International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-024-02233-0

ORIGINAL ARTICLE

Proportional periodic sampling for cross-load bearing fault diagnosis

Jianbo Zheng[1] · **Bin Jiang**[1] · **Chao Yang**[1]

Received: 7 July 2023 / Accepted: 28 May 2024

Abstract
Bearing vibration data under various loads shows different distributions, which leads to the poor performance of existing deep learning methods in performing cross-load fault diagnosis tasks. As a result, cross-load deep transfer learning methods that can reduce the impact of distribution differences have become the current research priority. However, existing transfer learning methods fall short in fully leveraging the bearing data characteristics, resulting in unreasonable sampling strategy, complex or inefficient network structure. These limitations hamper the performance improvement of cross-load fault diagnosis methods. To address the above drawbacks, a transductive convolution transfer learning (TCTL) method based on proportional periodic sampling is proposed. First, according to the periodic characteristic of bearing vibration data, the proportional periodic sampling is conducted on all vibration data to construct high-quality sample sets, which can greatly reduce the number of model parameters and ensure the excellent diagnostic accuracy of the model. Second, considering the short length of vibration samples and minor distribution differences between various loads, the cross-load convolutional neural network (CL-CNN) model is proposed to extract fault features from the constructed samples efficiently. Third, utilizing the characteristic that the vibration data under each load satisfies the clustering assumption, the cross-load multi-objective loss is used to effectively supervise the CL-CNN model to extract more domain-invariant features. Finally, compared to ablation and latest methods, TCTL boasts higher average value and lower standard deviation of accuracy on both datasets, proving the effectiveness of TCTL.

Keywords Intelligent fault diagnosis · Transfer learning · Maximum mean discrepancy · Convolutional neural network

1 Introduction

In the industrial environment, manufacturing enterprises must perform high-performance (high-precision and stable) fault diagnosis for the mechanical equipment undertaking production tasks. Bearings, as fundamental components of machinery, find extensive application across various equipment, particularly in rotating machinery [1]. When bearings malfunction, they generate abnormal vibration data, which not only diminishes the overall performance of the machinery but can also pose serious risks to personnel safety [2]. Therefore, there is considerable research and practical significance in achieving high-performance fault diagnosis for bearings.

On the one hand, with the gradual implementation of the industrial internet of things, numerous low-cost sensors are deployed to industrial machinery, thus obtaining vibration-sensing data that can accurately describe the operation status of machinery. On the other hand, deep learning techniques have made remarkable achievements in natural language processing and computer vision, showing strong feature extraction and pattern recognition capabilities [3, 4]. Moreover, deep learning has shown the progressiveness in extracting fault features from the vibration-sensing data efficiently and comprehensively [5, 6]. Consequently, based on the collected vibration data, the intelligent diagnosis of bearing faults using deep learning technology has emerged as a prominent research focus, garnering widespread attention from both academia and industry [7–11].

Jianbo Zheng and Chao Yang contributed equally to this work.

✉ Bin Jiang
jiangbin@hnu.edu.cn

Jianbo Zheng
zhengjianbo@hnu.edu.cn

Chao Yang
yangchaoedu@hnu.edu.cn

[1] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, Hunan, China

Published online: 18 June 2024

For bearing fault diagnosis, the vibration data in the datasets has three characteristics.

1. Multiple operating loads: the load of bearing is switched between multiple fixed loads according to the production demand, so there are multiple vibration data under different loads in the same dataset.
2. Different distribution between various loads: the changes of operating load directly affects the bearing deformation degree, resulting in difference vibration data distribution [12].
3. Limited sampling points under each load: in some existing public datasets, vibration data is collected only for a short time at each load, resulting in a small number of samples sampled from the limited vibration data.

Numerous studies have highlighted two key prerequisites for achieving high-performance deep learning methods [13–15]. One is that there should be sufficient samples, and the other is that the distribution of training data and test data should be consistent. Due to the small number of samples and the different distribution of vibration data under different loads, the performance of existing fault diagnosis methods based on deep learning is limited under cross-load conditions. Recognizing the challenges posed by these conditions, Zhao et al. achieve effective fault diagnosis under cross-load conditions by using common transfer learning theories and techniques [16].

Transfer learning [17, 18], as one of the machine learning methods, involves transferring knowledge acquired from a source domain to a target domain by leveraging data or task similarities. Transfer learning-based methods, which alleviate the impact of distribution difference between the source and target domains by reducing the Maximum Mean Difference (MMD) [19] distance between two domains, are generally divided into two categories: inductive transfer learning (also known as domain adaptation), where both the source and target domains possess labels, and transductive transfer learning, where labels are available only for the source domain.

Since fault labels of vibration data can only be obtained by stopping the machine for inspection, which is unacceptable in the actual production environment, the vibration data collected during the operation of the machine are unlabeled [20]. Moreover, in conjunction with the descriptions in the literatures [7, 13, 21], this paper defines cross-load fault diagnosis as accurate fault classification based on labeled data of one load for unlabeled data of another load. By treating the load with labeled data as the source domain and the load with unlabeled data as the target domain, cross-load fault diagnosis can be effectively accomplished using transductive transfer learning-based methods. High-performance cross-load fault diagnosis holds significant promise for enhancing the intelligent operation and maintenance of wind turbines operating under variable loads and harsh environmental conditions.

The vibration data has four characteristics for transductive transfer learning-based cross-load fault diagnosis.

1. Bearing vibration data has periodic characteristic: the fault patterns of bearing is closely related to the rotation period, so the length of each sample is generally fixed as the vibration data collected within 1 to 2 rotation periods [22]. Moreover, Zheng et al. first proposed a standard periodic sampling strategy for fault diagnosis, and achieved good fault diagnosis performance [5]. However, the length of periodic samples may be very long, which will greatly increase the computational resources required by the diagnostic model. In order to effectively shorten the sample length while ensuring the diagnostic performance, based on the periodic characteristics of vibration data, this paper assumes that the vibration data with a partial rotation period has sufficient fault information for high-performance cross-load fault diagnosis.
2. The length of samples is short: the sample length of an image is 3072 ($32 \times 32 \times 3$) in computer vision, while the sample length of the vibration sample is generally set to 1024 (32×32). For example, Ghorvei et al. simply set the sample length to 1024, ignoring the utilization of periodic features [23]. This paper sets the sample length as 420 or 500, and the short sample length is particularly prominent. It is inappropriate to directly apply the models in computer vision for fault diagnosis of vibration samples, so it is necessary to propose an efficient feature extraction model for the short sample length.
3. The difference between various loads is minor in distribution and bearing fault categories are cross-load: because the minor difference in distribution between various loads, deep learning methods achieve promising results under cross-load conditions, with an accuracy of around 90%. However, to achieve higher performance with accuracy above 99%, it is necessary to reduce the impact of minor differences in distribution among various loads. For instance, leveraging the correlation alignment (CORAL) metric, Wang et al. proposed a Deep Feature Correlation Matching Network (DFCMN) method to achieve high-precision cross-load fault diagnosis [24]. Furthermore, across the same dataset, the bearing fault categories remain consistent under various loads. This consistency allows the same network model to efficiently extract domain-invariant features from bearing vibration data under different loads, thereby enabling effective classification.

