Article:

Liu, Z.-H. orcid.org/0000-0002-6597-4741, Chen, L., Wei, H.-L. orcid.org/0000-0002-4704-7346 et al. (3 more authors) (2023) A tensor-based domain alignment method for intelligent fault diagnosis of rolling bearing in rotating machinery. Reliability Engineering & System Safety, 230. 108968. ISSN 0951-8320

https://doi.org/10.1016/j.ress.2022.108968

Accepted for publication by *Reliability Engineering & System Safety* on 7 November 2022

A Tensor-based Domain Alignment Method for Intelligent Fault Diagnosis of Rolling Bearing in Rotating Machinery

Zhao-Hua Liu[a,*], Liang Chen[a], Hua-Liang Wei[b], Fa-Ming Wu[c], Lei Chen[a], Ya-Nan Chen[c]

([a]School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

[b]Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3JD, U.K.

[c]Wind Power Business Division, CRRC Zhuzhou Electric Locomotive Institute Co. Ltd., Zhuzhou 412001, China)

Abstract—**Fault diagnosis of rolling bearings plays a pivotal role in modern industry. Most existing methods have two disadvantages: 1) The assumption that the training and test data obey the same distribution; and 2) They are designed for vector representation which is unable to characterize the important structure of the rolling bearings data of interest. To overcome these drawbacks, this paper proposes a novel tensor based domain adaptation method. Firstly, this method uses the time domain signals, the frequency domain signals, and the Hilbert marginal spectrum and integrates them into a third-order tensor model. Secondly, these three types of signals are split into two parts: the source and target domain data; all the representative features are identified in the source domain. Thirdly, a tensor decomposition method is used to decompose the features into a series of third-order tensors, and several alignment matrices are defined to align the representation of the two domains to the tensor invariant subspace. Then, the alignment matrices and the tensor subspace are jointly optimized to realize the adaptive learning. Finally, the feature tensor is reconstructed into a matrix form to realize the fault diagnosis through the classifier. Extensive experiments are conducted on a public dataset and a dataset collected from our own laboratory; experimental results show the satisfactory performance of the proposed method.**

Index Terms—**Tensor representation, subspace learning, tensor alignment, fault diagnosis, domain adaptation, transfer learning, rolling bearings, rotating machinery.**

I. Introduction

Fault diagnosis of rotating machinery has become an increasingly critical part of reliability engineering and system safety. In industrial applications, early fault diagnosis of rotating machinery can ensure reliable and safe operations [1], this is because rolling bearings, as a key component of most rotating equipment, are extremely vulnerable to wear and failure due to high pressure, heavy loads, and long-term operation under variable working conditions. These factors and many other uncertainties can lead to faults in

*Corresponding author. Tel/Fax number: 86-15874160936.
E-mail address: zhaohualiu2009@hotmail.com (Z.-H. Liu), 13069302167@163.com (L. Chen), w.hualiang@sheffield.ac.uk (H.-L. Wei), wufm@csrzic.com (F.-M. Wu), chenlei@hnust.edu.cn (L. Chen), chenyn@csrzic.com (Y.-N. Chen).

rolling bearings and may further cause serious accidents. In order to avoid major economic losses and catastrophic incidents, and strengthen the reliability of equipment and system safety, it is extremely important to develop effective and robust intelligent fault diagnosis methods [2].

Intelligent fault diagnosis methods for rotating machinery have attracted a great deal of attention in recent years [3]-[6]. Given that the fault information is usually hidden in vibration signals, the intelligent diagnosis methods based on vibration signals and data-driven modelling such as support vector machines (SVM) [7], stacked variant autoencoder [8], long short-term memory (LSTM) [9], and convolutional neural networks (CNN) [10], have become a research focus [11]. Most existing machine learning methods show excellent performance under a common hypothesis: the labeled training data and the unlabeled test data are subject to the same distribution. However, in practice such an assumption is not always satisfied, meaning that many machine learning methods may not work well for fault diagnosis tasks. For example, a machine learning model trained on a training dataset (in the source domain) may not work on a test dataset (a new dataset in the target domain) if there exists a large discrepancy in the data distribution properties in the two domains [12].

In order to solve this challenge, domain adaptation (DA), as a method of transfer learning, provides a new learning strategy for the establishment of knowledge transfer from source domain to target domain. DA is concerned with exploring domain-invariant features across distribution differences, and applying machine learning models built in the source domain to the target domain where the data distribution is different but related to that in the source domain [13]. In [14], Ainapure *et al.* proposed a fault diagnosis method by using CNN to extract features automatically, and the maximum mean discrepancy (MMD) was used to align the data distribution. To improve the model generalization ability, noise was added to the health condition label data in model training process. Sharma *et al.* [15] proposed a quick learning mechanism based on a DA method to realize the fault diagnosis of rolling bearing. In their diagnosis model, the MMD was minimized by the Net2Net transformation, and the target data pattern can be quickly captured. Schwendemann *et al*. [16] proposed a new DA method based on CNN and layered MMD. This method constructs an intermediate domain by converting the data into the time frequency domain based on the windowed envelope, and then uses CNN to learn the feature representation.

These DA methods have achieved satisfactory diagnosis results in transfer tasks under different situations. One of the assumptions of these DA methods is that there must be a close relevance between the source and target domains, that is, there must be a certain commonality to adapt. Many methods explore the commonality by learning invariant subspace, and achieve the desired effect [17]. However, the existing fault diagnosis methods based on DA and other machine learning methods mainly focus on the vector data. Therefore, when applying these technologies to process structured high-dimensional representation, the data must be vectorized first. Although this partly solves the problem, the vectors or the matrices cannot naturally and effectively represent and store the important information of structural data. In fact, the vectorization may even lead to the curse of dimensionality and increase the computational complexity [18].

Recently, new methods based on tensor data representation and tensor decomposition have been successfully applied to fault diagnosis. A tensor space is a high-order generalization of the low-rank space. The tensor representation can well preserve the multi-linear relationship of the data, which cannot be captured by the conventional vector and matrix representation. Therefore, the tensor decomposition, with its outstanding potential information extraction capabilities, has attracted more and more attention. Hu *et al*. [19] proposed a multi-dimensional signal fault diagnosis method for rolling bearings based on the tensor decomposition. The vibration signal was established as a third-order tensor, and the Tucker decomposition was used to filter the tensor to obtain an ideal target tensor. This method can well remove the noise and retain the fine features of the data as much as possible. Luo *et al*. [20] designed an adaptive sequential fault diagnosis framework that combined tensor decomposition layer and deep neural network model. This method not only improved the performance of neural network structure, but also enhances its feature expression ability. He *et al*. [21] designed a tensor classifier based on the flexible and displaceable convex hull of kernels. The main idea was to transform the optimal separation hyperplane problem between the kernel flexible convex hull and the replaceable convex hull of the tensor sample set into a quadratic programming problem, and use the replaceable factor to improve the interference of outliers. This method is effective for solving the problems of small sample and multi-source signal classification. Zhao *et al.* [22] proposed a rolling bearing fault diagnosis method based on the segment tensor rank decomposition, which used the segment tensor rank-(Lr, Lr, 1) decomposition to obtain the sub-tensors of source signals. Then, the source signal was reconstructed by combining the mode information of the sub-tensor.

