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 Process. 110 (2018) 349–367.
[21] X. Li, W. Zhang, Q. Ding, J.-Q. Sun, Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation, J. Intell. Manuf. 31 (2020) 433–452.
[22] Y. Zhang, X. Li, L. Gao, L. Wang, L. Wen, Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning, J. Manuf. Syst. 48 (2018) 34–50.
[23] C. Bunkhumpornpat, K. Sinapiromsaran, C. Lursinsap, Safe-Level-SMOTE: safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem, Proceedings of Advances in Knowledge Discovery and Data Mining, Berlin, Heidelberg, 2009, pp. 475–482.
[24] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, X. Li, Machinery fault diagnosis with imbalanced data using deep generative adversarial networks, Measurement 152 (2020) 107377.
[25] S. Shao, P. Wang, R. Yan, Generative adversarial networks for data augmentation in machine fault diagnosis, Comput. Ind. 106 (2019) 85–93.
[26] X. Gao, F. Deng, X. Yue, Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty, Neurocomputing (2019), https://doi.org/10.1016/j.neucom.2018.10.109.
[27] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, W. Jiang, Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks, Neurocomputing 315 (2018) 412–424.
[28] X. Zhao, M. Jia, A novel unsupervised deep learning network for intelligent fault diagnosis of rotating machinery, Struct. Health Monitor. (2020), 1475921719897317.
[29] P. Liang, C. Deng, J. Wu, G. Li, Z. Yang, Y. Wang, Intelligent fault diagnosis via semi-supervised generative adversarial nets and wavelet transform, IEEE Trans. Instrum. Meas. (2019) 1.
[30] K. Zhang, B. Tang, Y. Qin, L. Deng, Fault diagnosis of planetary gearbox using a novel semi-supervised method of multiple association layers networks, Mech. Syst. Signal Process. 131 (2019) 243–260.
[31] Y. Hu, P. Baraldi, F.D. Maio, E. Zio, A systematic semi-supervised self-adaptable fault diagnostics approach in an evolving environment, Mech. Syst. Signal Process. 88 (2017) 413–427.
[32] J. Zheng, H. Wang, Z. Song, Z. Ge, Ensemble semi-supervised fisher discriminant analysis model for fault classification in industrial processes, ISA Trans. 92 (2019) 109–117.
[33] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S.H. Kia, H. Henao, G. Capolino, Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems, IEEE Trans. Industr. Electron. (2018) 1.
[34] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, A semi-supervised diagnostic framework based on the surface estimation of faulty distributions, IEEE Trans. Industr. Inf. 15 (3) (2019) 1277–1286.
[35] Z. Zhang, S. Li, J. Wang, Y. Xin, Z. An, General normalized sparse filtering: a novel unsupervised learning method for rotating machinery fault diagnosis, Mech. Syst. Signal Process. 124 (2019) 596–612.
[36] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, Proceedings of 32nd International Conference on Machine Learning, vol. 37, 2015, pp. 97–105.
[37] X. Li, W. Zhang, N. Xu, Q. Ding, Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places, IEEE Trans. Industr. Electron. (2019) 1.
[38] 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. Industr. Electron. (2018) 1.
[39] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Industr. Electron. 66 (7) (2019) 5525–5534.
[40] B. Zhang, W. Li, X. Li, S. Ng, Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks, IEEE Access (2018) 1.
[41] 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.
[42] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly-accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Industr. Inf. (2018) 1.

[43] T. Han, C. Liu, W. Yang, D. Jiang, Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions, ISA Transactions.
[44] 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 Process. 122 (2019) 692–706.
[45] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Industr. Electron. 64 (3) (2017) 2296–2305.
[46] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434.
[47] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, D.J. Inman, Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, J. Sound Vib. 388 (2017) 154–170.
[48] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recogn. Lett. 31 (8) (2010) 651–666.
[49] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, Mech. Syst. Signal Process. 64–65 (2015) 100–131.
[50] D. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
[51] M. Caron, P. Bojanowski, A. Joulin, M. Douze, Deep clustering for unsupervised learning of visual features, in, Proceedings of ECCV 2018, Cham, 2018, pp. 139–156.
[52] J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, The 33rd International Conference on Machine Learning, vol. 48, 2016, pp. 478–487.