4. The vibration data under different loads all conform to the clustering assumption: the clustering assumption means that samples with the same labels are very close while samples with different labels are far away, i.e., the decision boundaries between samples with various labels are clear.

However, most transductive transfer learning-based cross-load fault diagnosis methods lack the utilization of vibration data characteristics, resulting in the following three defects.

- Unreasonable sampling strategy and incomplete use of vibration data: some methods ignore the importance of the sample sampling strategy, which significantly affects the model performance. Additionally, the utilization of only partial vibration data to construct a balanced sample set may omit certain crucial fault patterns. For example, Ding et al. constructed a balanced sample set for cross-load fault diagnosis, without using all the data [25]. These defects result in limitations in the performance of existing methods.
- Complex network structure: many methods ignore the characteristics of short sample length and minor differences in distribution. Multi-layer complex network structures such as Resnet18 [26] and VGG16 [27] are used for cross-load bearing fault diagnosis, resulting in limited feature extraction ability for short samples and an enormous waste of computing resources.
- Inefficiency loss function: some methods ignore the characteristic that unlabeled samples conform to the clustering assumption, resulting in the inefficiency of the designed loss function, which limits the improvement of the performance of the model [16].

To overcome the three defects above, according to the characteristics of bearing vibration data, a Transductive Convolutional Transfer Learning (TCTL) method based on proportional periodic sampling is proposed for cross-load fault diagnosis. The contributions are summarized below.

- First, based on the assumption that a subset of vibration data of a rotational period contains ample fault information, the proportional periodic sampling method is proposed to construct short sample sets with sufficient fault information and more sampling numbers from all original vibration data, which is the first successful attempt to utilize the periodic characteristics of bearing vibration data for cross-load fault diagnosis.
- Second, this paper proposes a simple Cross-Load Convolutional Neural Network (CL-CNN) model designed for efficient feature extraction, capitalizing on the characteristics of short sample length, minor distribution differences, and cross-load fault categories. Furthermore, leveraging the clustering assumption inherent in vibration data across various loads, a cross-load multi-objective loss function is proposed. This function integrates cross-entropy loss in the source domain with labels, MK-MMD distance, and cross-entropy loss in the target domain with pseudo-labels to supervise the CL-CNN model to obtain better classification boundaries. Ultimately, the collaborative integration of CL-CNN with the cross-load multi-objective loss enables the efficient extraction and classification of fault features within the sample sets.
- Third, comprehensive experimental results on two public datasets demonstrate the superiority of the proposed TCTL method.[1]

2 Background

2.1 Formal definition of cross-load fault diagnosis

Transfer learning modeling involves the definition of domain D and task T. Domain D consists of feature space χ and marginal distribution $P(X)$, namely $D = \{\chi, P(X)\}$, where X is vibration samples. Task T consists of label space Y and label discriminant function $f(.)$, namely $T = \{Y, f(.)\}$. The function of transfer learning is to transfer the knowledge of the source domain D^s to the target domain D^t. In this paper, the feature spaces χ^s, χ^t of the source and target domains are the same, that is $\chi^s = \chi^t$.

For cross-load fault diagnosis, the samples in the source domain are labeled with g labels in the label space Y^s. The target domain samples are unlabeled, but have a potential label space Y^t that is consistent with the label space Y^s. This means that Y^s and Y^t have the same set of labels, denoted $Y^s = Y^t \in \{L_1, L_2, \cdots, L_g\}$. Since the label discriminant function $f(.)$ used in the source domain is the same as the one used in the target domain, the source domain task $T^s = \{Y^s, f(.)\}$ is the same as the target domain task $T^t = \{Y^t, f(.)\}$.

First, the labeled source domain D^s is defined as the following formula:

$$D^s = (x_i^s, y_i^s) | i \in (1, n), x_i^s \in \chi^s, y_i^s \in Y^s \tag{1}$$

where n is the number of labeled samples. x_i^s and y_i^s are the feature and the label of the i-th sample in D^s respectively.

Next, the unlabeled target domain D^t is defined by the following formula:

$$D^t = \{x_i^t | i \in (1, m), x_i^t \in \chi^t\} \tag{2}$$

[1] https://github.com/IWantBe/TCTL-for-cross-load-fault-diagnosis.

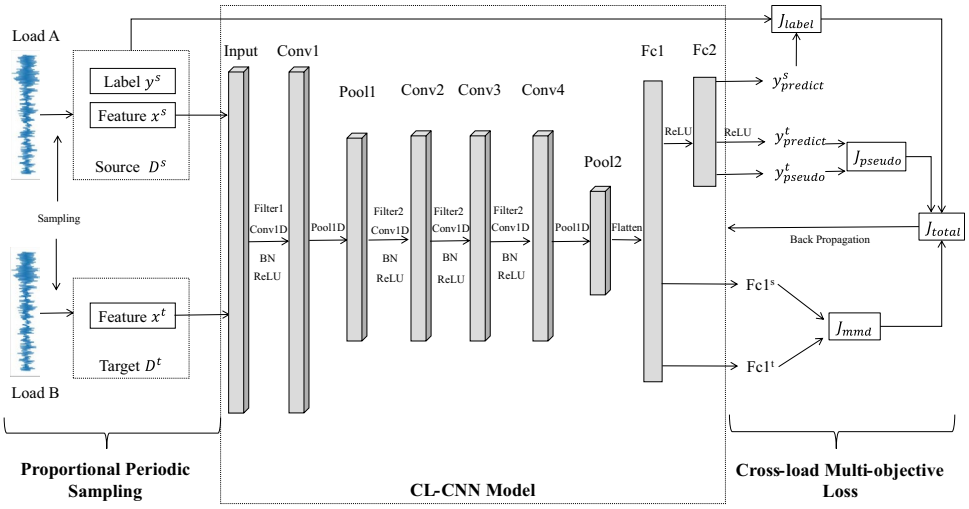

Fig. 1 The overall architecture of the TCTL method. The TCTL method consists of three parts. Firstly, it uses proportional periodic sampling to obtain more short samples. Secondly, it employs the simple CL-CNN model to carefully and comprehensively extract fault features. Lastly, it uses cross-load multi-objective loss to supervise the model's training

where m is the number of unlabeled samples, and x_i^t is the feature of the i-th sample in D^t.

Finally, the process of transductive transfer learning-based cross-load fault diagnosis is defined as follows: given the labeled source domain D^s, source domain classification task $T^s = \{Y^s, f(.)\}$, unlabeled target domain D^t and target domain classification task $T^t = \{Y^t, f(.)\}$, in the case $D^s \neq D^t$ and $T^s = T^t$, the label discriminant function $f(.)$ is optimized by utilizing the knowledge present in both D^s and D^t, and then the optimized $f(.)$ is used to classify the unlabeled samples in D^t.

2.2 MK-MMD

When applying transfer learning, it's important to measure the difference in data distribution between the source and target domains. This can be accomplished by calculating the distance or similarity between them. The metric used to measure this difference directly affects the method's final performance, so it's important to choose a well-designed measure. The most commonly used metric in transfer learning is the MMD distance. This metric calculates the distance between the feature distribution of source domain samples and that of target domain samples in the Reproducing Kernel Hilbert Space (RKHS).