Presently, tensor based methods are mainly used in image recognition and related fields [23]-[25], and relatively less applications in fault diagnosis have been reported in the literature. The traditional fault diagnosis methods based on DA only consider the distribution difference between the source and target domains, without considering the following important key point: the conventional multi-dimensional signals cannot fully retain and reflect the data integrity under the vector representation. In addition, the existing tensor correlation methods only use the tensor decomposition techniques, which may not be able to sufficiently reveal and represent the feature fault classification purpose. To overcome these weaknesses and improve the diagnostic ability of the classifier to be trained in the source domain, there is a strong need to establish an effective unified tensor representation of multi-dimensional signals and global domain adaptation, so as to reduce the difference in domain distribution and improve the accuracy of fault diagnosis.

To effectively address the problem of cross domain fault diagnosis and strengthen the reliability and safety of equipment, this paper proposes a novel tensor domain adaptation (TDA) method named joint-domain alignment based invariant tensor subspace learning. Firstly, the method establishes a unified tensor representation model for multi-dimensional signals. Secondly, the Tucker decomposition is used to obtain tensor subspace and tensor subspace representation. Finally, an alignment matrix is introduced to achieve the domain adaptation.

Different from the existing literature, this paper proposes a novel tensor-based domain alignment method for intelligent fault diagnosis of rolling bearing. The main contributions of this work are summarized as follows:

1) In order to effectively retain and reflect the important structure and internal relationship of multidimensional data, a unified tensor representation model is established, which provides the data reliability guarantee for better feature recognition and model training.

2) A novel TDA framework is proposed to solve the problem of data distortion caused by the forced alignment of shared subspace in vector representation in traditional DA methods. The proposed method uses DA in the learning process of invariant tensor subspace, to make the learned subspace have stronger representation ability. Furthermore, an alternating optimization algorithm with orthogonal constraints is used to efficiently train the TDA model.

The remainder of this paper is organized as follows. Section II provides the preliminaries. The proposed method is presented in Section III. In Section IV, the comparative study is conducted. Finally, the conclusions are arranged in Section V.

II. PRELIMINARIES

A. Problem Formulation

A domain $\mathcal{D}=\{\mathcal{X},P(X)\}$ comprises a feature space of inputs, $\mathcal{X}$, and a marginal probability distribution of inputs, $P(X)$.

A task $\mathcal{T}=\{\mathcal{Y},f(x)\}$, associated with a specific domain $\mathcal{D}$, is defined as a label space $\mathcal{Y}$ and a target prediction function $f(x)$. From a probabilistic viewpoint, $f(x)$ can be considered as a conditional distribution $P(y|x)$, where $y\in\mathcal{Y}$.

In this paper, the situation is a set of labeled source domain $\mathcal{D}_s=\{\mathcal{Z}_i,y_i\}_{i=1}^{n_s}$, where $\mathcal{Z}_i\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ is an N-th order tensor and $y_i\in\{1,2,\ldots,C\}$ is the corresponding class label. Similarly, a set of unlabeled target domain $\mathcal{D}_T=\{\mathcal{Z}_j\}_{j=1}^{n_t}$ is given. The aim is to infer the class label $y_j\in\{1,2,\ldots,C\}$ of a target based on a model trained with data in the source domain.

B. Basic Theory of the Tensor

A tensor is a high-order generalization of vectors and matrices, which is widely used in data mining, machine learning, computer vision, and other fields [26].

The N-th order tensor $\mathcal{Z}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ is a N-dimensional data array, with elements being denoted by $z_{i_1\cdots i_k\cdots i_n}$, where $1\le i_n\le I_N$, for $k=1,\ldots,N$. The *model-n* product of a tensor $\mathcal{Z}$ with a matrix M is defined as $\mathcal{Z}\times_n M$. The operator $\times_n$ represents the matrix multiplication performed along the n-th mode. Similarly, the product can be performed equivalently by matrix multiplication $\left(\mathcal{Z}\times_n M\right)_{(n)}=MZ_{(n)}$, where $Z_{(n)}$ is the *model-n* matrix unfolding.

C. Tensor Decomposition

Tensor decomposition is a powerful computing technique that decomposes the raw data and extracts the valuable features from the data. In this paper, the Tucker decomposition [27] is used to generate the tensor subspace. For a *N*-th order tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, a core tensor and a set of factor matrices along each mode can be obtained by Tucker decomposition. The tensor can be expressed as follows:

$$\mathcal{Z} = \mathcal{C}\prod_{i=1}^{N} \times U^{(i)} = \mathcal{C};\mathcal{U} \;, \tag{1}$$

where $\mathcal{C} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_N}$ is the core tensor, denoting the tensor subspace representation of $\mathcal{Z}$, and $U^{(n)} \in \mathbb{R}^{I_n \times d_n}$ represents the factor matrix of the *n*-th mode. The last part of (1), $\mathcal{C};\mathcal{U}$, is a simplified representation of Tucker decomposition, where $\mathcal{U}$ denotes the tensor subspace.

Specifically, for a third-order tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, a third-order core tensor and three second-order factor matrices can be obtained by Tucker decomposition, and the associated tensor decomposition model is shown in Fig. 1.

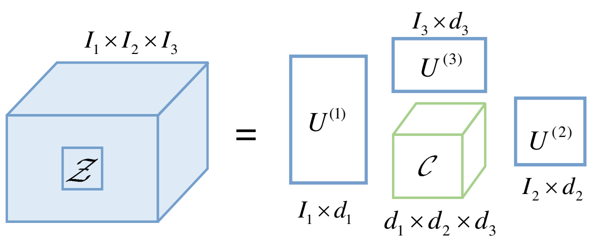


Fig. 1. Third-order tensor factorization.

D. Feature Extraction through the Tensor Factorization

At present, large amounts of data are available from heterogeneous sources in the field of fault diagnosis and other research fields as well. Constructing effective mathematical models to uniformly represent the complex heterogeneous data, and designing efficient feature extraction algorithm to extract the high-quality core features from low-quality raw data, are important for data-driven fault diagnosis.

In order to overcome the shortages of the traditional methods and make good use of the available data, this work constructs the multi-source data into a tensor representation to build more effective diagnosis models based on the tensor decomposition. Compared with the vector representation, the tensor representation has the following advantages:

(1) The data is converted to high-order tensor data and multi-dimensional data with larger data volume; the better characteristics can be obtained without the loss of original data information.

(2) In the feature extraction, the tensor representation can fully consider the association between the data and effectively retain the structural features of the original samples. The new features extracted by the tensor decomposition under orthogonal constraints

have better quality.

(3) The tensor can overcome the curse of dimensionality issue encountered when using the traditional vectorization representation methods.

This paper uses the tensor Tucker decomposition method to project the tensor representation of the data into a tensor subspace, and then transform the tensor into a core tensor of the same order but lower dimensions. Finally, the core tensor is vectorized and reconstructed to feature vectors.

E. Domain Adaptation with Subspace Learning

Recently, the subspace learning based DA has shown good performance in visual data analysis [28], [29]. As mentioned in section II-A, the source and target domain are part of the same *L*-dimensional space, even though they have different data distributions. A principal component analysis approach can be applied to the two subspace to obtain two orthogonal mapping matrices, denoted by X_s and X_T, respectively, where $X_S, X_T \in \mathbb{R}^{L\times l}$, $X_S X_S^T = I_l$, $X_T X_T^T = I_l$, and I_l is the *l*-dimensional identity matrix.

