IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 February 2024, accepted 7 March 2024, date of publication 14 March 2024, date of current version 22 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3375400

RESEARCH ARTICLE

Rolling Bearing Fault Diagnosis Using Deep Transfer Learning Based on Joint Generalized Sliced Wasserstein Distance

NA LEI[1], JIPENG CUI[1], JICHENG HAN[2], XIAN CHEN[3], AND YOUFU TANG[1]

[1]Mechanical Science and Engineering Institute, Northeast Petroleum University, Daqing 163318, China
[2]Guangdong Zhuhai Golden Bay LNG Ltd., Zhuhai, Guangdong 519090, China
[3]BYD Company Ltd., Shenzhen, Guangdong 518119, China

Corresponding author: Na Lei (leina9316d@163.com)

This work was supported by the Natural Science Foundation of Northeast Petroleum University under Grant 2018QNL-28.

ABSTRACT The big data of rolling bearings for on-site monitoring usually contains very few failure samples and easily affected by noise and monitoring errors, so it is difficult to extract and identify useful fault information in normal samples. In addition, the rolling bearing samples of field test are un-labeled dataset of unknown fault types. If the existing fault diagnosis approaches are directly used for extraction and identification, it is easy to cause misjudgment or missing judgment. To solve this problem, a novel intelligent fault diagnosis approach using deep transfer learning based on joint generalized sliced Wasserstein distances (JGSWD) deep transfer learning is proposed. Firstly, the joint discrepancy between the data from real-case scenarios (DRS) and the data from laboratory equipment (DLE) is minimized by calculating the generalized sliced Wasserstein distances. Following, the marginal and conditional dataset distribution between source domain and target domain is balanced by using the dynamic domain alignment. Then, the top K correlated pseudo labels are calculated for reducing the conditional distribution and improving better transfer capability. Finally, the deep transfer learning from laboratory bearing dataset to field bearing dataset is carried out. The result shows that the proposed JGSWD method can achieve 97.56% fault diagnosis accuracy, which is higher than the other methods. Therefore, it is a practical semi-supervised learning approach for bearing fault diagnosis with small samples.

INDEX TERMS Fault diagnosis, rolling bearing, transfer learning, generalized sliced Wasserstein distance.

I. INTRODUCTION

Rolling bearing is one of the critical parts in rotating machinery which has been widely applied in many necessary engineering fields, such as aerospace, rail transportation, petrochemical industry and so on. Rolling bearing usually works in complex and harsh environments with heavy loads, high temperatures, corrosion and high operating rates to fulfill actual production needs, in addition, the complex mechanical structure and changing operating conditions lead to the failure of rolling bearings [1]. Many serious consequences will be brought by the damage of rolling bearing for example, the working performance and motion accuracy might be reduced and increased vibration and wear, or even lead to production stoppage, economic losses and casualties [2]. Therefore, it is significant to develop a more scientific and efficient method to diagnose the fault of the rolling bearings.

Over the recent years, many fault diagnosis methods have been proposed that provide useful solutions for bearing diagnosis problems. It can be rouvaluable,ghly divided into the following categories: mechanism-based fault diagnosis methods, signal processing-based fault diagnostic methods, and data-driven fault diagnosis methods. The mechanism-based fault diagnosis method is mainly to establish a dynamic model of the rolling bearing system to analyze the mechanism

The associate editor coordinating the review of this manuscript and approving it for publication was Hassen Ouakad.

of its fault. The advantage is that it can reveal the essence of the rolling bearing fault, but the disadvantage is that the method involves more theoretical knowledge and slower research progress.The fault diagnosis method based on signal processing is to analyze and study the vibration signal of the rolling bearing, its advantage is that it can find a more accurate fault location, but it is difficult to diagnose the fault of the rolling bearing accurately due to the noise of the signal [3].The data-driven fault diagnostic method is to use a large amount of rolling bearing fault data to establish and train a network model to classify and identify its faults. The advantage is that ideal fault identification results can be achieved without a deep and specific understanding of the fault diagnosis object, but the disadvantage is that a large amount of labeled data is required to train the neural network. The methods are mainly divided into intelligent diagnostic methods based on traditional machine learning and intelligent diagnostic methods based on deep learning. This paper specifically researches the application of clever methods based on deep learning in fault diagnosis of rolling bearings. Deep learning fault diagnosis methods have the advantage of classifying and predicting the bearing fault among machine learning [4], [5]. Nevertheless, deep learning methods always lead to bad performance with unlabeled data. Data collected from real-case scenarios (DRS) are always unlabeled, so it isn't easy to train a reliable intelligent diagnostic model. It is unrealistic to label the data with a large amount of workforce [6].

It is convenient to collect sufficient labeled data from laboratory equipment (DLE) by simulating the work of real-case scenarios. Thus, a deep learning model can be simply built to diagnose the DRS with the help of the massive labeled data. However, the result is unsatisfactory if using the deep learning model to train the DRS directly when the two data have different feature distribution.

Transfer learning, which can provide an excellent deep learning methodology for tackling the above problem. The concept of transfer learning is transferring knowledge from one domain to a different but related domain. In other words, the DLE is set as the source domain with a rich-labeled and the DRS is designated as the target domain with a scarce-labeled, the transfer learning utilizes the knowledge from the source domain to improve the fault prediction performance of the model for the target domain [7], [8].

Intelligent fault diagnosis of rolling bearings can be implemented through various strategies. Based on deep learning, [9] proposed a technique based on multi-synchrosqueezing transform (MSST) and deep residual convolution neural network to identify weak faults of rolling element bearing (REB) accurately. Reference [10] had been implementing the AdaBN into the deep convolutional neural networks which increase the efficiency of fault diagnosis. Reference [11] built a Dual-stage Attention-based Recurrent deep learning structure utilizes the image processing to alleviate the imbalance rate of rolling bearings datasets. Reference [12] proposes a deep residual network based IFD method of planetary gearboxes in cloud environments, it is valuable to incorporating the beneficial characteristics of different algorithms for promoting the efficiency of mechanical fault diagnosis and prediction [13]. Deep meta-learning and variational auto-encoder (DML-VAE) are applied for coupling fault diagnosis of rolling bearing under variable working conditions in this paper [14]. The above intelligent fault diagnosis methods have significantly increased the accuracy and efficiency of rolling bearing fault diagnosis, but these methods cannot perform unsupervised learning and rely on a large amount of labeled data. Reference [15] proposed an intelligent fault diagnosis method base on small imbalanced data but it is not to finish unsupervised diagnosis. Based on transfer learning, the central objective of the transfer learning is minimizing distribution distance between the source domain and the target domain, which is called domain adaptation [16]. Many effective methods have been proposed for domain adaptation. Reference [17] propose a feature-based transfer neural network (FTNN) with the help of multi-layer domain adaptation and pseudo label learning, the FTNN model can reduce the feature discrepancy between the source domain and the target domain. Meanwhile, the FTNN model proposed by Yang can diagnose the fault of rolling bearings with massive unlabeled data effectively.Still.the pseudo label learning offered,in the FTNN model can't get precise label. [18] Propose a multi-layer domain adaptation CNN network using a multi-kernel maximum mean discrepancies (MMD) metric. Reference [19] using IJDA mechanism and I-Softmax loss, a deep discriminant transfer learning network (DDTLN) was constructed to realize fault migration diagnosis. Reference [20] propose a transfer learning approach with LSTM neural networks for bearing fault diagnosis, and which uses the joint distribution adaptation (JDA) [21] to reduce the differences in probability distributions between a source domain dataset and target domain dataset. The approach is robust greatly to noise and has a great performance under different noise levels. Reference [22] constructs a multiple-scale feature learner with the help of the modified ResNet-50 to shorten the conditional distribution distance between the two domains for bearing fault diagnosis, the model can extract discriminative and robust features from the two domains and avoid the loss of information, the conditional distribution calculation of the above two methods without using measurement criteria and quickly leads to the problem of gradient disappearance. Reference [23] propose a applications of unsupervised deep transfer learning to CWRU, which use the JAN, DAN as the measurement criteria between the source domain and the target domain, [24] propose a self-attention ensemble lightweight transfer learning model achieve higher accuracy with small training samples, but these methods can't conduct the end-to-end training. Reference [25] propose a transfer learning approach with adversarial networks for bearing fault diagnosis combine with the knowledge mapping-based method, the results show the proposed method proves ideal generalization ability, but the accuracy of classification needs to be improved. Reference [26] propose an unsupervised domain adaptation with Join Sliced Wasserstein Distance to provide an end-to-end training and higher accuracy, but the calculation of Sliced Wasserstein Distance is very complicated.