Assume that there are n samples in D^s and m samples in D^t, the MMD distance is calculated as follows:

$$J_{mmd}\left(D^s, D^t\right) = \sqrt{\left\| \frac{1}{n}\sum_{i=1}^{n} f\left(x_i^s\right) - \frac{1}{m}\sum_{i=1}^{m} f\left(x_i^t\right) \right\|_{\mathcal{H}_k}^2} \tag{3}$$

where $\mathcal{H}_k$ is the RKHS using the characteristic kernel k, and the Gaussian kernel is the most commonly used kernel. $f(.)$ is the mapping function, and $f(x_i^s)$, $f(x_i^t)$ are the mapping of the i-th sample x_i^s, x_i^t on the RKHS, respectively.

Next, we take the linear combination of multiple kernels k to obtain an optimal multi-kernel K, and the formula is as follows:

$$K = \sum_{u=1}^{d} \beta_u k_u, \sum_{u=1}^{d} \beta_u = 1, \beta_u \geq 0 \tag{4}$$

where d is the number of kernels, k_u is the u-th kernel, and β_u is the coefficient of the u-th kernel.

Finally, the MMD distance is calculated using the optimal multi-kernel K, denoted as the MK-MMD distance [28].

3 Proposed method

3.1 The overall architecture of TCTL

The overall architecture of the TCTL method proposed in this paper is shown in Fig. 1.

First, load A and load B are considered as the labeled source and unlabeled target domains, respectively. Then, based on the assumption that the vibration data with a partial rotation period has sufficient fault information for high-performance cross-load fault diagnosis, the vibration data in the source and target domains are sampled in a proportional period to obtain high-quality sample sets with more short samples. The labeled source domain samples set $D^s = \{x^s, y^s\}$, and the unlabeled target domain samples set $D^t = \{x^t\}$ are sampled, respectively.

Second, according to the characteristics that the sample length is short and bearing fault categories are cross-load, both the labeled feature x^s and the unlabeled feature x^t are

input into the same simple CL-CNN model for detailed feature extraction.

Third, based on the characteristic that the source domain samples and target domain samples meet the clustering assumption, the CL-CNN model adopts pseudo-labelling technology to label unlabeled features x^t with the pseudo-labels y^t_{pseudo}. Moreover, after every round of model training, the y^t_{pseudo} of unlabeled features x^t is updated by the model with the latest parameters.

Fourth, the CL-CNN model is supervised by the efficient multi-objective loss J_{total}, which is composed of J_{label}, J_{pseudo}, and J_{mmd}. This helps the model to learn in a direction that not only extracts more domain-invariant features in samples but also reduces the impact of different data distribution between the source and target domains. After training the CL-CNN model, the unlabeled samples of the target domain are fed into the model for fault classification, completing the cross-load fault diagnosis task.

3.2 Proportional periodic sampling of vibration data

The bearing has g operating states (normal or faulty states). Under a load, the raw vibration data with N sampling points is described as follows:

$$D = \left\{ \left(\left[X_i^1, X_i^2, \cdots, X_i^N \right], y_i \right) \right\}, i \in \{1, 2, \cdots, g\} \tag{5}$$

where D contains the vibration data and status labels in g operating states. X_i^N represents the N-th sampling point of the vibration data under state i. y_i represents the status label under state i, and $y_i \in \{label_1, label_2, \cdots, label_g\}$.

The number of sampling points in a rotation period, n, can be calculated by the Revolutions Per Minute (RPM) and sampling frequency f.

$$n = (60 \times f)/RPM \tag{6}$$

The shorter the sample length, the more samples can be sampled from vibration data collected within a particular length of time, which can effectively enhance the versatility of the deep learning model. Moreover, shorter samples can significantly reduce deep learning model parameters, which lowers resource requirements for deployment.

However, due to differences in sampling frequency and bearing speed, there is a significant difference in the number of sample points calculated for the rotation period. In order to obtain more short samples for high-performance fault diagnosis, based on the assumption that vibration data with partial rotation period length has sufficient fault information, the length of each sample is set as a portion of the sampling points collected during one rotation cycle of the bearing. The length L of each sample is defined as follows:

$$L = \alpha \times n, \alpha \in (0, 1] \tag{7}$$

Considering that the total number of vibration sample points in public datasets is limited, it is necessary to use the sliding window sampling strategy to expand the number of samples available for training. It is important to consider both the sample length (L) and the sliding stride (S) when constructing high-quality samples. To define the sample set, samples are extracted from the original vibration data based on the sample length L and sliding stride S, using the following formula:

$$\begin{cases} sample\_set = \{x_i, y_i\}, i \in (1, g) \\ x_i = \left\{ x_i^1, x_i^2, \cdots \right\} \\ x_i^j = \left[X_i^{start}, \cdots, X_i^{start+L} \right] \\ x_i^{j+1} = \left[X_i^{start+S}, \cdots, X_i^{(start+L)+S} \right] \end{cases} \tag{8}$$

where x_i^j represents the j-th samples under state i. $start$ is initialized to 0, the sample length L and sliding stride S are adjusted as needed.

3.3 Fault feature extraction using CL-CNN

This paper uses the same network model for extracting fault features from both the source and target domain samples, as the size of the input (length of samples) and output (number of labels) is consistent in both domains. The feature extraction process using the CL-CNN is as follows.

The process begins with the sample features being fed into a convolutional layer Conv1. This layer uses one-dimensional convolution (Conv1D) with a convolution kernel Filter1, batch normalization (BN), and ReLU operations. Following this, a pooling layer with one-dimensional max-pooling (Pool1D) operation is used to obtain a shallow feature map Pool1.

Then, the Pool1 is input into two one-dimensional convolutional layers (Conv2 and Conv3). Both Conv2 and Conv3 perform Conv1D with convolution kernel Filter2, BN, and ReLU. Additionally, a one-dimensional max-pooling layer with Pool1D is used to obtain the high-level feature map Pool2.

Finally, Pool2 is flattened and fed into two fully connected layers to model the mapping relationship between the high-level feature map and fault types, thereby effectively extracting features and classifying faults.

When designing the CL-CNN model, two special treatments are adopted to extract more representative and comprehensive fault features, given the short sample length.

- To extract fault features that better represent the data, this paper proposes that $Filter1 > Filter2$. The rationale behind this is that the sample length is short and $Filter1$

is used to extract features from these short samples. Therefore, more filters are required to extract the most representative fault features hidden in short samples.
- To extract comprehensive fault features more carefully, the stride of Conv1D is set to 1.

3.4 Cross-load multi-objective loss function

The loss function J_{total} used in this paper consists of three components, J_{label}, J_{mmd} and J_{pseudo}. Each of these components is explained in detail below.