The strategy of most existing DA-based subspace learning methods is to project the source and target data into a shared subspace or construct a set of intermediate representations. This can lead to the loss of information. To overcome this issue, this paper proposes to project the domain data into the respective subspaces of both the source and target domains, so that the source subspace coordinate system is aligned with the target subspace coordinate system. It then directly compares the source and target samples in their respective subspaces. Specifically, an alignment matrix *M* is introduced to realize the alignment from X_s to X_T, and the subspaces after projection are $X_s M$ and $X_T M$ respectively. The matrix *M* is learned by minimizing the Bregman matrix divergence as follows:

$$F(M) = \left\| X_S M - X_T \right\|_F^2 \tag{2}$$

$$\tilde{M} = \arg\min_{M}(F(M)) = X_S^{-1} X_T \tag{3}$$

where $\|\bullet\|_F^2$ is the Frobenius norm.

Similarly, for tensor subspaces U_S and U_T, the domain discrepancy can be defined as $\left\| U_S M - U_T \right\|_F^2$. According to definition (1), it can be inferred that the arrangement of the tensor is equivalent to the arrangement of the tensor subspace, so $\mathcal{Z}_S; \mathcal{M}$ can be expressed as:

$$\mathcal{Z}_S \prod_{i=1}^{N} \times M^{(i)} = \mathcal{C}_S \prod_{i=1}^{N} \times (M^{(i)} U_S^{(i)}) \tag{4}$$

III. Proposed Method

As mentioned earlier this study attempts to align the data distribution in different domains while preserving the important structure of the data. In this section, a TDA fault diagnosis model for variable working conditions is proposed based on the tensor decomposition and subspace learning approaches described in the previous section.

A. Proposed TDA Model

In the traditional transfer learning, the metrics such as MMD [12], A-distance [30], and Wasserstein distance [31] are often used to measure the difference in probability distribution between different domains. However, most of the existing DA methods based on these distance metrics are only applicable to vectors, and these metrics may not be able to preserve and reflect important structures for some applications. In addition, some existing DA methods are implemented under the premise that there is a shared subspace between the source and target domains. This is reasonable when the domain difference is not large; but if the difference is large, these methods may not work well for aligning the shared subspace.

In order to address the above problems, this paper proposes a TDA method, together with an invariant subspace learning scheme, to directly adapt the tensor representation of the source and target domains. The structure of the proposed method is shown in Fig. 2. In this study, the tensor representation model of data is constructed firstly, and then the data is divided and mapped into shared space. Next, an approach of tensor Tucker decomposition with orthogonal constraints is used to extract the hidden features of the data and apply them to the computational model of joint-domain alignments based tensor invariant subspace learning. Finally, the trained model is directly applied to the target datasets to achieve the classification of unlabeled data.

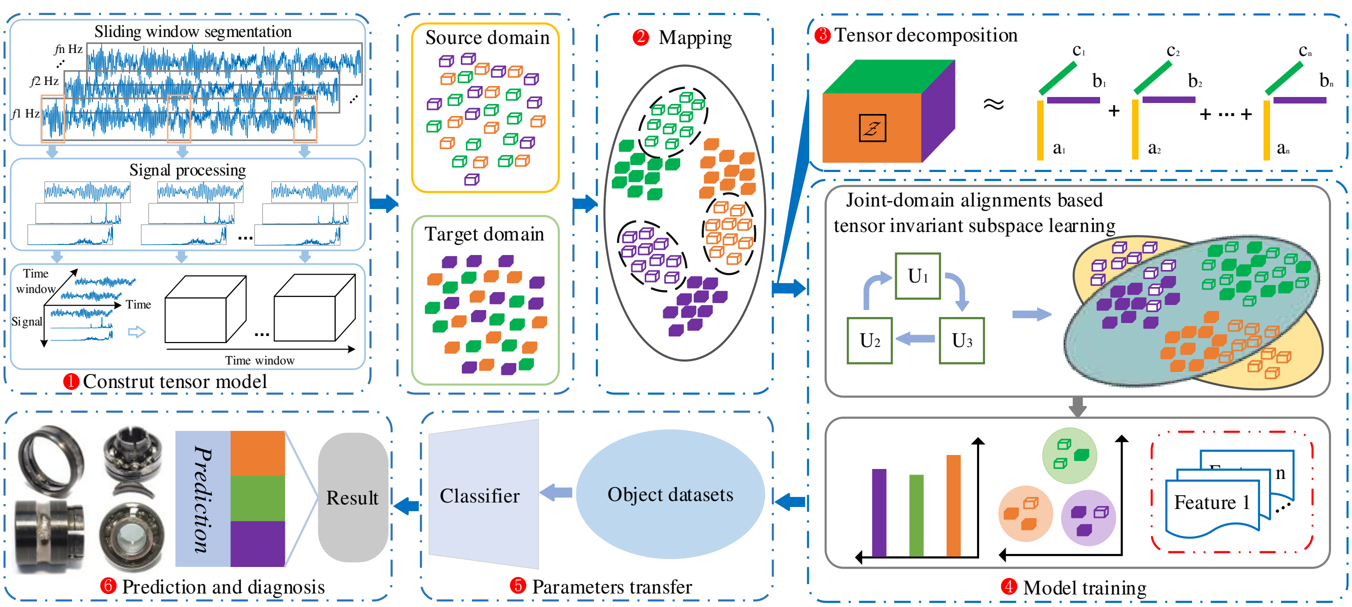


Fig. 2. Framework of the proposed TDA method

1) Construction of Tensor Model

Most raw vibration data contain tens of thousands of sampled points. In order to make full use of the information contained in the sampled signals, this study adopts a sliding rectangular window approach to divide the monitoring data under different working conditions into multiple time window segments, where each segment has H samples. After that, The fast Fourier transform and Hilbert-Huang Transform are used to analyze the data in each time window to obtain the time domain, frequency domain, and marginal spectra of the signals of different scales. Finally, a third-order tensor model of multichannel $Signals \times time\ window \times time$ is constructed. The process of tensor construction is shown in the first step in Fig. 2.

2) Joint-Domain Alignments Based Tensor Invariant Subspace Learning

Different from the existing DA methods based on subspace learning, the proposed model uses the tensor decomposition to obtain the tensor subspace. In the learning process of tensor subspace, the joint-domain alignment is integrated into the process and a dynamic distribution alignment is introduced. This method can not only align the shared subspace of the source and target domains at the same time, and alleviate the problem of excessive data distortion caused by forced alignment, but also overcomes the limitation of the traditional vector presentation where the vectorized feature space may not fully characterize the important structure of the data. The process of joint-domain alignments based tensor invariant subspace learning is shown in Step 4 of Fig. 2, where U_i represents the factor matrix of the *i*-th mode obtained by the Tucker decomposition, which is a second-order matrix.