In this paper, enlightened by generative adversarial nets (GAN) [27] and Generalized Sliced Wasserstein Distances [28], an approach namely Joint generalized sliced Wasserstein distances guided transfer learning (JGSWD) is proposed in this article that utilizes the generalized sliced Wasserstein distance for decreasing complex computational cost and create top K correlated pseudo labels to calculate a better conditional distribution. The JGSWD is consisting of four structures: a feature generator composes with CNN [29], a joint feature, a discriminator and a classifier, the network of the JGSWD will be optimized to minimize different feature distribution between the two domains. The main contributions of this paper are as follows

1) Joint generalized sliced Wasserstein distances guided transfer learning(JGSWD) is proposed in this article as an application of unsupervised intelligent fault diagnosis.

2) The joint distribution discrepancy is measured via calculating generalized sliced Wasserstein distances which decrease computational costs and improve the training speed of the model.

3) The top K correlated pseudo labels are proposed for ensuring more accurate prediction of target domain labels.

4) A dynamic domain alignment is introduced in the article for aligning marginal and conditional distributions and using the loss to tune hyperparameters.

II. THEORETICAL BASIS

A. THE BASIC DEFINITIONS OF TRANSFER LEARNING

Transfer learning method can apply knowledge and experience in familiar fields to related unknown fields, and can solve the problem that training and test data are subject to different distributions, resulting in difficult training and low performance of the model. Therefore, it is suitable for fault diagnosis in fields such as small training data, scene changes and task changes.

In a fault diagnosis problem, there is a labeled source dataset $X^s = \left\{\left(x_i^s, y_i^s\right)\right\}_{i=1}^{n^s}$ of n^s samples from the source domain D^s, domain D^s is composed of two parts: a feature space X and a marginal distribution $P(X)$. In other words, $D^s = \{X, P(X)\}$. And the symbol denotes an instance set, which is defined as $X = \{x_i \in X\}$, and an unlabeled target dataset $X^t = \left\{\left(x_j^t\right)\right\}_{j=1}^{n^t}$ of n^t samples from the target domain D^t, $D^t = \{X\}$. It is assumed that the two domains follow different data distributions. The main point of the problem is to extract the transferable features from the two domains. Then the transfer learning method proposed in the paper is mainly to minimize the joint discrepancy between the two domains.

B. GENERALIZED SLICED WASSERSTEIN DISTANCES

There are many existing works using Wasserstein distance in deep networks for transfer learning and achieving excellent performances. The advantage of choosing Wasserstein distance as the metric criterion to calculate the joint feature distribution is that even if the feature distributions of the source domain and the target domain do not overlap or overlap very little, the distance can still measure the distance of the two distributions in space, and will not cause the problem of gradient disappearance. However, it is difficult to measure the high dimensional distribution, and there are many methods to accelerate the calculation of Wasserstein distance, such as slicing Wasserstein distance, which is calculated by the probability distribution related to the linear fragment and pull transform [30]. Although slicing Wasserstein distances requires less computational complexity, the number of randomly selected linear projections is very large, and there is no guarantee that the selected linear projections will provide an effective assessment of feature distribution distances.Inspired by the generalized radon transform, a new measurement distance called generalized sliced Wasserstein distance is proposed in the paper by extending the linear slicing to the non-linear slicing of probability measures to reducing the number of required projections.

The generalized radon transform is evolved from the classical radon transform which extends the integration on a hyperplane to the integration on a hypersurface like the manifolds. The classical radon transform can be defined as follows:

$$\mathcal{R}P_{S/T}(t, w) = \int_{R^d} P_{S/T}(x)\delta(t - \langle x, w\rangle)dx \tag{1}$$

The $\mathcal{R}P_{X/T}$ represents the infinite set of integrals projecting the high-level distribution onto the low-dimensional distribution hyperplane of R^d. The $P_{S/T}$ means the probability distribution function of the source and target domains. $\forall w \in S^{d-1}$, and the S^{d-1} stands the unit sphere of d-dimensional space. $\forall t \in \boldsymbol{R}$, $\delta()$ represents one-dimensional Dirac delta function. $\langle x, w\rangle$ represents the Euclidean inner-product.

The generalized radon transform can be defined as:

$$GP_{S/T}(t, w) = \int_{R^d} P_{S/T}(x)\delta(t - g(x, w))dx \tag{2}$$

The radon transform is a particular case of the generalized radon transform when $g(x, w) = \langle x, w\rangle$, meanwhile, the generalized sliced Wasserstein distances can be defined via using the generalized radon transform, W denote the d-dimensional Wasserstein distances:

$$GSW(P_S, P_T) = \int_{S^{d-1}} W(GP_S(., w), GP_T(., w))d\omega \tag{3}$$

C. JOINT DISTRIBUTIONS BETWEEN DIFFERENT DOMAINS

In order to solve the problem of different probability distributions between the training data and the test data in mechanical fault diagnosis, the focus of the approach is to measure the joint distribution discrepancy between the DLE and DRS. The joint distributions are composed by the marginal distributions and conditional distributions, The marginal discrepancy D_{mar} between the two domains is described as: $D_{mar} = GSW(P_S, P_T)$, however, the target domain is unlabeled data and the trained feature generator under the source domain data performs poorly in the domain. In order to predict higher accurate target label, the pseudo labels are proposed to keep unlabeled target data in lots of researches. In this paper, the top k correlated labels are used to mark the unlabeled target data. The conditional discrepancy D_{con} between the two domains is described as: $D_{con} = \sum_{c=1}^{c} \left[\frac{1}{m'} \sum_{m=1}^{m'} GSW\left(P_S\left(y_m^{S^c} \mid x_m^{S^c}\right), P_T\left(y_m^{T^c} \mid x_m^{T^c}\right)\right)\right]$, which the c denotes the number of categories, m' is the number of samples in the Cth category in the source or target domain. The $y_m^{T^c}$ can be calculated via the top k correlated labels.