J_{label} is the cross-entropy loss of source domain samples with actual labels, and the model can learn more fault knowledge from the labeled samples in D^s by minimizing J_{label}. The formula for calculating J_{label} is as follows:

$$J_{label} = -\sum_{i=1}^{n} p^s\left(x_i^s\right) \log\left(q^s\left(x_i^s\right)\right) \tag{9}$$

where n is the number of samples in x^s, $p^s(x_i^s)$ and $q^s(x_i^s)$ are the distribution of the actual label y^s and the predicted label $y^s_{predict}$ of the i-th sample, respectively.

J_{mmd} is the MK-MMD distance between the feature map $Fc1^s$ and the feature map $Fc1^t$. $Fc1^s$ and $Fc1^t$ are the feature maps extracted from the source and target domain samples using the CL-CNN model, respectively. The smaller the MK-MMD distance is, the smaller the distribution difference between the source domain samples and the target domain samples on the $Fc1$ layer, and the better the transfer learning effect will be. By minimizing J_{mmd}, the model can reduce the distribution difference between the source and target domains. The formula for calculating J_{mmd} is given by formula (3).

J_{pseudo} is the cross-entropy loss of target domain samples with pseudo-labels [29], which is the key component that distinguishes it from traditional transfer learning losses. Because unlabeled samples in the target domain that meet the clustering assumption cannot be directly used to train the CL-CNN model, this paper proposes to first apply pseudo-labeling technology to label the unlabeled samples with pseudo-labels and then jointly trains the model with labeled samples. By minimizing J_{pseudo}, the model can obtain improved classification decision boundaries. The formula for calculating J_{pseudo} is as follows:

$$J_{pseudo} = -\sum_{i=1}^{m} p^t\left(x_i^t\right) \log\left(q^t\left(x_i^t\right)\right) \tag{10}$$

where m is the number of samples in x^t, $p^t(x_i^t)$ and $q^t(x_i^t)$ are the distribution of the pseudo-label y^t_{pseudo} and the predicted label $y^t_{predict}$ of the i-th sample, respectively.

The detailed process of labeling unlabeled samples with pseudo-labels is as follows: since the number of neurons in the fully-connected layer $Fc2$ is equal to the number of labels g, the output of $Fc2$ is a vector $V_{Fc2} = [v_0, \cdots, v_{g-1}]$. Then, the vector V_{Fc2} is converted into the one-hot code v_q representing a fault category according to (11), and the v_q is regarded as the pseudo-labels y^t_{pseudo} of unlabeled samples.

$$v_q = \begin{cases} 1, q = \text{argmax}\left(V_{Fc2}\right) \\ 0, \text{otherwise} \end{cases} \tag{11}$$

where $q \in \{0, 1, \cdots, g-1\}$.

Finally, the formula of J_{total} is as follows:

$$J_{total} = J_{label} + \gamma \times J_{mmd} + \beta \times J_{pseudo} \tag{12}$$

where γ and β are the weight parameters that represent the importance of MK-MMD distance and cross-entropy loss of pseudo-labeled samples, respectively.

4 Experiment settings

4.1 Datasets description

The bearing vibration data under three different loads in the JiangNan University (JNU) dataset [29], and the vibration data under four different loads in "12k Driving-end Bearing Fault Data" of Case Western Reserve University (CWRU) dataset [30, 31], which have richer fault types, are used to verify the superiority of the TCTL method. The sampling frequency of JNU is 50 kHz, and there are three loads with rotating speeds of 600, 800, and 1000 Revolutions Per Minute (RPM). Moreover, each load has four categories of vibration data. The sampling frequency of the CWRU is 12 kHz, and there are four loads with rotating speeds of 1797, 1772, 1750, and 1730 RPM, respectively. After removing some unavailable data, there are ten categories of vibration data under each load of CWRU.

Next, we use the sampling frequency and RPM to determine the number of sampling points n for one cycle of rotation in JNU and CWRU using formula (6). For JNU, the number of sampling points n_J for one cycle under three loads are $n_{J600} = 5000$, $n_{J800} = 3750$ and $n_{J1000=}3000$ respectively. For CWRU, the number of sampling points n_C for one cycle under four loads is $n_{C0} = 401$, $n_{C1} = 406$, $n_{C2} = 412$ and $n_{C3} = 416$ respectively. That n is greater than the number of sampling points collected by rotating one cycle under all load conditions, this paper sets $n_J = 5000$ and $n_C = 420$. The relevant information of the two datasets is shown in Table 1.

4.2 Cross-load fault diagnosis tasks and parameters settings

In this paper, the cross-load fault diagnosis task A → B is defined as training the model with labeled samples of

Table 1 Description of datasets

Datasets	Load	RPM	Points in a cycle	n	Category
CWRU	0	1797	401	420	10
	1	1772	406		
	2	1750	412		
	3	1730	416		
JNU	600	600	5000	5000	4
	800	800	3750		
	1000	1000	3000		

Table 2 The table of cross-load fault diagnosis tasks

Datasets	Cross-load fault diagnosis tasks		
CWRU	$0 \rightarrow 1$	$0 \rightarrow 2$	$0 \rightarrow 3$
	$1 \rightarrow 0$	$1 \rightarrow 2$	$1 \rightarrow 3$
	$2 \rightarrow 0$	$2 \rightarrow 1$	$2 \rightarrow 3$
	$3 \rightarrow 0$	$3 \rightarrow 1$	$3 \rightarrow 2$
JNU	$600 \rightarrow 800$	$600 \rightarrow 1000$	$800 \rightarrow 600$
	$80 \rightarrow -600$	$1000 \rightarrow 600$	$1000 \rightarrow 800$

load A and unlabeled samples of load B, and then testing the model's performance with unlabeled samples of load B. The arrow starts in the source domain and ends in the target domain. There are six cross-load fault diagnosis tasks for the JNU dataset, namely, $600 \rightarrow 800$, $600 \rightarrow 1000$, $800 \rightarrow 600$, $800 \rightarrow 1000$, $1000 \rightarrow 600$ and $1000 \rightarrow 800$. There are twelve types of cross-load fault diagnosis tasks in the CWRU dataset, namely, $0 \rightarrow 1$, $0 \rightarrow 2$, $0 \rightarrow 3$, $1 \rightarrow 0$, $1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 0$, $2 \rightarrow 1$, $2 \rightarrow 3$, $3 \rightarrow 0$, $3 \rightarrow 1$, $3 \rightarrow 2$. The cross-load fault diagnosis tasks of two datasets are shown in Table 2.

In this paper, all experiments are conducted five times to eliminate randomness interference. For the deep learning framework, PyTorch 1.9.0 + cu102 is used. The model's epochs for CWRU and JNU are set to 100. During the initial stage of training when $epoch \leq 2$, the model produces many mislabeled labels among the pseudo-labeled samples, which could harm the final performance of the model. To avoid this, the weight β of the cross-entropy loss for pseudo-labeled samples is set to 0. However, when $epoch > 2$, the model has good diagnostic performance after preliminary training and produces only a few mislabeled labels among the pseudo-labels. As a result, β is set to 0.25.

The weight γ of MMD distance is set to 0.7. The learning rate is 0.001, the batch size is 32, and the activation function is Relu. The first convolutional layer's filters are set to $Filter1 = 64$ to obtain more representative fault features, while the filters of the second, third, and fourth convolutional layers are set to $Filter2 = 32$. The convolution kernel is set to 20, the padding is set to "same", and the pooling kernel is set to 2. The convolution stride is set to 1 to obtain more comprehensive fault features.