As mentioned in section II, given N_s labeled samples $\{\mathcal{Z}_s^n, y_s^n\}_{n=1,\ldots,N_s}$ from the source domain, each sample represents a *K*-model tensor $\mathcal{Z}_s^n \in \mathbb{R}^{n_1 \times \ldots n_K}$, and all the samples of source domain can be stacked into a (*K*+1)-order tensor $\mathcal{Z}_s \in \mathbb{R}^{n_1 \times \ldots n_K \times N_s}$. Similarly, N_t samples $\{\mathcal{Z}_t^n\}_{n=1,\ldots,N_t}$ from the target domain are stacked as a $(K+1)$-order tensor. Based on (2) and (3), the error of tensor subspace learning model can be written as:

$$\begin{gathered}\mathcal{L}(\mathcal{U},\mathcal{C}_s,\mathcal{C}_t,\mathcal{M}) = \| \; \mathcal{Z}_s;\mathcal{M} \;-\; \mathcal{C}_s;\mathcal{U} \; \|_F^2 + \\ \| \; \mathcal{Z}_t;\mathcal{M} \;-\; \mathcal{C}_t;\mathcal{U} \; \|_F^2, \; s.t. \; \forall k, U^{(k)T}U^{(k)} = I\end{gathered} \tag{5}$$

where $\mathcal{M} = \{M^{(k)}\}_{k=1,\ldots,k}$ denotes a set of alignment matrices, and $\mathcal{C}_s$ and $\mathcal{C}_t$ represent the tensor subspace representation of $\mathcal{Z}_s$ and $\mathcal{Z}_t$ respectively. $\mathcal{U}$ is the invariant tensor subspace, which is related to a dynamic distribution alignment under tensor representation, and $U^{(k)}$ is column-wise orthogonal. Here *I* is an identity matrix.

However, without preserving the original data variance after alignment, it can cause the projected data to cluster to a single point [18], [32]. Therefore, it is necessary to adopt an orthogonal constraint on $\mathcal{M}$ to keep the data variance as much small as possible. The expression of the constraint condition is:

Accepted for publication by *Reliability Engineering & System Safety* on 7 November 2022

$$\mathcal{F}(\mathcal{M}) = \left\| [\![\mathcal{Z}_s; \mathcal{M}; \mathcal{M}^T]\!] - \mathcal{Z}_s \right\|_F^2, \quad s.t.\ \forall k, M^{(k)T} M^{(k)} = I \tag{6}$$

where $M^{(k)}$ is row-wise. It is worth noting that the regularization term reflects the quality of the source data reconstructed by $\mathcal{M}$.

B. Softmax Classifier for Fault Diagnosis

This work uses the widely used softmax classifier, which is easy to implement and fast to calculate [33]. Given a labeled training set $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in \mathbb{R}^{N\times 1}$ and $y_i \in (1,2,\ldots,C)$ are sample sand the corresponding labels. For each input sample x_i, the classifier calculates the probability $P(y_i = c \mid x_i)$ for each label of $c = 1,2,\ldots C$ according to the hypothesis function, and uses the label corresponding to the maximum probability as the class of the sample. Specifically, the hypothesis function is defined as:

$$J_\theta(x_i) = \begin{bmatrix} p(y_i = 1 \mid x_i;\theta) \\ p(y_i = 2 \mid x_i;\theta) \\ \vdots \\ p(y_i = C \mid x_i;\theta) \end{bmatrix} = \frac{1}{\sum_{c=1}^{C} e^{\theta_c^T} x_i} \begin{bmatrix} e^{\theta_1^T} x_i \\ e^{\theta_2^T} x_i \\ \vdots \\ e^{\theta_C^T} x_i \end{bmatrix} \tag{7}$$

where $\theta = [\theta_1, \theta_2, \ldots, \theta_C]^T$ represents the regression model parameters. The ingenious design is term $\sum_{c=1}^{C} e^{\theta_c^T} x_i$ normalizes the distribution, which enables the classifier to ensure that the probability estimation output for each sample is positive and the sum is 1.

Based on the hypothesis function, the softmax classifier uses the cross-entropy function $\mathcal{L}_y$ as the loss function [34], and defines $\mathcal{L}_y$ as:

$$\mathcal{L}_y = -\frac{1}{N}\left[\sum_{i=1}^{N} \sum_{j=1}^{C} 1\{y_i = c\} \log \frac{e^{\hat{\theta}_c^T x_i}}{\sum_{j=1}^{C} e^{\hat{\theta}_j^T x_i}} \right] \tag{8}$$

where $1\{y_i = c\}$ represents the indicator function. If the condition is true, it returns 1; otherwise the return value is 0.

C. Objective Function

By combining (5) and (6), the overall objective function of the TDA model can be obtained as follows:

$$\min_{\mathcal{U},\mathcal{G}_s,\mathcal{G}_t,\mathcal{M}} \left\| \mathcal{Z}_s;\mathcal{M} - \mathcal{C}_s;U \right\|_F^2 + \left\| \mathcal{Z}_t;\mathcal{M} - \mathcal{C}_t;\mathcal{U} \right\|_F^2 + \lambda \left\| [\![\mathcal{Z}_s;\mathcal{M};\mathcal{M}^T]\!] - \mathcal{Z}_s \right\|_F^2, \quad s.t.\ \forall k, V^{(k)T} U^{(k)} = I, M^{(k)T} M^{(k)} = I \tag{9}$$

where λ denotes the weight of the regularization term.

Since $\mathcal{M}$ and $\mathcal{U}$ influence each other in (6), it is difficult to perform the joint optimization. The general approach is to decompose the original problem into sub-problems, which are solved by alternately updating $\mathcal{U}$, $\mathcal{M}$ and the core tensor $\mathcal{C}$ until the objective function converges.

1) $\mathcal{V}$ *Sub-Problem*

The prior condition for the $\mathcal{U}$ sub-problem as follows: given $\mathcal{M}$, update $\mathcal{U}$, $\mathcal{C}_s$ and $\mathcal{C}_t$, so the equation (9) can be reduced to:

$$\begin{aligned} &\min_{\mathcal{U},\mathcal{G}_s,\mathcal{G}_t} \left\| \hat{\mathcal{Z}}_s - \mathcal{C}_s;\mathcal{U} \right\|_F^2 + \left\| \hat{\mathcal{Z}}_t - \mathcal{C}_t;\mathcal{U} \right\|_F^2, \\ &s.t. \, \forall k, U^{(k)T}U^{(k)} = I. \end{aligned} \tag{10}$$

This problem can be solved effectively by Tucker decomposition. When the optimal $\tilde{\mathcal{U}}$ is found, $\mathcal{C}_s$ and $\mathcal{C}_t$ can be easily obtained by the following transformation:

$$\mathcal{C}_{s/t} = \mathcal{Z}_{s/t} \prod_{k=1}^{K} \times \tilde{U}^{(k)T} \tag{11}$$

2) $\mathcal{M}$ *Sub-Problem*

The prior condition for the $\mathcal{M}$ sub-problem is as follows: supposing $\mathcal{U}$, $\mathcal{C}_s$, and $\mathcal{C}_t$ are fixed, update $\mathcal{M}$, so the problem is formulated as:

$$\begin{aligned} &\min_{\mathcal{U},\mathcal{G}_s,\mathcal{G}_t,\mathcal{M}} \left\| \mathcal{Z}_s;\mathcal{M} - Q_s \right\|_F^2 + \left\| \mathcal{Z}_t;\mathcal{M} - Q_t \right\|_F^2 + \lambda \left\| [[\mathcal{Z}_s;\mathcal{M} \; ;\mathcal{M}^T]] - \mathcal{Z}_s \right\|_F^2, \\ &s.t. \, \forall k, M^{(k)T}M^{(k)} = I \end{aligned} \tag{12}$$

where Q_s and Q_t are known and fixed, and the equation (12) can be effectively solved by using an optimization method with orthogonality constraints [35].