D. TOP K CORRELATED LABEL

Aiming at the problem that the target domain sample data lacks labels and the conditional distribution is difficult to calculate, the top K correlated label is proposed in the paper, which calculates the similarity of the sample feature space in the source domain and the target domain.The number is denoted as similar samples. The similarity of probability between the x_i^T and the x_j^T will be estimated by, which can be written as:

$$p_{i|j}^{\mathcal{M},\mathcal{C}} = p_{i|j}^{\mathcal{M}} \bullet p_{i|j}^{\mathcal{C}} = \frac{\| F(\mathcal{X}_i^T) - F\left(\mathcal{X}_j^T\right) \|_2 + \| F(\mathcal{X}_i^T)^{\mathcal{C}} - F\left(\mathcal{X}_j^T\right)^{\mathcal{C}} \|_2}{\sum_{i=1}^{n^t}\left(\| F(\mathcal{X}_i^T) - F\left(\mathcal{X}_j^T\right) \|_2 + \| F(\mathcal{X}_i^T)^{\mathcal{C}} - F\left(\mathcal{X}_j^T\right)^{\mathcal{C}} \|_2\right)} \tag{4}$$

The $\bullet$ denote the probability conjunct function and the $F(\cdot)$ represent features generator function, meanwhile, this probability matrix in the target domain can be expressed by $M = \otimes_{i=1}^{m_t} \oplus_{j=1}^{m_t} p_{i|j}^{\mathcal{M},\mathcal{C}}$, the K label measures the index of the most similar sample in the target domain and the y_k represent the first k index probability matrix that is operated by first descending sorting. Thus the target label loss can be written as:

$$\mathcal{L}_{\mathcal{K}}\left(\mathcal{Y}_{pred}, \mathcal{Y}_{\mathcal{K}}\right) = \frac{1}{\mathcal{M}_{\mathcal{T}}} \sum_{i/j=1}^{\mathcal{M}_{\mathcal{T}}} \mid \mathcal{Y}_{pred}^{i/j} - M\left(\mathcal{Y}_{\mathcal{K}}^{i/j}\right) \mid \tag{5}$$

The $\mathcal{Y}_{pred}^{i/j}$ represent the predicted label from the classifier, the $\mathcal{Y}_{\mathcal{K}}^{i/j}$ represent top K correlated label matrix and the M is aimed to select the most frequent labels in the correlated label matrix. Therefore, the joint distribution can be defined as follows:

$$\mathcal{D}_j = D_{mar} + D_{con} = GSW\left(\mathrm{P}_S, \mathrm{P}_T\right) + \sum_{c=1}^{c}\left[\frac{1}{m'}\sum_{m=1}^{m'} GSW\left(\mathrm{P}_S\left(y_m^{s^c} \mid x_m^{s^c}\right), \mathrm{P}_T\left(M\left(y_{\mathcal{K}}^{m}\right) \mid x_m^{t^c}\right)\right)\right] \tag{6}$$

E. DYNAMIC DOMAIN ALIGNMENT

Many previous methods only considered the marginal distributions alignment and the conditional distributions alignment separately, or aligning the marginal distribution and conditional distribution in a way with the same specific gravity, without considering that the importance of these two ways may be different in real applications, which cause the problem of declining the transfer ability between two other domains. The marginal distributions and the conditional distributions will be dynamically aligned by calculating a balance factor in the paper. The $\mu \in [0, 1]$ is denoted an adaptive balance factor to align the marginal distribution and conditional distribution dynamically, the formula can be expressed as: $\overline{\mathcal{D}} = \mu D_{mar} + (1-\mu) D_{con}$. The adaptive balance factor is updated by calculating the domain alignment loss, which will be introduced in the next section.

III. ESTABLISH THE DEEP MIGRATION MODEL OF JOINT COUNTERMEASURE

A. THE COMPOSITION OF JGSWD NETWORK

The data obtained in laboratory equipment or the data with tags on the real-case scenarios are used as the labeled source domain (DLE), and the unlabeled data measured by other laboratory equipment or the real-case scenarios are used as the target domain (DRS), it is obviously that the two domains have different distributions. Inspired by the adversarial network, a new approach called Joint generalized sliced Wasserstein distances guided transfer learning (JGSWD) is proposed, the transfer learning overview framework is given in Figure 1, of which consist a feature generator, a joint feature, a domain discriminator and a classifier. The feature generator which can be implemented by a convolutional neural network with batch normalization (BN) layer to extract the domain invariant features, which can speed up the training process and improve network generalization performance, and better preserve the different feature distributions of each domain, the detailed layers and size of feature generator is illustrated in Table 1. Only the marginal features can be extracted in the feature generator in general, and many methods ignore the calculation of the conditional features. A joint feature is added to general diagnostic neural network structure, which is used for computing the joint distribution, which is composed of conditional distribution and marginal distribution. The conditional distribution can be calculated by generating label of the DRS via utilizing the top K correlated label, and the prediction accuracy of labels can be

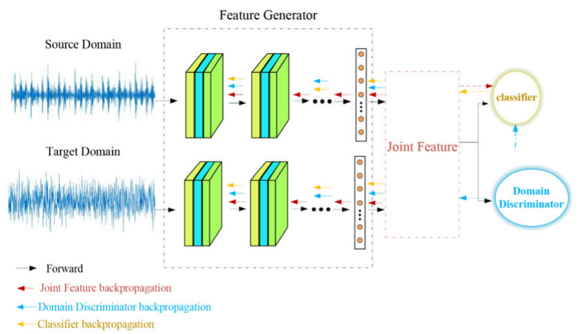


FIGURE 1. The framework of JGSWD.

TABLE 1. The parameters of the CNN.

Layers	Parameters
Input	1*1024
Conv1	16*1010 kernel-size=15 BN,RELU
Conv2	32*1008 kernel-size=3 BN,RELU
Max Pooling	32*504 kernel-size=2
Conv3	64*502 kernel-size=3 BN,RELU
Conv4	128*500 kernel-size=3 BN,RELU
Adaptive Max Pooling	128*4
Fc1	256*1

improved. The dynamic domain alignment is used to balance the marginal and conditional distribution to improve the transferability of the model. Then the joint distributions can be calculated by the joint generalized sliced Wasserstein distance criterion to decrease computational costs and improve the training speed of the model. The domain discriminator is used to distinguish the different domains for more comprehensive domain adaptation, which forms an adversarial domain loss function. In addition, the final label would be predicted in the classifier.

B. TRAINING PROCESS

The feature generator can be expressed by a function F with parameter θ_g that process the input big data directly. Input the labeled source domain into the feature generator to pre-training, obtain the initial parameters, and then input the unlabeled target data into the pre-training network. Given an instance $x \in R^m$ from either domain, the feature extractor learns a function $F_g : R^m \rightarrow R^d$ that maps the instance to a d-dimensional representation with corresponding network parameters θ_g. The source domain data set $\{X^s, Y^s\}$ consists of the original data X^s and its related label Y^s composition. While the target domain data $\{X^t\}$ consists of the unmarked target domain data X^t composition. To align joint distributions between different domains, the joint feature has been used to estimate the joint generalized sliced Wasserstein distances by utilizing the Eq 6. Then build the classifier discriminative loss function. The discriminator $F_c : R^d \rightarrow R^l$ is used to compute the softmax prediction with parameter θ_c where is the number of classes. The classified losses are written as follows:

$$\mathcal{L}_c\left(F_c\left(F_g\left(x^s\right)\right), y^s\right) = -\frac{1}{\mathcal{M}^s}\sum_{i=1}^{\mathcal{M}^s}\sum_{c=1}^{C} 1\left(y_i^{sc}\right)\cdot \log F_c\left(F_g\left(x_i^{sc}\right)\right) \tag{7}$$

Simultaneously, the function F_d is learned in the domain discriminator to distinguish the different domains with parameters θ_d and the loss function is defined as follows:

$$\mathcal{L}_d\left(F_g\left(x^s\right), F_g\left(x^t\right)\right) = -\frac{1}{\mathcal{M}^s}\sum_{i=1}^{\mathcal{M}^s}\log\left(1-F_d\left(F_g\left(x_i^s\right)\right)\right) - \frac{1}{\mathcal{M}^t}\sum_{i=1}^{\mathcal{M}^t}\log\left(F_d\left(F_g\left(x_i^t\right)\right)\right) \tag{8}$$