4.3 Proportional periodic sampling strategy

Literature [5] confirms that the sliding stride considerably affects the performance of the diagnostic model. Following the work in [5], we set the sliding stride as half of the sampling length, represented as $S = L/2$. Next, we investigate how the sample length affects the performance of the model. The effect of sample length on model performance is reflected in two aspects. One is to affect the number of parameters in the model, and the other is to influence the diagnostic accuracy of the model. The average accuracy and the number of parameters of the proposed TCTL method with different sample lengths L on twelve tasks in the CWRU dataset are shown in Table 3.

Table 3 The effect of sample length on parameters and accuracy

Length	105	210	256	420	840	1024
Parameters	336k	502k	579k	**841k**	1514k	1808k
Average accuracy (%)	97.85	99.12	99.36	**99.41**	99.38	99.00

The bolding merely highlights the best performance

Three conclusions can be drawn from Table 3.

1. The larger the length of each sample, the higher the number of parameters in the TCTL model: For instance, a sample length of 420 has 846k parameters, which is significantly lower than the 1808K parameters for a sample length of 1024.
2. The sample length significantly affects the performance of the model: When the sample length L is set to $L_C = 1 \times n_C = 420$, the average accuracy of the model is the highest at 99.41%. However, when the sample length L_C is set to 1024 according to the convention, the average accuracy drops to 99.00%.
3. The hypothesis that the vibration data with a partial rotation period has sufficient fault information for high-performance cross-load fault diagnosis has been effectively confirmed: Even when the sample length is 256 (partial rotation cycle), the diagnostic accuracy is still high at 99.36%, slightly lower than 99.41% of a rotation cycle. Moreover, this hypothesis

Table 4 Unbalanced sample set sampling from CWRU

CWRU	Faulty samples									Normal samples	Total
	Ball			IR			OR				
	7	14	21	7	14	21	7	14	21		
Load = 0	582	579	579	576	579	580	579	579	581	1160	6374
Load = 1	577	580	578	579	579	577	581	580	579	2303	7513
Load = 2	577	579	580	580	579	579	577	579	581	2308	7519
Load = 3	577	580	580	584	578	579	582	579	579	2311	7529

Table 5 Unbalanced sample set sampling from JNU

JNU	Faulty samples			Normal samples	Total
	Ball	IR	OR		
Load = 600	2000	2000	2000	6004	12,004
Load = 800	2000	2000	2000	6004	12,004
Load = 1000	2000	2000	2000	6004	12,004

is also confirmed in the JNU dataset after extensive experiments, and the TCTL method performs excellently when the sample length of JNU is set as $L_J = 0.1 \times n_J = 500$ with the proportion $\alpha = 0.1$.

4.4 Description of sample sets

The vibration data collected under each load of CWRU includes one normal state and three types of faulty states—outer ring (OR), inner ring (IR), and ball. Each type of faulty state has three severity levels with fault diameters of 7, 14, and 21 miles, respectively. Similarly, the vibration data collected under each load of JNU also includes one normal state and three types of faulty states—OR, IR, and ball. Next, according to the proportion of $\alpha = 1$ and $\alpha = 0.1$, the sample lengths of CWRU and JNU are set as 420 and 500, respectively. Then, the sampling stride S of JNU and CWRU are set to $L/2$ to obtain more samples, that is, $S_C = 210$ and $S_J = 250$. Finally, 6374, 7513, 7519, and 7529 samples are sampled from all the vibration data of four loads in CWRU, and 12,004 samples are sampled from all the vibration data of each load in JNU. The precise quantity of ten categories of samples in the CWRU sample set and four in the JNU sample set are shown in Tables 4 and 5, respectively.

As can be seen from Tables 4 and 5, the sample sets constructed based on all data under each load have two characteristics. Firstly, the number of normal and various fault samples is unbalanced, with the number of normal samples exceeding the amount of each type of fault sample in both sample sets. Secondly, in the CWRU sample set, the total number of samples under *load* = 0 is more than 1000 less than that under other loads. The above two characteristics harm cross-load fault diagnosis.

Table 6 Comparison result on CWRU dataset

Method Length L	DANN 1024	JSWD 1024	FT-IDJ 512	1D-LDSAN 512	Our TCTL 420
0-1	99.53	97.56	97.94	99.65	99.61
0-2	95.50	96.06	97.46	99.45	99.59
0-3	84.43	90.09	98.63	99.64	99.85
1-0	97.39	94.53	97.66	99.79	98.83
1-2	98.50	96.58	98.79	99.66	99.59
1-3	88.17	97.57	98.49	96.27	99.86
2-0	92.70	92.95	96.36	98.90	99.51
2-1	93.76	97.40	99.02	99.45	99.66
2-3	90.72	97.31	99.73	99.80	99.86
3-0	79.19	88.87	94.71	97.11	97.83
3-1	76.71	91.65	97.53	95.94	99.15
3-2	86.37	93.79	98.83	99.19	99.59
Average accuracy	90.25	94.48	97.93	98.74	**99.41**
SD	7.13	2.94	1.29	1.37	**0.55**

The bolding merely highlights the best performance

5 Experiment result and analysis

Based on the sample sets constructed by proportional periodic sampling, this section first compares the TCTL method with several latest methods to verify its overall effectiveness, and then performs ablation experiments to demonstrate the effectiveness of each part of the TCTL method.

5.1 Overall comparison performance

In this part, the TCTL is compared with several latest methods on CWRU and JNU, and detailed results are shown in Table 6 and Table 7, respectively.

First, the TCTL method achieves the best performance. In Table 6, the average accuracies of the DANN ($L_c = 1024$),

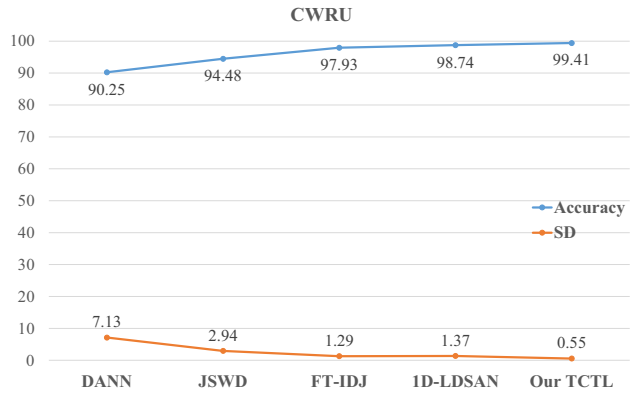

Fig. 2 Comparison of accuracy between TCTL and latest methods on CWRU

JSWD [29] ($L_C = 1024$), FT-IDJ [32] ($L_C = 512$), 1D-LDSAN [33] ($L_C = 512$) and TCTL ($L_C = 420$) methods on twelve cross-load fault diagnosis tasks are 90.25%, 94.48%, 97.93%, 98.74% and 99.41%, respectively, with standard deviations (SD) of 7.13, 2.94, 1.29, 1.37, and 0.55. In the CWRU dataset, the TCTL method uses the shortest sample length to achieve the best performance of cross-load fault diagnosis, with the highest accuracy and minimum standard deviation. In Table 7, the average accuracies of the BARTL [34] ($L_J = 2500$), JSWD [29] ($L_J = 1024$), SASMN [35] ($L_J = 1024$), FT-IDJ [32] ($L_J = 512$) and TCTL ($L_J = 500$) methods on the JNU dataset are 98.73%, 87.94%, 95.80%, 88.56% and 99.39%, respectively, with SD of 0.80, 3.89, 2.66, 6.64, 0.35. In the JNU dataset, the TCTL method also achieves the best performance of cross-load fault diagnosis based on the shortest sample length. Moreover, the average accuracies and the SD of TCTL and the latest method on CWRU and JNU are shown in Figs. 2 and 3, respectively.