In summary, the variable alternate iteration optimization method is shown in Algorithm 1.

Algorithm 1: Joint-domain alignments based tensor invariant subspace learning

---

Input: : tensor set $\mathcal{Z}$ in both domains;

Output: reconstructed $\hat{\mathcal{Z}}$

Initialize: tensor subspace $\mathcal{U}$;

alignment matrices $\mathcal{M}$;

regularization parameter λ;

dim of subspace $\{d_1, d_2, d_3\}$;

maximum iteration *T*.

1 Calculate the loss and accuracy according to (9);

2 **for** $t = 1$ to T **do**

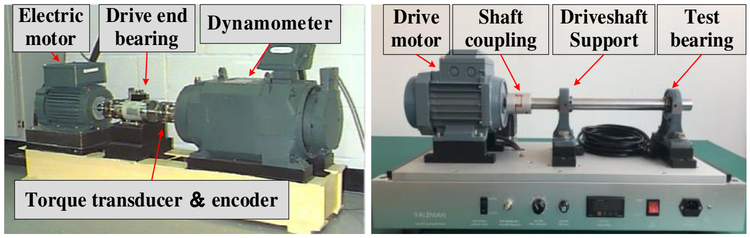
```
    fix 𝓜 to solve 𝓤 via (10);
    calculate the loss and accuracy;
    fix 𝓤, 𝓒_s, 𝓒_t and update each M of 𝓜 by solving (12);
    calculate the loss and accuracy;
    check the convergence;
    if equation (5) convergence then
        break;
    end
    t = t + 1;
end
return reconstructed Ẑ
```

IV. Experiments

Condition monitoring and fault diagnosis are important for rolling bearings that is key components of many industrial systems. This study uses two real datasets of rolling bearings to verify the performance of the proposed method

A. Datasets and Experiment Setups

1) Public Dataset: The dataset is acquired from the bearing data center of Case Western Reserve University (CWRU) [36], which is recognized as a standardized bearing dataset; the test bench is shown in Fig. 3(a). The collected vibration data mainly include three types of data: drive end accelerometer data (DE), fan end accelerometer data (FE), and base accelerometer data (BA). The data were sampled with 12KHz and 48KHz for the drive end bearing experiments. All fan end bearing data were collected at a frequency of 12KHz. The data used in this paper are all collected at 12KHz frequency, including a set of normal data (NR), only containing FE and DE, and three types of fault data, namely fault of inner race (FIR), fault of outer race (FOR), and fault of rolling ball (FRB). Each type of fault status includes three fault diameters, namely, 0.007 inch, 0.014 inch, and 0.021 inch. The selected samples include the vibration data of the bearing under the motor load of 0~3hp (corresponding to the motor speed of 1797 RPM, 1772 RPM, 1750 RPM, and 1730 RPM, respectively), so the four domains can be constructed according to the load. Each domain contains 10 types of fault information, and each type of fault contains a series of samples composed of the vibration data. With these data, twelve DA tasks (from DT1 to DT12) are constructed and used to verify the performance of the proposed method. The details of the twelve DA tasks are shown in Table I, where for each task, two domains are generated according to the different motor loads. The source domain consists of the data obtained from four bearing states (NR, FIR, FOR, and FRB) under the same motor load with the three fault diameters (0.007, 0.014, 0.021 inch), and the labels are defined as {1, 2,...,10}.Similarly, the target domain also has 10 types of data generated under another motor load. The goal is to predict the status label of the unlabeled target domain data.



(a) (b)

Fig. 3. Rolling bearing test platform. (a) CWRU test platform. (b) PT100 test platform.

TABLE I

TASKS AND THEIR ASSOCIATED FAULT LABELS BASED ON CWRU DATASET

Task	Domain shift	NR label	0.007 inches fault label	0.014 inch fault label	0.021 inch fault label	Task	Domain shift	NR label	0.007 inch fault label	0.014 inch fault label	0.021 inch fault label
DT1	0hp→1hp	1	2, 3, 4	5, 6, 7	8, 9, 10	DT7	1hp→0hp	1	2, 3, 4	5, 6, 7	8, 9, 10
DT2	0hp→2hp	1	2, 3, 4	5, 6, 7	8, 9, 10	DT8	2hp→0hp	1	2, 3, 4	5, 6, 7	8, 9, 10
DT3	0hp→3hp	1	2, 3, 4	5, 6, 7	8, 9, 10	DT9	3hp→0hp	1	2, 3, 4	5, 6, 7	8, 9, 10
DT4	1hp→2hp	1	2, 3, 4	5, 6, 7	8, 9, 10	DT10	2hp→1hp	1	2, 3, 4	5, 6, 7	8, 9, 10
DT5	1hp→3hp	1	2, 3, 4	5, 6, 7	8, 9, 10	DT11	3hp→1hp	1	2, 3, 4	5, 6, 7	8, 9, 10
DT6	2hp→3hp	1	2, 3, 4	5, 6, 7	8, 9, 10	DT12	3hp→2hp	1	2, 3, 4	5, 6, 7	8, 9, 10

2) *PT100 Platform Test Dataset:* This dataset was collected from our laboratory's rolling bearing failure test platform, and contained vibration data of six health states, namely NR, FIR, FOR, FRB, fault of bearing cage (FBC), and fault of composite bearing (FCB). For each state, the data of the bearing running at three different speeds (1800RPM, 2100RPM, and 2400RPM) were collected at 48KHz. The collected data include the coupling end and the non-driving end (that is, the test bearing end). The test platform is shown in Fig. 3(b). With these data, the four domains can be constructed, and each domain contains six types of fault information. The other six DA tasks are used for further experimental verification, and the details of the six tasks are shown in Table II, where the five types of fault labels (2, 3, 4, 5, and 6) represent FIR, FOR, FRB, FBC, and FCB, respectively. The source and target domains are generated under different bearing speeds, respectively.

B. *Comparison Methods*

In order to confirm the effectiveness of the proposed method, its performance was compared with that of many traditional methods without using domain adaptation and mainstream domain adaptation schemes.

1) Without DA: In this work, SVM, Softmax logistic regression, k-means [37], and CNN [38] are chosen to use, which are four conventional methods that can be utilized to realize the fault diagnosis. These classifiers are trained directly with labeled source data and applied to the target domain.

TABLE II

TASKS AND THEIR ASSOCIATED FAULT LABELS BASED ON PT100 PLATFORM TEST DATASET

Task	Domain shift	NR label	Fault label
PT1	1800 RPM→2100 RPM	1	2, 3, 4, 5, 6
PT2	1800 RPM→2400 RPM	1	2, 3, 4, 5, 6
PT3	2100 RPM→2400 RPM	1	2, 3, 4, 5, 6
PT4	2100 RPM→1800 RPM	1	2, 3, 4, 5, 6
PT5	2400 RPM→1800 RPM	1	2, 3, 4, 5, 6
PT6	2400 RPM→2100 RPM	1	2, 3, 4, 5, 6

2) Correlation Alignment (CORAL) [39]: As a popular domain adaptive method, the CORAL method explores the second-order statistics of the distribution of source and target domains, and aligns the distribution of input features to minimize domain shift.

3) Joint Distribution Adaptation (JDA) [40]: The main idea of this method is to jointly adapt the marginal distribution and the conditional distribution in the process of dimensionality reduction to minish the difference between domains and construct the new features. This is somehow similar to the idea used in the present study.