In the end, a dynamic distribution alignment is applied to balance the marginal and conditional distribution to improve the generalization ability of the model, the domain alignment loss is expressed as follows:

$$\mathcal{L}_{\mathcal{DA}} = \arg\min \mathcal{L}_c\left(F_c\left(F_g\left(x^s\right)\right), y^s\right) + \eta \parallel F_c \parallel_K^2 + \lambda\left(\mu D_{mar} + (1-\mu) D_{con}\right) \tag{9}$$

The η, λ denote regularization parameters, $\parallel F_c \parallel_K^2$ represent a squared norm of F_c, and μ denote the balance factor

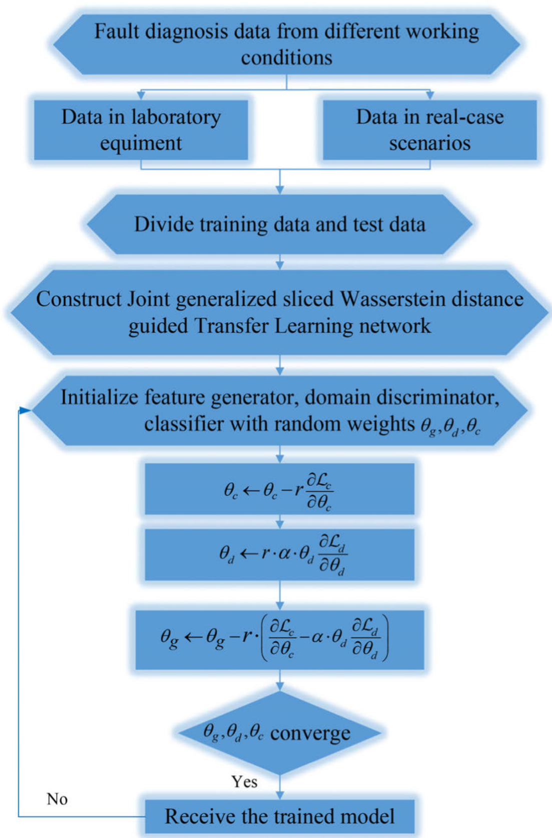


FIGURE 2. The Flow Chart of JGSWD.

to align the marginal distribution and conditional distribution dynamically. The detailed algorithm of the combination is given in Figure fig2 and the entire objective function of the model proposed in the paper is expressed as follows:

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_{\mathcal{K}} + \mathcal{L}_d + \mathcal{L}_{\mathcal{DA}} \tag{10}$$

IV. EXPERIMENTAL

To evaluate the ability of the transferable ability of the proposed approach in the field of mechanical fault diagnosis, experimental verification under different data is being conducted. Furthermore, comparing our approach with other fault diagnosis methods, JGSWD achieves better performance.

A. DATA DESCRIPTION

Bearing Data of Xi'an Jiaotong University [31]: The rolling bearing failure test bench comprises AC motor, motor speed controller, support shaft, two support bearings (heavy roller bearings), hydraulic loading system, etc., as shown in Figure 3. The bearing type is LDK UER204. The bearing dataset is composed by three different speed conditions: 2100, 2250, 2400, each speed condition contains inner ring, outer ring and cage faults, etc. The details information can be seen in table 2.

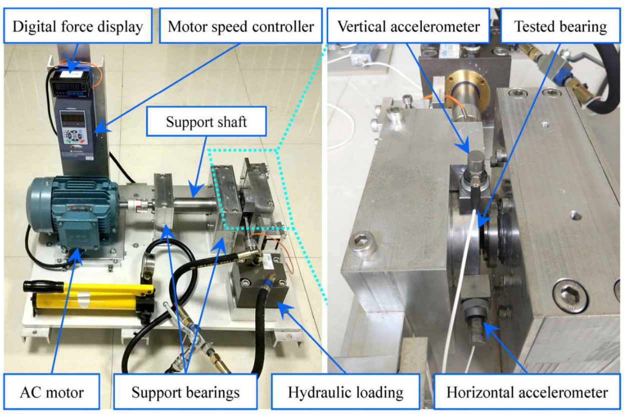


FIGURE 3. The rolling bearing failure test Bench Xi 'an Jiaotong University.

TABLE 2. Bearing data of Xi'an Jiaotong University.

Data	A	B	C
Rotating Speed	2100	2250	2400
Health Status	Health/Defect	Health/Defect	Health/Defect
Defect Location	IF/OF/RF/CF	IF/OF/RF/CF	IF/OF/RF/CF

Bearing Data of Southeast University: The test bed consists of a motor, a motor controller, a planetary gearbox, a reduction gearbox, a brake and a brake controller, as shown in Figure 4. The data set is mainly divided into bearing data set and gear box data set. Each fault type corresponds to the two working conditions: speed 1200rpm-load 0 Nm and speed 1800rpm-load 7.32Nm. The gearset and bearingset are described in table 3.

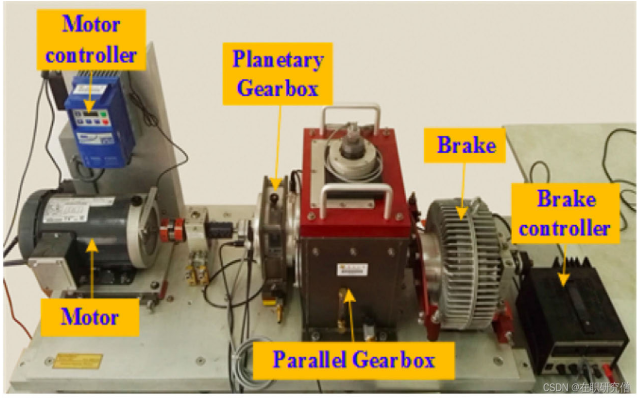


FIGURE 4. The test bed of Southeast University.

TABLE 3. Dataset from Southeast University.

Data	A	B
Rotating Speed	20_0/30_2	20_0/30_2
Health Status	Health/Defect	Health/Defect
Dataset	Bearingset	Gearset
Defect Location	IF/OF/RF/CF	Chip/Miss/Surface/Root

Bearing Data of real-case scenarios [32]: The dataset is collected from a compressor with a bearing power of 500kW under changing loads and operating conditions, as shown in Figure 5. The fault category is divided into fatigue, pitting,

FIGURE 5. The compressor of real-case scenarios.

TABLE 4. Bearing data of real-case scenarios(DRS).

Data	DRS
Rotating Speed	800-1500
Health Status	fatigue, pitting, plastic deform and indentations
Defect Location	IF/OF

plastic deform, indentations in inner race and outer race. The data detail is shown in Table 4.

B. COMPARATIVE APPROACHES

The proposed approach is mainly compared with conditional adversarial domain adaptation (CDAN), domain adversarial neural network (DANN) [33], deep correlation alignment (CORAL) [34] and joint maximum mean discrepancy (JMMD). The purpose of our proposed method (JGSWD) and the comparison method is mainly to reduce the domain discrepancy.

1) S-ONLY

The source domain data is used to train the convolutional network and tested on the target test data directly without using the domain adaptive method.

2) CDAN

The CDAN is used to extract the high-level features of the source domain and the target domain and capture the multimodal structure of data features by multi-linear mapping, which measures the divergence between two probability distributions.

3) DANN

The DANN architecture includes a deep feature extractor, a deep label predictor and a domain classifier. The minimax optimization is used to solve the data coming from similar but different distributions.