Compared with some of the latest methods, TCTL shows best performance due to the following two reasons. (1) By taking the proportional periodic sampling to sample all vibration data under each load in the datasets, more short samples containing sufficient fault information are obtained, which can effectively enhance the stability of the TCTL method (less SD). (2) The CL-CNN proposed has a strong ability in feature extraction of short vibration samples, and the multi-target loss used can help the model learn better decision boundaries. Hence, TCTL has a stronger fault classification ability (higher accuracy).

Next, the TCTL has stronger universality. According to Tables 6 and 7, The average accuracies of the JSWD and FT-IDJ on CWRU are 94.48% and 97.93%, respectively, while the average accuracies on JNU are 87.94% and 88.56% respectively. The performance of the latest methods varies greatly on different datasets, indicating that their universality is poor. However, The average accuracies of the

Table 7 Comparison result on JNU dataset

Method	BARTL	JSWD	SASMN	FT-IDJ	Our TCTL
Length L	2500	1024	1024	512	500
600–800	99.50	90.32	98.38	86.01	99.18
600–1000	99.25	90.22	96.25	85.57	98.93
800–600	97.87	84.32	93.34	89.83	99.51
800–1000	99.12	91.27	97.35	88.96	99.75
1000–600	97.37	80.87	91.21	70.82	99.07
1000–800	99.25	90.365	98.29	89.84	99.88
Average accuracy	98.73	87.94	95.80	88.56	**99.39**
SD	0.80	3.89	2.66	6.64	**0.35**

The bolding merely highlights the best performance

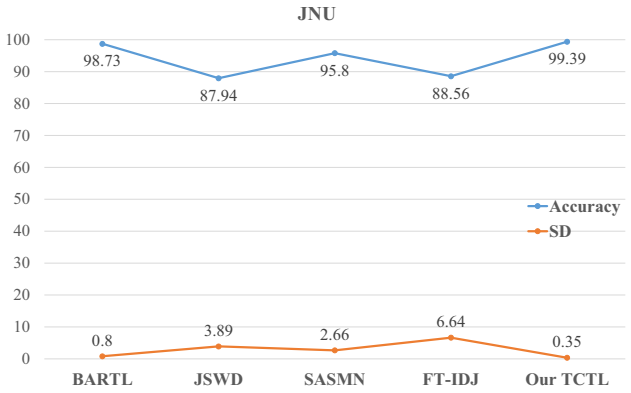

Fig. 3 Comparison of accuracy between TCTL and latest methods on JNU

TCTL method on CWRU and JNU are 99.41% and 99.39%, respectively, showing strong universality.

To sum up, on CWRU and JNU datasets, the proposed TCTL method is comprehensively superior to the latest methods, with higher diagnostic accuracy, more stable accuracy, and stronger universality, which makes TCTL have higher application significance.

5.2 Ablation experiments

The ablation experiments compare the following six methods, which are described below.

- A1: CL-CNN + cross-entropy loss $J_{total}(\gamma = 0, \beta = 0)$.
- A2: CL-CNN + dual-objective loss with cross-entropy and Coral metric.
- A3: CNN_base [16] + dual-objective loss J_{total} with cross-entropy and MK-MMD metric ($\gamma = 1, \beta = 0$).
- A4: CL-CNN + dual-objective $J_{total}(\gamma = 1, \beta = 0)$.
- A5: CL-CNN + balanced multi-objective loss $J_{total}(\gamma = 1, \beta = 1)$.

Table 8 Accuracy of ablation methods performed on CWRU

Method	A1	A2	A3	A4	A5	TCTL
0-1	93.34	95.85	98.51	99.06	99.65	99.61
0-2	92.05	92.17	99.74	99.12	99.57	99.59
0-3	78.44	84.77	93.72	99.46	99.87	99.85
1-0	96.23	98.09	93.33	95.77	99.61	98.83
1-2	99.51	99.55	100	98.62	99.59	99.59
1-3	93.80	92.69	99.94	99.85	99.87	99.86
2-0	92.97	94.32	93.72	94.41	99.67	99.51
2-1	92.36	92.53	99.61	99.52	99.66	99.66
2-3	94.18	99.83	100	99.87	99.87	99.86
3-0	78.96	89.55	92.18	92.21	98.84	97.83
3-1	80.04	86.60	99.09	99.59	96.76	99.15
3-2	86.80	93.75	100	99.57	99.59	99.59
Average accuracy	89.89	93.31	97.49	98.09	99.38	**99.41**
SD	6.80	4.73	3.19	2.42	0.83	**0.55**

The bolding merely highlights the best performance

Table 9 Accuracy of ablation methods performed on JNU

Method	A1	A2	A3	A4	A5	TCTL
600–800	97.66	98.04	98.33	98.88	99.8	99.18
600–1000	95.80	91.01	97.20	96.27	99.17	98.93
800–600	96.63	97.61	96.21	98.44	99.07	99.51
800–1000	99.38	98.63	99.01	99.56	99.31	99.75
1000–600	93.62	94.92	91.98	97.51	99.79	99.07
1000–800	99.56	99.43	97.78	99.37	99.03	99.88
Average accuracy	97.11	96.61	96.75	98.34	99.36	**99.39**
SD	2.07	3.13	2.53	1.14	0.35	**0.35**

The bolding merely highlights the best performance

- TCTL: CL-CNN + cross-load multi-objective loss $J_{total}(\gamma = 0.7, \beta = 0.25)$.

The accuracies of the six ablation methods on the cross-load fault diagnosis tasks of CWRU and JNU datasets are shown in Tables 8 and 9, respectively.

From the results in Table 8, we can see that, on twelve cross-load fault diagnosis tasks in the CWRU dataset, the average accuracies of the A1, A2, A3, A4, A5 and TCTL methods are 89.89%, 93.31%, 97.49%, 98.09%, 99.38%, and 99.41%, respectively. Moreover, the standard deviations (SD) of the six methods are 6.80, 4.73, 3.19, 2.42, 0.83, and 0.55, respectively. There is a rule of $A1 < A2 < A3 < A4 < A5 < TCTL$ in precision and stability. As the results shown in Table 9, on six cross-load fault diagnosis tasks of the JNU dataset, the average accuracies of the A1, A2, A3, A4, A5, and TCTL are 97.11%, 96.61%, 96.75%, 98.34%, 99.36%, and 99.39%, respectively. And the six methods' SD are 2.07, 3.13, 2.53, 1.44, 0.35, and 0.35, respectively. There is a rule of $A2 < A3 < A1 < A4 < A5 < TCTL$ in performance. Moreover, the average accuracies of TCTL and ablation methods on CWRU and JNU are shown in Fig. 4.