4) Subspace Alignment (SA) [41]: This method creates the subspace for the source and target domains, and adjusts the basis of the subspace globally by learning a mapping function to align the source subspace with the target subspace. The method is regularized in nature, so no regularization parameters need to be adjusted.

5) Easy Transfer Learning (EasyTL) [42]: The EasyTL method learns non-parametric transfer features and classifiers by using the intra-domain structure. This process does not require model selection and hyperparameter adjustment.

6) Stratified Transfer Learning (STL) [43]: The STL method is designed to solve the cross-domain diagnosis problem that only considers global domain transfer. In the process of global domain transfer, STL uses the internal affinity of the class to realize the knowledge transfer within the class, and realizes the conversion of the same class in the source and target domains into the same subspace.

The proposed method, together with all the above seven methods, is applied to the two datasets: CWRU dataset and PT100 dataset. The performances of the compared methods are presented in following two sections.

C. Performance Comparison of Different Methods on the CWRU Dataset

1) Diagnosis Results

The detailed results of the accuracy of the nine methods for twelve cross-domain fault diagnosis tasks are shown in Table III, where the column "Average" represents the average accuracy of each method for all the 12 tasks. For a more intuitive comparison, the detailed comparison results are also represented by histograms, as shown in Fig. 4. From Table III and Fig. 4, it can be noticed that the accuracy of the proposed method in different tasks is above 95%, and even reaches 100% in tasks DT4 and DT10; these confirm the effectuality of the proposed TDA fault diagnosis framework. The four methods without using DA technology show lower accuracy on most tasks. For example, the highest accuracy of SVM and Softmax is 75.536% and 84.536% for the 12 tasks, respectively. The k-means method has the lowest average accuracy among all the methods; it is only 62.025%, which is 36.259% lower than the proposed method. For the CNN method, its accuracy for task DT2 reaches 88.875%, which is highest among the first four methods. In addition, the average accuracy of the CNN method is also higher than that of other methods with DA technology, but is still 19.304% lower than that of the proposed method. It is worth mentioning that the DA based methods train models in the source domain with labeled samples, and perform fault diagnosis in the target domain directly. For such a scenario, it is difficult to achieve more accurate classification results when the data of the source and target domains do not follow the same distribution.

Among the other five DA methods, the average accuracy of JDA and SA is slightly better than the other three methods. The best accuracy rate of JDA reaches 97.036%, but its accuracy rate fluctuates around 85% in general. For STL, in tasks DT8-DT12, the fluctuation of the test results is about 68%. The reason for such a fluctuation phenomenon is that the negative migration occurred in the learning process of the model parameters. These DA methods utilize the traditional vector representation to learn fault information of rolling bearings, they have limited capability for fault diagnosis and may not always show satisfactorily high diagnostic accuracy in real applications. Different from the traditional DA methods, the proposed TDA fault diagnosis method explores the power of tensor decomposition to fully make use of the sampling signals of rolling bearings,

TABLE III

TESTING RESULTS (%) OF DIFFERENT METHODS ON THE CWRU DATASET

Method	DT1 (%)	DT2 (%)	DT3 (%)	DT4 (%)	DT5 (%)	DT6 (%)	DT7 (%)	DT8 (%)	DT9 (%)	DT10 (%)	DT11 (%)	DT12 (%)	Average (%)	Time (s)
SVM	75.536	58.750	73.054	71.250	77.304	59.339	67.786	56.857	52.982	51.929	64.589	54.696	**63.673**	**2.377**
Softmax	80.411	71.232	84.536	70.321	71.054	67.964	70.161	70.143	84.536	69.643	79.268	70.161	**74.119**	**1.119**
k-means	56.036	42.857	58.750	71.357	71.375	57.143	57.375	57.339	73.446	71.429	70.464	57.054	**62.052**	**12.782**
CNN	78.929	88.875	80.196	72.571	63.893	73.821	81.411	74.357	85.893	80.643	75.946	86.161	**78.558**	**137.463**
CORAL	86.589	81.625	**92.661**	77.929	83.089	70.018	80.321	75.393	**93.268**	79.321	84.411	76.107	**81.728**	**2.896**
JDA	85.268	88.482	85.714	85.161	85.482	83.196	86.089	83.304	**97.036**	83.518	88.321	85.107	**86.390**	**13.113**
SA	85.839	83.054	85.304	**91.339**	85.661	89.929	83.714	85.286	88.804	89.679	86.179	85.643	**86.702**	**13.830**
EasyTL	79.500	76.696	**90.768**	79.500	80.268	70.429	71.732	67.036	67.839	69.982	69.982	67.893	**74.302**	**5.211**
STL	85.500	81.643	82.571	82.196	85.696	83.232	85.536	82.821	83.554	81.893	80.911	82.036	**83.132**	**33.400**
Proposed	**99.625**	**97.261**	**96.255**	**100.000**	**97.025**	**98.042**	**94.382**	**98.609**	**95.024**	**100.000**	**96.308**	**95.666**	**97.350**	**90.903**

Accepted for publication by *Reliability Engineering & System Safety* on 7 November 2022

method

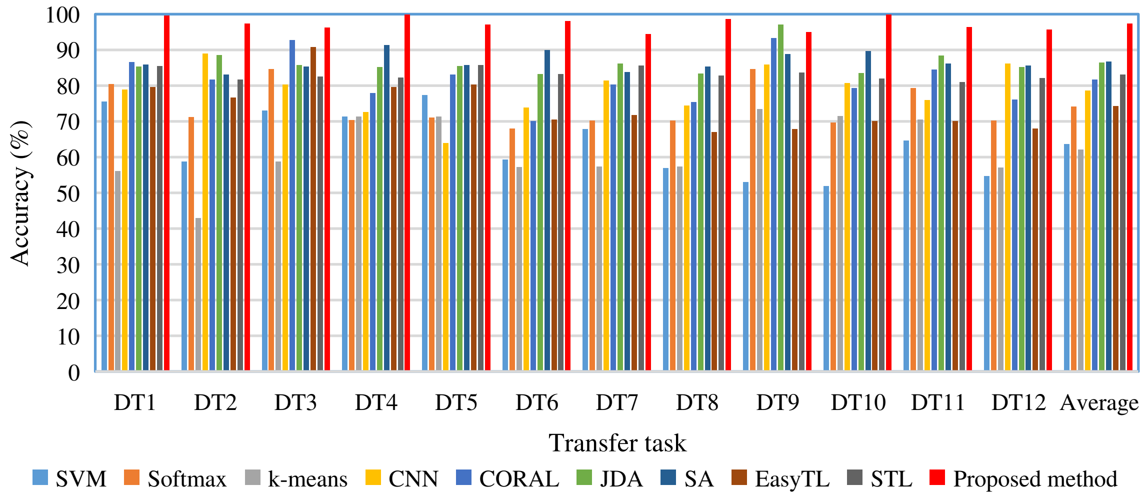


Fig. 4. Classification results of the twelve tasks on the CWRU dataset.

and hence obtains more knowledge about the fault information. Moreover, the tensor decomposition technology is applied to obtain the tensor subspace. Dynamic distribution alignment is introduced to dynamically align the shared subspace of the source and target domains simultaneously to alleviate excessive data distortion caused by forced alignment, overcoming the limitation of the traditional vector representation. Therefore, the proposed method can identify fault categories more accurately, and reaches an average accuracy rate of 97.350% in all the tasks, which is 12.281% higher than that of SA method, and 31.019% higher than that of EasyTL method. Overall, the proposed method performs better in cross-domain diagnosis tasks and has better stability performance; this is because the data representation of the model and the subspace learning of joint domain alignment provide advantages for data validity and feature mining.