4) CORAL

The CORAL uses a linear transformation method to align the second-order statistical features of the source and target domain distributions, it relies on linear transformations and is not end-to-end training.

5) JMMD

JMMD is a deep transfer learning method adopted by Long et al. [35], further proposes joint maximum mean discrepancy to align the domain shift and comprehensively utilizes the low-level and high-level features of the convolutional neural network to combine Domain adaptation to better realize feature matching.

C. IMPLEMENTATION DETAILS

The convolutional network in this article is composed by three convolutional layers with Batch Normalization and a fully connected layer as the basic network architecture, the details can be seen in Table 1. In contrast with our method, the S-only without the domain critic structure, the MMD method mainly changes the measurement metric of our proposed approach to the MMD metric. The DANN tries to distinguishing the source and target features by using an adversarial representation learning approach. The MLDAN method used the Multi-layer MMD metric based on MMD.

For XJTU-SY bearing Data, there are three different working conditions: A, B, C, six transfer tasks for bearing dataset have been seen in Table 5, for the bearing dataset and gear dataset from Southeast University, there are two transfer tasks have been constructed in Table 6. The XJTU-SY bearing dataset and the dataset from Southeast University can be chosen as the DLE, the transfer learning from DLE to DRS can be shown in Table 7.

TABLE 5. Six transfer tasks for bearing dataset.

Transfer task	A → B	A → C	B → A	B → C	C → A	C → B
Source speed	2100	2100	2250	2250	2400	2400
Target speed	2250	2400	2100	2400	2100	2250
Source sample	3125	3125	3150	3150	3134	3134
Target sample	3150	3134	3125	3134	3125	3150

TABLE 6. Transfer task from bearingest to gearset.

Transfer task	bearingset → gearset
Source speed	30_2
Target speed	30_2
Source sample	3492
Target sample	3492

TABLE 7. Transfer learning from DLE to DRS.

Transfer task	XJTU-SY → DRS	gearset → DRS
Source speed	2100,2250,2400	30_2
Target speed	800-1200	800-1200
Source sample	6200	3492
Target sample	3000	3000

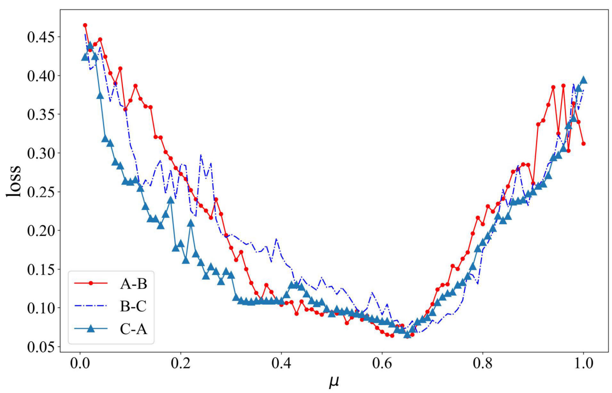
V. RESULTS AND DISCUSSION

A. ADAPTIVE WAVEFORM DECOMPOSITION ALGORITHM

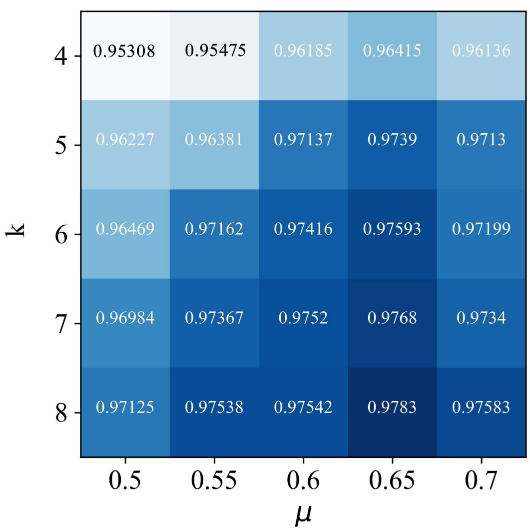
The selection of optimal parameter: A→B, B→C, C→A transferring learning task are choosed as the object of parameter selection experiment to select the optimal k and μ. The model achieves the best effect when $\mu \in [0.5, 0.7]$ shown in Figure 6, and the highest classification prediction accuracy is achieved when k=8.

Different working condition: As shown in Table 8 can see the comparison results of our experiments on these approaches. It is obviously that the JGSWD method proposed in this paper outperforms other compared approaches on the seven tasks, the S-only has the lowest accuracy among these approaches. In addition, the JGSWD also can as the adversarial adaptation approaches, which achieves better performance than the DANN.

The B→C transferring learning task is choosed as the object of t-SNE visualization in randomly, the results of their feature visualization are shown in Figure 7. In these figures, the fault in inner race, outer race, roll and cage are represented by IF, OF, RF and CF respectively, the four different fault categories are represented by four different colors and the circle shape points represent the target domain, on the contrary, the other shape points represent the source domain. It is clear that the feature mapping of the JGSWD is achieving better than other approaches, which cluster circle shape and cross shape points together and classify the ten colors easily.



a. μ



b. **k**

FIGURE 6. The selection of optimal parameter.

TABLE 8. Classification results of SEVEN transfer tasks.

Transfer task	S-only	DANN	CORAL	CDAN	DDC	JMMD	JGSWD
A → B	0.9627	0.9729	0.9731	0.9760	0.9561	0.9789	0.9891
A → C	0.9607	0.9638	0.9688	0.9729	0.9502	0.9777	0.9946
B → A	0.961	0.9654	0.9648	0.9732	0.9571	0.9785	0.9943
B → C	0.9605	0.9715	0.9727	0.9830	0.9617	0.9835	0.9967
C → A	0.9650	0.9662	0.9716	0.9742	0.9602	0.9789	0.9919
C → B	0.9642	0.96102	0.9721	0.9741	0.9593	0.9801	0.9925

Transfer learning from bearingset to gearset: The classification prediction results of proposed method (JGSWD) and comparison methods are shown in Figure 8. The results show that the final classification and recognition of the transferring experiment using the S-only, CORAL,DDC, CDAN, DANN and JMMD methods are not good, the accuracy only reach around 50%. On the contrary, the recognition accuracy of the model is improved using the proposed method (JGSWD) that can reaching about 89.56%. Moreover the experiment is visualized through the t-SNE method, the feature visualization of the unprocessed bearingset and gearset can be seen in Figure 9. It is can be seen clearly that the feature distributions of the bearingset and gearset are different. Then feature visualization after using the approach proposed in this paper is seen in Figure 10 which have an excellent clustering result for the target domain.

Transfer learning from DLE to DRS: The challenge of this experiment lies in the fault bearing data is obtained in different mechanical equipment and the target domain is unlabeled data, which leads to other distribution. The XJTU-SY Bearing Dataset and gearset are choosed as the DLE, dataset collected from a compressor which work in real scenarios as the DRS. The specific dataset is shown in Table 4, and the classification performance of the transfer experiment from XJTU-SY. Bearing Dataset to the DRS in comparison approaches has been shown in the Figure 11 that get the classification accuracy is up to 96%, in the same way the classification accuracy of transfer experiment from gearset to the DRS can get around 80.67%. Moreover, The XJTU-SY Bearing Dataset →DRS transferring learning task is choosed as the object of t-SNE visualization in Figure 12, the results show the proposed method (JGSWD) could achieve better clustering and make the feature distribution of the source and target domains more consistent.