Next, we can draw the following conclusions:

- The cross-load fault diagnosis performance of the A1 method using only CL-CNN is poor: The vibration data distribution varies under different loads, which results in poor performance of the A1 method that solely uses CL-CNN. Thus, it is crucial to mitigate the impact of distribution differences by reducing the MK-MMD distance. Additionally, the A1 method's lowest accuracy in the two datasets is 89.89%, indicating that the distribution difference is minor.
- The CL-CNN has stronger feature extraction ability: Both A3 and A4 methods use a balanced dual-objective loss function considering the MK-MMD distance. In terms of performance, the A4 method using CL-CNN is better than the A3 method using CNN_base, which proves that the CL-CNN structure has stronger feature extraction ability in cross-load fault diagnosis tasks. The reason for this is that CL-CNN is capable of obtaining

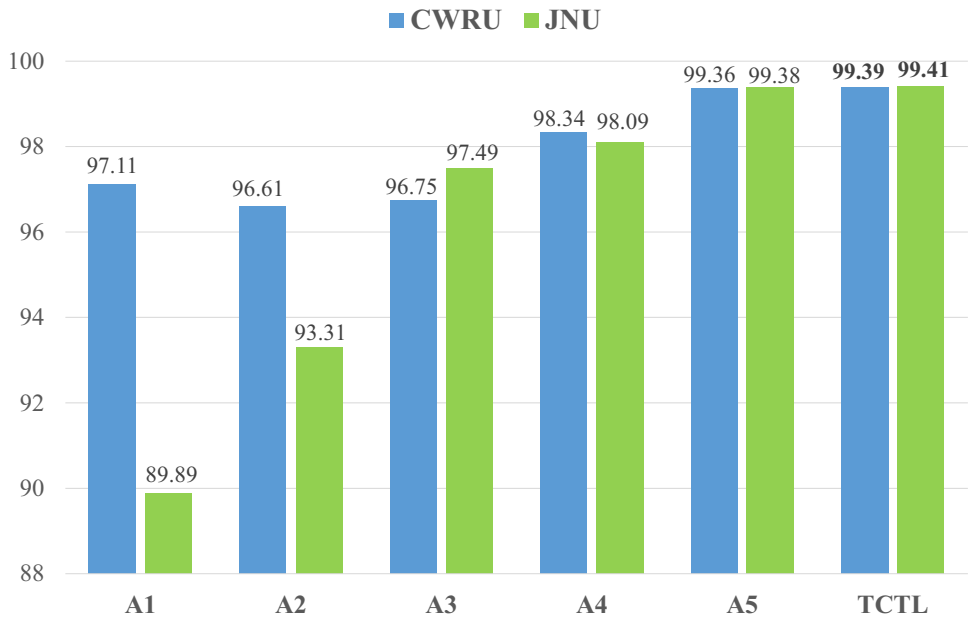

Fig. 4 Comparison of accuracy between TCTL and ablation methods on CWRU and JNU

more representative fault features by setting the Filter1 of the first convolutional layer to 64, and it is also able to achieve more detailed fault feature extraction by setting the convolution stride to 1.

- Cross-load multi-objective loss function can effectively supervise the model to perform cross-load fault diagnosis: The A1, A4, A5, and TCTL methods all use the CL-CNN, and there are rules with $A1 < A4 < A5 < TCTL$ in CWRU and JNU datasets. $A1 < A4$ shows that the negative effect caused by the distribution difference between various load samples can be reduced by minimizing the MK-MMD metric between them, which is helpful for CL-CNN to extract domain-invariant features. $A4 < A5$ indicates that the pseudo-labelling technology can exploit the fault information in the unlabeled target domain to obtain better performance for cross-load fault diagnosis. $A5 < TCTL$ demonstrates that using cross-load (imbalanced) multi-objective loss can achieve optimal fault diagnosis performance. Specifically, on the JNU dataset, there is a rule of $A2 < A3 < A1$. The reason for $A3 > A2$ may be that distribution alignment based on the MK-MMD metric is better for short sample sets than based on the Coral metric. The reason for $A1 > A3$ may be that the fault features of short samples in JNU are more prominent, and CL-CNN is particularly suitable for fault classification of short samples.

Finally, it is worth noting that among twelve cross-load fault diagnosis tasks in CWRU, the accuracy of task $3 \rightarrow 0$ is the worst at 97.83%. To further analyze the reasons for its poor accuracy, the confusion matrix is calculated when the TCTL performs task $3 \rightarrow 0$ with the worst accuracy of 94.75%, as shown in Fig. 5.

As can be observed in Fig. 5, the reason for the poor accuracy of task $3 \rightarrow 0$ is that many faulty samples with *label* = *IR*_14 are misclassified as normal samples with *label* = *Normal*. There are mainly two reasons. One is that the features extracted from these two types of samples are very similar, and the model is difficult to distinguish them. The other is that in the target domain, the number of normal samples with *label* = *Normal* is 1160, twice the number of fault samples with *label* = *IR*_14 (576), which makes the model more inclined to divide fault samples with

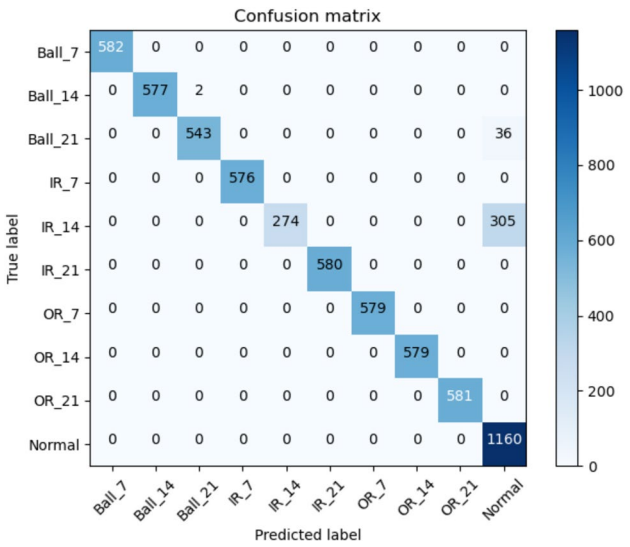

Fig. 5 Confusion matrix of the task $3 \rightarrow 0$

label = *IR*_14 into normal samples with *label* = *Normal*. The above reasons cause a non-negligible negative impact on the performance of the TCTL method.

6 Conclusion

This paper proposes the TCTL method for bearings cross-load fault diagnosis. The experimental results on two public datasets with multiple operating loads show that the TCTL performs best cross-load fault diagnosis performance through the proportional periodic sampling, fault feature extraction based on CL-CNN, and cross-load multi-objective loss function. Moreover, this paper confirms that vibration data from a partial rotation period has sufficient fault information for high-precision cross-load fault diagnosis. In the future, we will further optimize cross-machine fault diagnosis methods based on the proportional periodic sampling strategy to make it more practical and efficient.