In addition, the average processing time of each method was also calculated and is shown in Table III (the last column). From this table, the Softmax method needs the shortest processing time (1.119s) and the CNN method needs the longest processing time (137.463s), this is because the CNN model has a large number of parameters to optimize in the training process. Compared with all the other methods (except CNN), the processing time of the proposed method is higher, and most of its processing time is spent on tensor alignment and subspace learning, so as to obtain better domain adaptation learning.

2) Feature Visualization

In order to perform a qualitative analysis of the proposed method and prove the transferability of features, the statistical method of t-distributed stochastic neighbor embedding (t-SNE) [44] is used to map the samples in the feature space to a two-dimensional space to realize the visualization representation of high-dimensional data. Taking the task DT3 as an example, the visualization effect of the original features and the extracted features of the four bearing states (NR, FIR, FOR, and FRB) are displayed in Fig. 5(a)-(d), where the legends "Original_S" and "Original_T" represent the original features of the source and target domains, respectively. Similarly, the legends "TDA_S" and "TDA_T" represent the source and target features extracted by the proposed method, respectively.

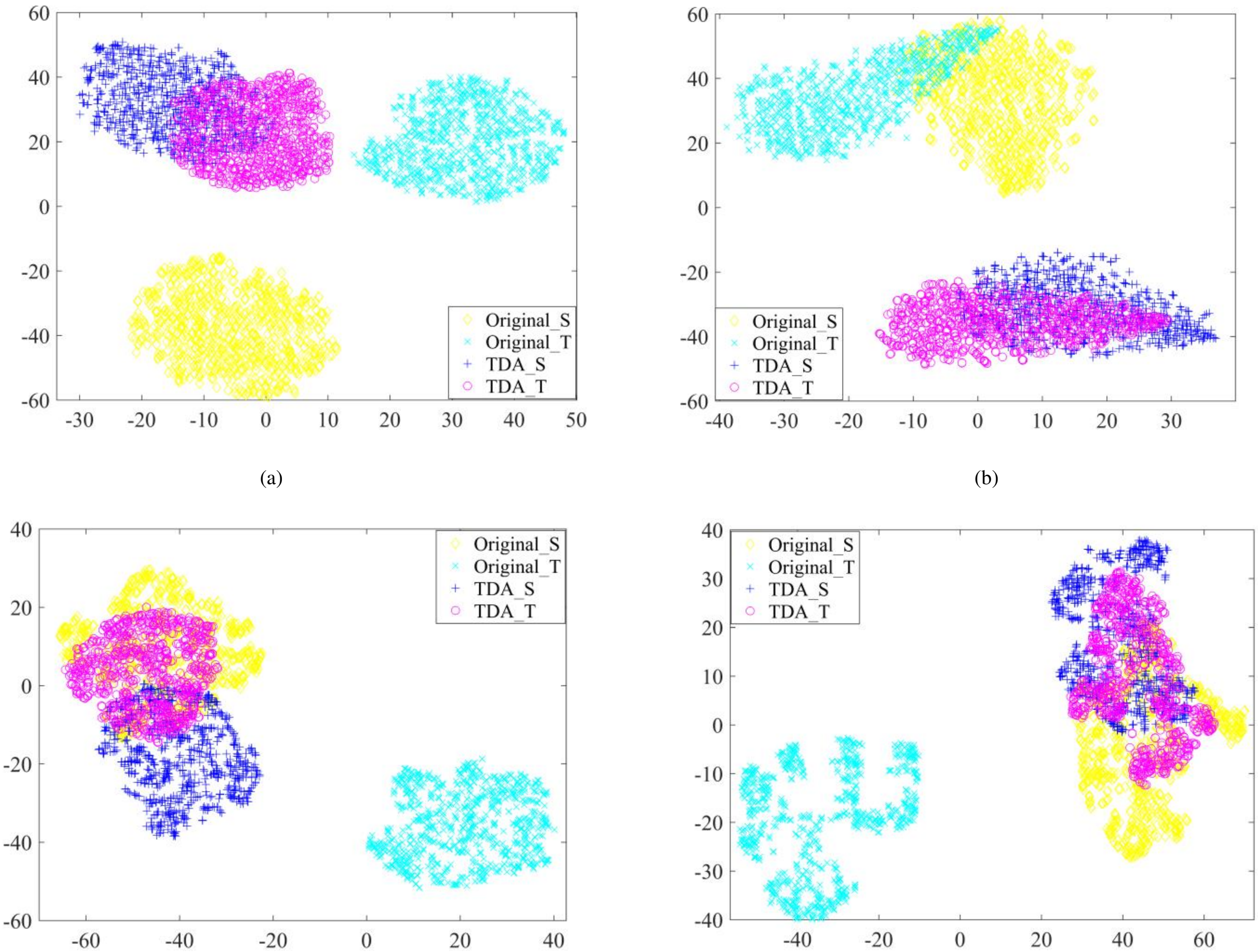


Fig. 5. Feature visualization via t-SNE. (a) NR; (b) FIR; (c) FOR; (d) FRB.

According to the visualization results, it can be clearly observed that the alignment effect of the features extracted by the proposed model in the domains is significantly improved compared with the original features. Specifically, it can be seen that only the FIR state has a small difference between the source and target domains, while the distribution of the other three health states has a large difference between the domains. The problem of this kind of domain difference is usually the main reason that many diagnostic methods perform poor. In the proposed TDA method, the domain differences are significantly reduced, especially in the categories of FIR and FRB, which are well aligned. This also indicates that the proposed method can cluster and align the data of the same categories and separate the data of different categories, thus effectively reducing the domain differences.

3) Parameter Analysis

Here the spatial factors are considered and the data are selected from DT1-DT6 tasks to perform the sensitivity analysis of the three parameters involved in the proposed method. Specifically, the three parameters are spatial mode dimensionality d (d=d_1=d_2), feature mode dimensionality d_3, and trade-off parameter λ. It should be noted that each sample is mapped to a 5×5×144 third-order tensor, where d_1=d_2≤5 refers to spatial locations, d_3≤144 corresponds to features, and the subspace dimension is set to 5×5×40.

a) Spatial Mode Dimensionality: In order to determine the effect of spatial information of the model, the experimental results under different spatial dimensions in the six DA tasks are drawn in Fig. 6. It can be observed that with the increase of *d*, the classification accuracy of the tasks DT3, DT5, and DT6 show an overall upward trend. It is very interesting that the classification performance in the tasks DT1, DT2, and DT4 is hardly affected by the change of *d*. In general, the increase of *d* is conducive to improving the classification accuracy, which means that preserving the original spatial pattern is necessary to find potential commonalities in low-dimensional subspaces.