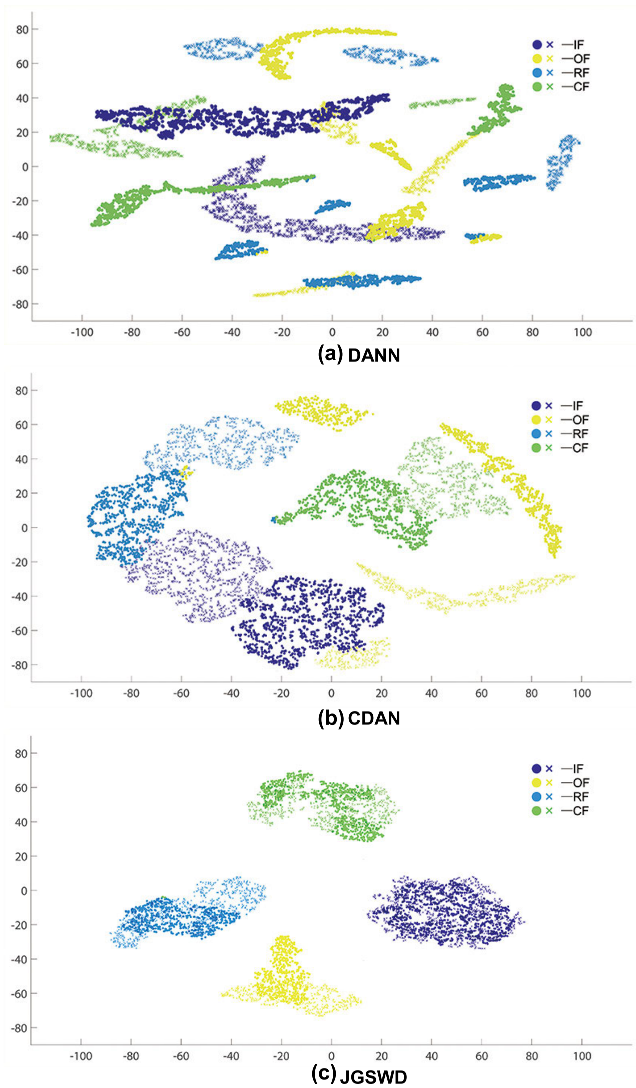


FIGURE 7. Feature visualization of the B to C task in XJTU-SY bearing dataset.

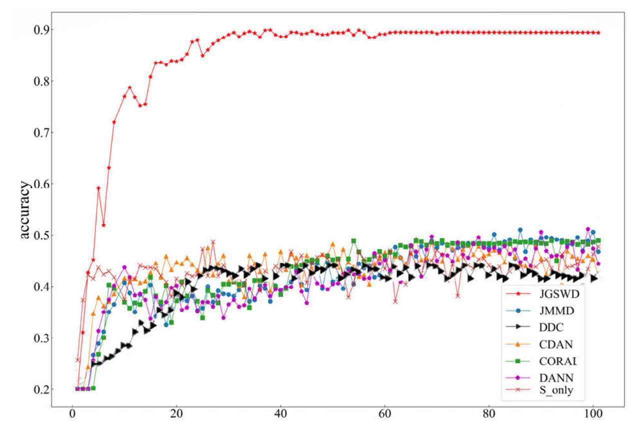


FIGURE 8. Classification result from bearingset to gearset.

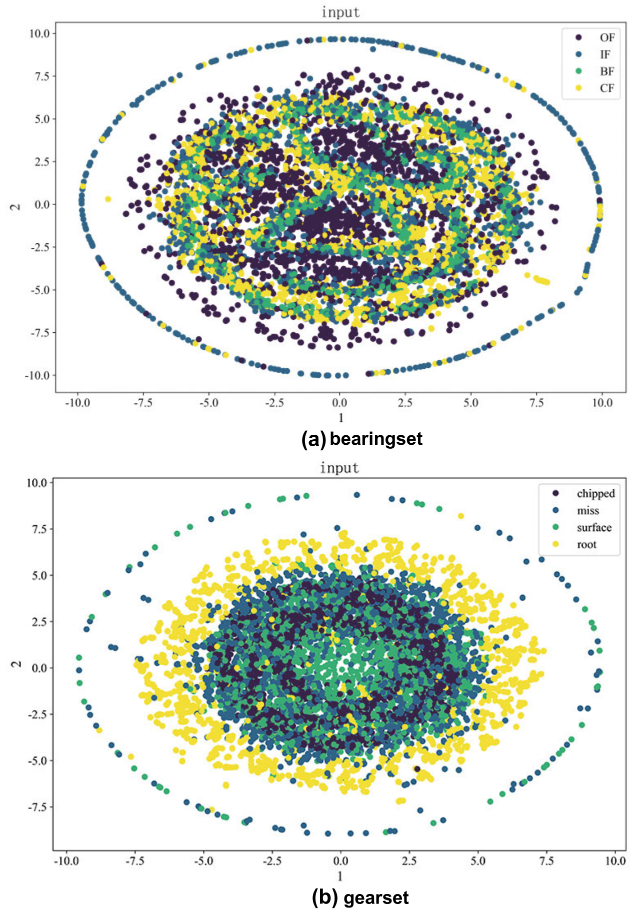


FIGURE 9. Feature visualization of the unprocessed data.

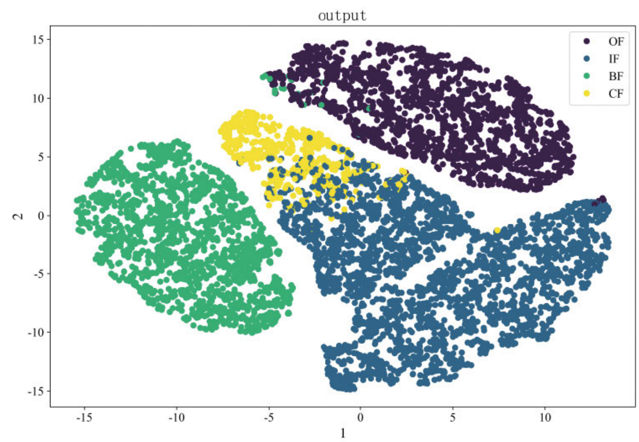
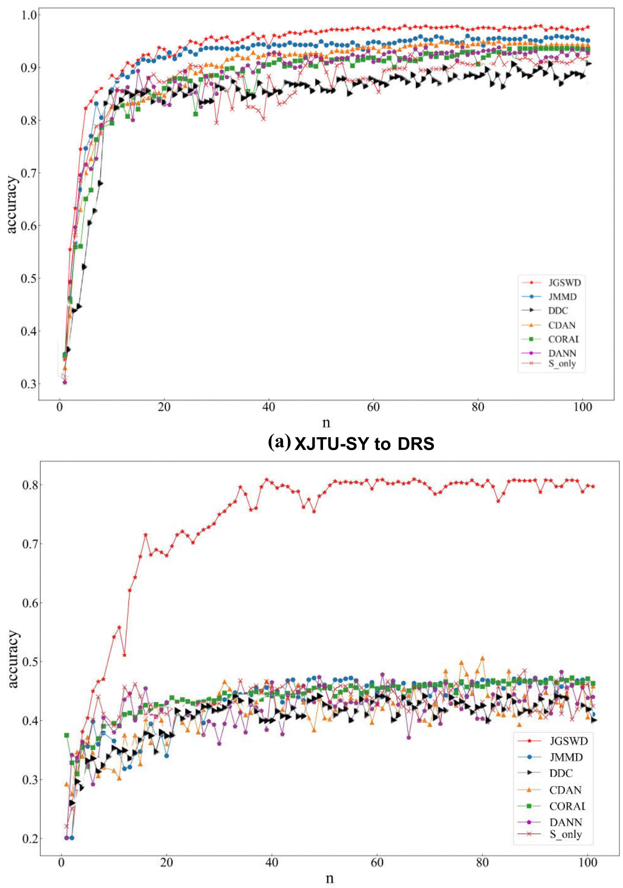


FIGURE 10. Feature visualization of the processed data through.