Acknowledgements This work was supported by National Key R &D Program of China under Grant no. 2020YFB1713003, National Natural Science Foundation of China under Grant no. 62072169, the Key R &D Program of Changsha under Grant no. KQ2004005 and the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant no. QL20230099.

Data availability The data used in this paper are available in public repositories.

References

1. Siddique A, Yadava G, Singh B (2005) A review of stator fault monitoring techniques of induction motors. IEEE Trans Energy Convers 20(1):106–114
2. Huang W, Cheng J, Yang Y (2019) Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection. Mech Syst Signal Process 114:165–188
3. Qiu S, Cui X, Ping Z, Shan N, Li Z, Bao X, Xu X (2023) Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: a review. Sensors 23(3):1305
4. Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In: 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, pp 117–122
5. Zheng J, Yang C, Zheng F, Jiang B (2022) A rolling bearing fault diagnosis method using multi-sensor data and periodic sampling. In: 2022 IEEE international conference on multimedia and expo (ICME). IEEE, pp 1–6
6. Han H, Yang D, Qin J (2022) Hidden features extraction and amplification based on eigenvalue imaging and gray-level grouping for bearing fault diagnosis. Int J Mach Learn Cybern 13(11):3555–3568
7. Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK (2020) Applications of machine learning to machine fault diagnosis: a review and roadmap. Mech Syst Signal Process 138:106587
8. Zhao K, Jiang H, Li X, Wang R (2021) Ensemble adaptive convolutional neural networks with parameter transfer for rotating machinery fault diagnosis. Int J Mach Learn Cybern 12:1483–1499
9. Chen J, Hao L, Li H, Zhang L (2022) Time-frequency characteristics analysis and diagnosis of rotating rectifier faults in multiphase annular brushless system. IEEE Trans Ind Electron 70(4):3233–3244
10. Im J-H, Kang J-K, Hur J (2022) Static and dynamic eccentricity faults diagnosis in pm synchronous motor using planar search coil. IEEE Trans Ind Electron 70(9):9291–9300
11. Li M, Yan C, Liu W, Liu X, Zhang M, Xue J (2023) Fault diagnosis model of rolling bearing based on parameter adaptive avmd algorithm. Appl Intell 53(3):3150–3165
12. Chen L, Li Q, Shen C, Zhu J, Wang D, Xia M (2021) Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions. IEEE Trans Ind Inform 18(3):1790–1800
13. Guo L, Lei Y, Xing S, Yan T, Li N (2018) Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data. IEEE Trans Ind Electron 66(9):7316–7325
14. Sun S, Gao J, Wang W, Du J, Yang X (2023) Aanet: adaptive attention network for rolling bearing fault diagnosis under varying loads. Int J Mach Learn Cybern 1–15
15. Tao Y, Jun Z, Zhi-hao Z, Yi Z, Fu-qiang Z, Xiao-zhi G (2022) Fault detection of train mechanical parts using multi-mode aggregation feature enhanced convolution neural network. Int J Mach Learn Cybern 13(6):1781–1794
16. Zhao Z, Zhang Q, Yu X, Sun C, Wang S, Yan R, Chen X (2021) Applications of unsupervised deep transfer learning to intelligent fault diagnosis: a survey and comparative study. IEEE Trans Instrum Meas 70:1–28
17. Long M, Zhu H, Wang J, Jordan,M.I (2017) Deep transfer learning with joint adaptation networks. In: International conference on machine learning. PMLR, pp 2208–2217
18. Chen M, Fu D.Y, Narayan A, Zhang M, Song Z, Fatahalian K, Ré C (2022) Perfectly balanced: Improving transfer and robustness of supervised contrastive learning. In: International conference on machine learning. PMLR, pp 3090–3122
19. Long M, Cao Y, Wang J, Jordan M (2015) Learning transferable features with deep adaptation networks. In: International conference on machine learning. PMLR, pp 97–105
20. Li J, Huang R, Chen J, Xia J, Chen Z, Li W (2022) Deep self-supervised domain adaptation network for fault diagnosis of rotating machine with unlabeled data. IEEE Trans Instrum Meas 71:1–9
21. Michau G, Fink O (2021) Unsupervised transfer learning for anomaly detection: application to complementary operating condition transfer. Knowl Based Syst 216:106816
22. Zilong Z, Wei Q (2018) Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification. In: 2018 IEEE 15th international conference on networking, sensing and control (ICNSC). IEEE, pp 1–6
23. Ghorvei M, Kavianpour M, Beheshti MT, Ramezani A (2023) Spatial graph convolutional neural network via structured subdomain adaptation and domain adversarial learning for bearing fault diagnosis. Neurocomputing 517:44–61
24. Wang B, Wang B, Ning Y (2022) A novel transfer learning fault diagnosis method for rolling bearing based on feature correlation matching. Meas Sci Technol 33(12):125006
25. Ding Y, Jia M, Zhuang J, Cao Y, Zhao X, Lee C-G (2023) Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions. Reliab Eng Syst Saf 230:108890
26. Wang R, Huang W, Wang J, Shen C, Zhu Z (2022) Multisource domain feature adaptation network for bearing fault diagnosis

under time-varying working conditions. IEEE Trans Instrum Meas 71:1–10
27. Wang H, Sun W, He L, Zhou J (2022) Rolling bearing fault diagnosis using multi-sensor data fusion based on 1d-cnn model. Entropy 24(5):573
28. Gretton A, Sejdinovic D, Strathmann H, Balakrishnan S, Pontil M, Fukumizu K, Sriperumbudur BK (2012) Optimal kernel choice for large-scale two-sample tests. Adv Neural Inf Process Syst 25
29. Chen P, Zhao R, He T, Wei K, Yang Q (2022) Unsupervised domain adaptation of bearing fault diagnosis based on join sliced wasserstein distance. ISA Trans 129:504–519
30. Smith WA, Randall RB (2015) Rolling element bearing diagnostics using the case western reserve university data: a benchmark study. Mech Syst Signal Process 64:100–131
31. Hendriks J, Dumond P, Knox D (2022) Towards better benchmarking using the cwru bearing fault dataset. Mech Syst Signal Process 169:108732
32. Qian C, Jiang Q, Shen Y, Huo C, Zhang Q (2021) An intelligent fault diagnosis method for rolling bearings based on feature transfer with improved densenet and joint distribution adaptation. Meas Sci Technol 33(2):025101
33. Zhang R, Gu Y (2022) A transfer learning framework with a one-dimensional deep subdomain adaptation network for bearing fault diagnosis under different working conditions. Sensors 22(4):1624
34. Hu Q, Si X, Qin A, Lv Y, Liu M (2022) Balanced adaptation regularization based transfer learning for unsupervised cross-domain fault diagnosis. IEEE Sens J 22(12):12139–12151
35. Zhang D, Wei Y, Wang B, Liu S (2021) Scale adaptive subdomain matching network for bearing fault diagnosis. Meas Sci Technol 33(2):025006