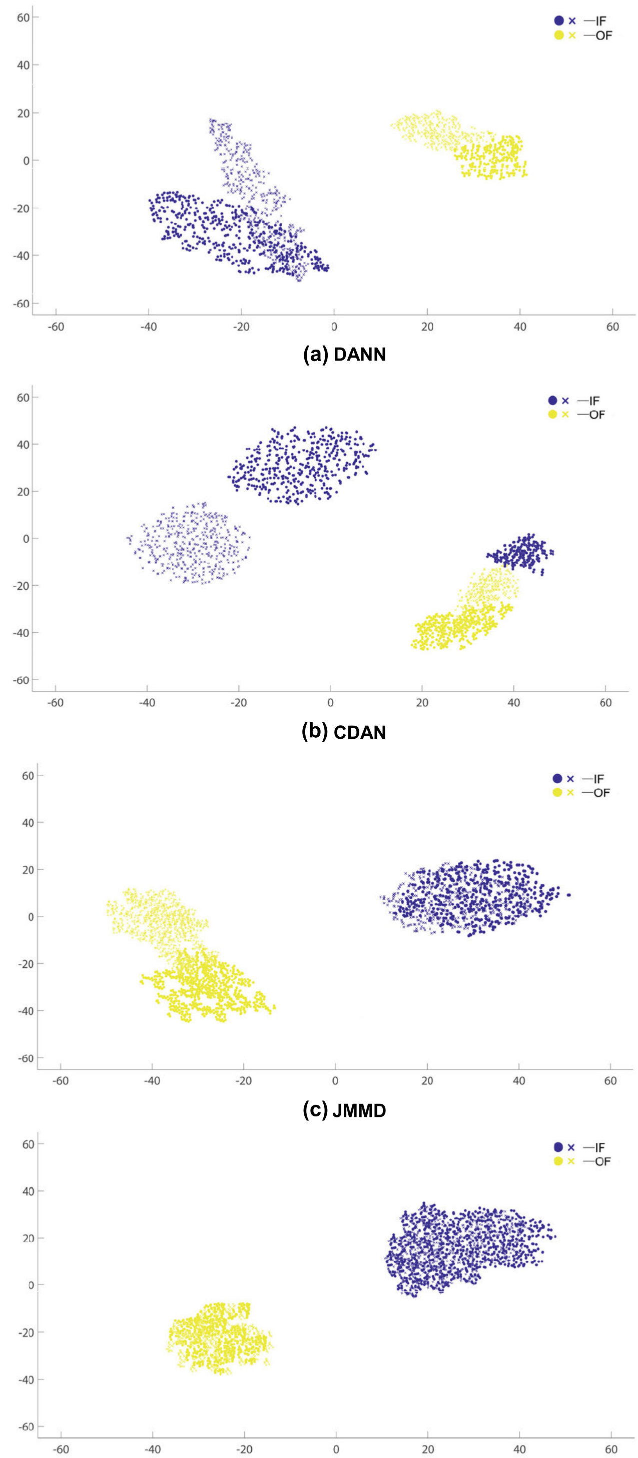
Ablation experiment: To better demonstrate the effectiveness of each proposed innovation on classification accuracy, an ablation experiment is presented in the article. *G*: generalized, *K*: Top K correlated label, *DA*: dynamic domain alignment. "JGSWD-*G*/*DA*/*K*" is implemented without generalized sliced Wasserstein distances, Top K correlated label, and dynamic domain alignment. As shown in Figure 13, it is obvious that each proposed innovation is effective in improving performance and the accuracy of classification prediction.



(a) XJTU-SY to DRS

(b) gearset to DRS

FIGURE 11. Classification results from DLE to DRS.



(a) DANN

(b) CDAN

(c) JMMD

(d) JGSWD

FIGURE 12. Transfer learning performance from DLE to DRS.

The time for predicting the fault type of test sample is greatly reduced.

To some extent, the proposed method is not only a bearing diagnosis model, but also a set of mechanical fault diagnosis methods. Ome Its application can be extended to other scenarios, such as spindle blade tools, gearboxes, and other mechanical equipment. We hope to expand its application in more scenarios in future work.

On the other hand, although the proposed method can achieve excellent and stable bearing diagnostic transmission performance in cross-airfield scenarios, there are still unresolved problems in some experimental results that need to be considered in future work. Firstly, the model is tested and applied in more actual factory scenarios to improve the portability of complex data. Secondly, through comparative analysis and erosion analysis, it is found that the accuracy of some tasks is relatively poor. In future work, when modifying network structure and adjusting network parameters, the stability or robustness of the diagnostic framework needs to be improved. This is critical for diagnostic systems to handle more types of transfer tasks/cases or real-world factory scenarios with complex data distribution. In the following studies, we will explore how to build advanced neural network structures and conduct a more detailed analysis of the formalization of distribution differences. Third, there is still a lack of consideration for the impact of model computation overhead, which can affect actual deployment in a real factory environment. Therefore, more efficient transmission algorithms need to be developed to reduce the computational overhead and realize the trade-off

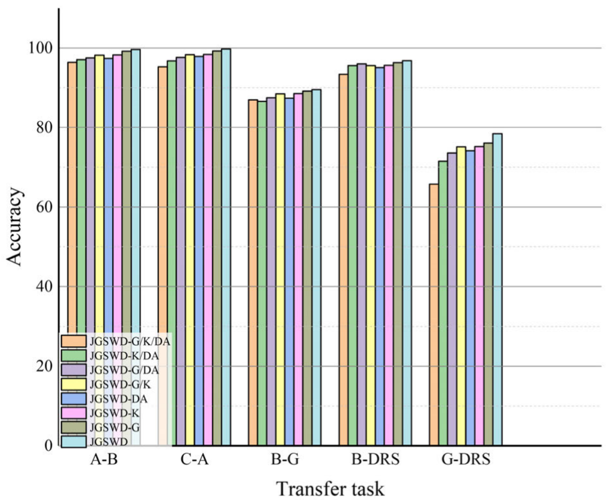


FIGURE 13. **Ablation experiment.**

between transmission performance and computational overhead.

VI. CONCLUSION

This paper proposes a deep transfer learning method for fault diagnosis of rolling bearings based on domain adaptive, joint generalized slice Wasserstein distance-guided transfer learning (JGSWD), which mainly solves the problem that DRS and DLE have different feature distributions. The joint difference between the two domains is minimized by dynamically balancing the edge distribution and the conditional distribution. In this paper, by calculating the Wasserstein distance of generalized slice, the complexity of calculation is significantly reduced and the speed of fault diagnosis is greatly improved. In addition, in order to reduce the conditional distribution and improve the transmission capability, the top K related pseudo-tags are calculated. The network structure is composed of feature generator, domain discriminator, joint feature and classifier. The feature generator is composed of multiple convolution layers. It adopts batch normalization method to extract transferable feature knowledge from DLE and DRS to make its feature distribution more stable, and then uses domain discriminator to reduce domain differences by using maximum and minimum adverssion idea. Using joint features to calculate joint differences. Finally, the traditional deep learning fault diagnosis is carried out by using the classifier. The validity of the JGSWD model is verified by the actual operation of XJTU-SY bearing data set, gear set and compressor bearing data set. The experimental results on the XJTU-SY bearing dataset show that the average classification accuracy of JGSWD method is about 99%. The experimental results of transfer learning from bearing data set to gear set show that the average classification accuracy of JGSWD method is about 89.56%. In addition, the classification accuracy of XJTU-SY bearing data set to DRS transfer experiment can reach 97%, similarly, the classification accuracy of JGSWD method from gear group to DRS transfer experiment can get 80.67%. Superior to the methods compared, this paper provides a new and effective method for fault diagnosis.

DATA AVAILABILITY

The data used to support the findings of this study are included within the article.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] M. Gan, C. Wang, and C. Zhu, "Fault feature enhancement for rotating machinery based on quality factor analysis and manifold learning," *J. Intell. Manuf.*, vol. 29, no. 2, pp. 463–480, Feb. 2018.
[2] C. Li, Y. Tang, and G. Xu, "SVMWPE-HLLE based fault diagnosis approach for rolling bearing," *IEEE Access*, vol. 11, pp. 4645–4659, 2023.
[3] J. Zhang and J. Wu, "A novel transfer dictionary learning strategy for rolling bearing fault identification with a mixed noise model," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, Aug. 2021.
[4] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.
[5] H. Pan, H. Xu, J. Zheng, and J. Tong, "Non-parallel bounded support matrix machine and its application in roller bearing fault diagnosis," *Inf. Sci.*, vol. 624, pp. 395–415, May 2023.
[6] W. Li, J. He, H. Lin, R. Huang, G. He, and Z. Chen, "A lightgbm-based multi-scale weighted ensemble model for few-shot fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, Jul. 2023.
[7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Jan. 2009.
[8] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.
[9] H. Wang and W. Du, "Multi-source information deep fusion for rolling bearing fault diagnosis based on deep residual convolution neural network," *Proc. Inst. Mech. Engineers, Part C, J. Mech. Eng. Sci.*, vol. 236, no. 13, pp. 7576–7589, Jul. 2022.
[10] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.
[11] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May 2017.
[12] X. Huang, G. Qi, N. Mazur, and Y. Chai, "Deep residual networks-based intelligent fault diagnosis method of planetary gearboxes in cloud environments," *Simul. Model. Pract. Theory*, vol. 116, Apr. 2022, Art. no. 102469.
[13] M. Huang, Z. Liu, and Y. Tao, "Mechanical fault diagnosis and prediction in IoT based on multi-source sensing data fusion," *Simul. Model. Pract. Theory*, vol. 102, Jul. 2020, Art. no. 101981.
[14] C. Che, H. Wang, R. Lin, and X. Ni, "Deep meta-learning and variational autoencoder for coupling fault diagnosis of rolling bearing under variable working conditions," *Proc. Inst. Mech. Engineers, Part C, J. Mech. Eng. Sci.*, vol. 236, no. 17, pp. 9900–9913, Sep. 2022.
[15] T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, and E. Xu, "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," *ISA Trans.*, vol. 119, pp. 152–171, Jan. 2022.
[16] S. Ben-David, J. Blitzer, and F. Pereira, "Analysis of representations for domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, no. 19, 2006, pp. 1–8.
[17] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, May 2019.
[18] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019.

[19] Q. Qian, Y. Qin, J. Luo, Y. Wang, and F. Wu, "Deep discriminative transfer learning network for cross-machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 186, Mar. 2023, Art. no. 109884.
[20] Z. Wu, H. Jiang, K. Zhao, and X. Li, "An adaptive deep transfer learning method for bearing fault diagnosis," *Measurement*, vol. 151, Feb. 2020, Art. no. 107227.
[21] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.
[22] X. Wang, C. Shen, M. Xia, D. Wang, J. Zhu, and Z. Zhu, "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Rel. Eng. Syst. Saf.*, vol. 202, Oct. 2020, Art. no. 107050.
[23] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–28, 2021.
[24] J. Li, Y. Liu, and Q. Li, "Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method," *Measurement*, vol. 189, Feb. 2022, Art. no. 110500.
[25] Q. Li, C. Shen, L. Chen, and Z. Zhu, "Knowledge mapping-based adversarial domain adaptation: A novel fault diagnosis method with high generalizability under variable working conditions," *Mech. Syst. Signal Process.*, vol. 147, Jan. 2021, Art. no. 107095.
[26] P. Chen, R. Zhao, T. He, K. Wei, and Q. Yang, "Unsupervised domain adaptation of bearing fault diagnosis based on join sliced Wasserstein distance," *ISA Trans.*, vol. 129, pp. 504–519, Oct. 2022.
[27] I. Goodfellow, J. Pouget-Abadie, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, no. 27, 2014, pp. 1–9.
[28] S. Kolouri, K. Nadjahi, and G. Rohde, "Generalized sliced Wasserstein distances," in *Proc. Adv. Neural Inf. Process. Syst.*, no. 32, 2019, pp. 1–12.
[29] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.
[30] J. Shen, Y. Qu, and Y. Yu, "Wasserstein distance guided representation learning for domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, p. 11784.
[31] D. Zhao, S. Liu, H. Du, L. Wang, and Z. Miao, "Deep branch attention network and extreme multi-scale entropy based single vibration signal-driven variable speed fault diagnosis scheme for rolling bearing," *Adv. Eng. Informat.*, vol. 55, Jan. 2023, Art. no. 101844.
[32] Y. Tang and F. Lin, "Fault feature extraction of reciprocating compressor based on adaptive waveform decomposition and lempel-ziv complexity," *IEEE Access*, vol. 7, pp. 82522–82531, 2019.
[33] H. Ajakan and P. Germain, "Domain-adversarial neural networks," 2014, *arXiv:1412.4446*.
[34] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. AAI Conf. Artif. Intell.*, vol. 30, no. 1, 2016, p. 10306.
[35] M. Long, H. Zhu, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2208–2217.

NA LEI received the master's degree from the Mechanical Science and Engineering College, Northeast Petroleum University, China, in 2006, and the Ph.D. degree in oil and natural gas engineering from the School of Petroleum Engineering, Northeast Petroleum University, in 2016. She is currently an Associate Professor of mechatronic engineering with the Mechanical Science and Engineering College, Northeast Petroleum University. Her current research interests include mechanical fault diagnosis and noise and vibration control.

JIPENG CUI received the bachelor's degree from the Mechanical Engineering College, Inner Mongolia University of Science and Technology, China, in 2019. He is currently pursuing the master's degree in mechatronic engineering with the Mechanical Science and Engineering College, Northeast Petroleum University. His current research interests include mechanical fault diagnosis and big data visualization.

JICHENG HAN received the bachelor's degree from the Architectural Engineering Institute, East University of Heilongjiang, China, in 2020, and the master's degree from the Mechanical Science and Engineering College, Northeast Petroleum University, China, in 2023. His current research interests include mechanical fault diagnosis and noise and vibration control.

XIAN CHEN received the bachelor's degree from the School of Mechanical and Electrical and Vehicle Engineering, East China Jiao Tong University, China, in 2020, and the master's degree from the Mechanical Science and Engineering College, Northeast Petroleum University, China, in 2023. Her current research interests include mechanical fault diagnosis and artificial intelligence.

YOUFU TANG received the master's degree from the Mechanical Science and Engineering College, Northeast Petroleum University, China, in 2006, and the Ph.D. degree in mechanical and electronic engineering from the School of Mechatronic Engineering and Automation, Shanghai University, China, in 2013. He is currently a Professor of engineering mechanics with the Mechanical Science and Engineering College, Northeast Petroleum University. His current research interests include mechanical fault diagnosis and noise and vibration control.