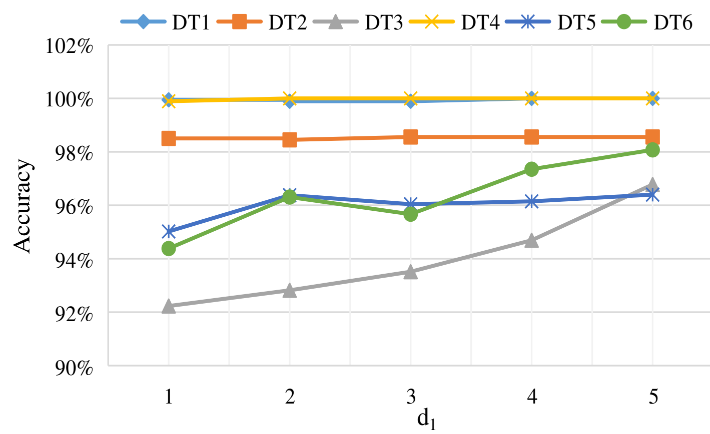


Fig. 6. Sensitivity of the spatial mode dimensionality

b) Feature Mode Dimensionality: The effect of feature mode dimensionality on cross-domain diagnosis performance is presented in Fig. 7. It can be observed that in different tasks, when the value of d_3 is around 40, the test accuracy is relatively higher. Specifically, when d_3 increases from 10 to 40, the tasks DT1, DT4, and DT5 show an increasing trend; the classification accuracy for the task DT5 is obviously faster than for other tasks. However, for the tasks DT2 and DT3 the accuracy decreases first and then increases.

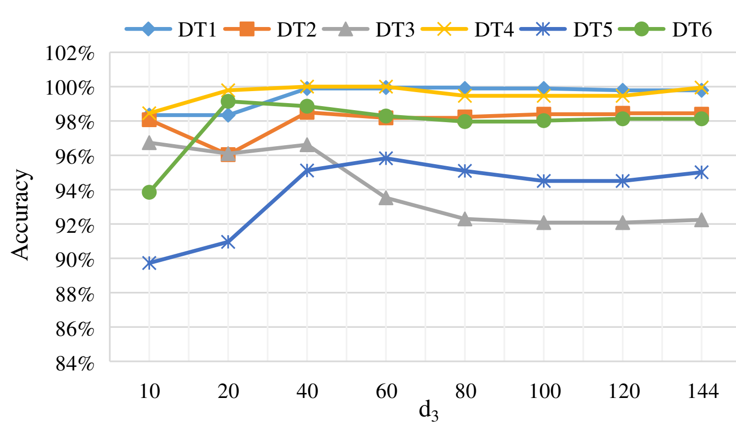


Fig. 7. Sensitivity of the feature mode dimensionality

As the value of d_3 continues to increase, the accuracy of the six tasks gradually becomes stable. Unfortunately, the accuracy for the task DT3 is about 4% lower when $d_3 = 80$ than that of $d_3 = 40$. Therefore, so as to efficaciously capture the information of the original data, the feature mode dimensionality is chosen as 40.

c) Trade-off Parameters : In order to clearly show the impact of the trade-off parameter on the model performance, the different λ values of TDA method are tested in this section, and the test results for different tasks are shown in Fig. 8. Obviously, when λ varies from $1e^{-5}$ to 1, the classification accuracy in the tasks DT1, DT4, DT5, and DT6 only fluctuates slightly, showing that the accuracy is not significantly affected. However, the tasks DT2 and DT3 both exhibited about 4% fluctuations. Overall, for all the tasks, the accuracy is higher than 95% when $\lambda \in [1e^{-4}, 1e^{-1}]$, which provides a basis for the selection of the parameter.

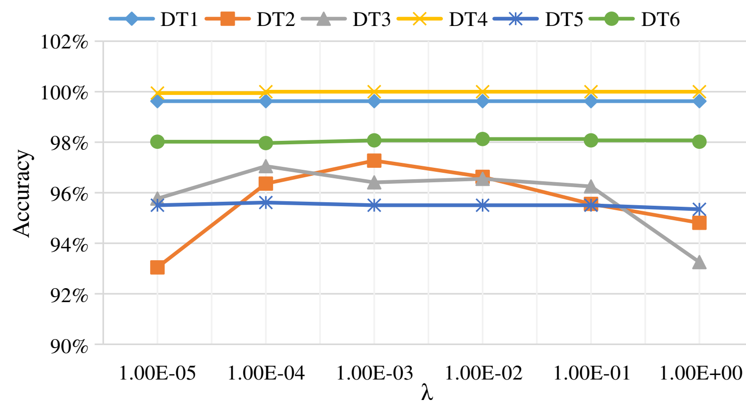


Fig. 8. Effects of tradeoff parameters of regularization terms

D. Validation on the PT100 Dataset and Performance Analysis

To further test the performance of the proposed method, the modelling experiments were carried out on the PT100 dataset. The experimental results generated by the proposed method and the nine compared methods on six transfer tasks (shown in Table II) are shown in Fig. 9, where the term "Average" represents the average accuracy of six tasks for each method.

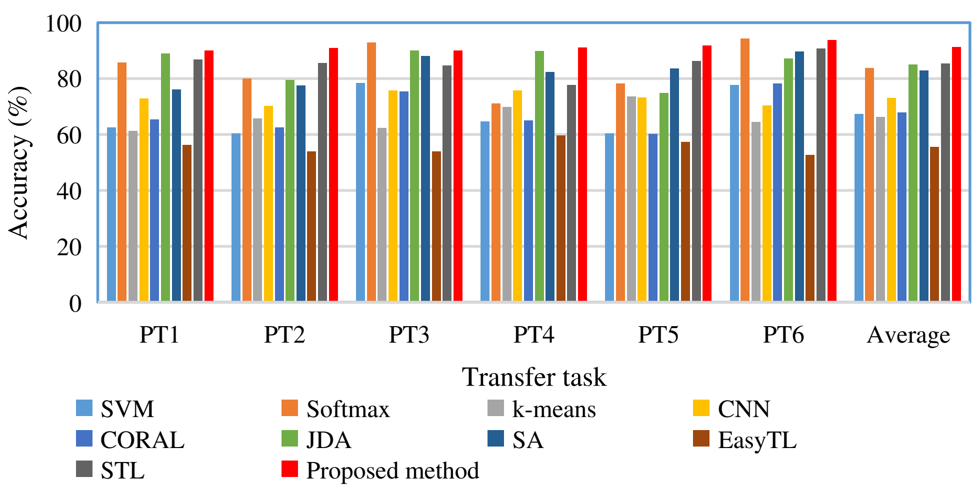
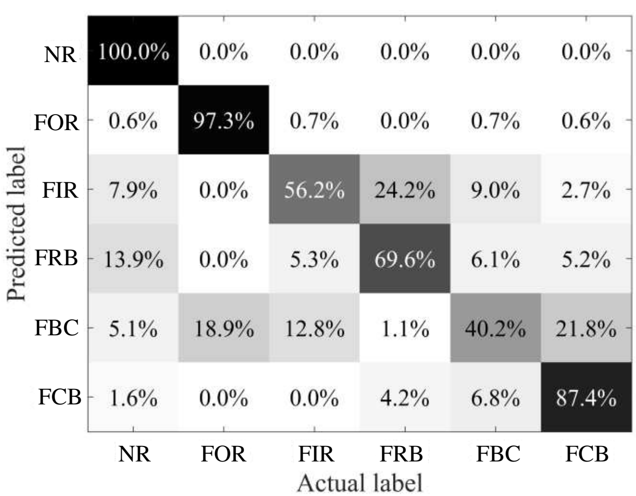


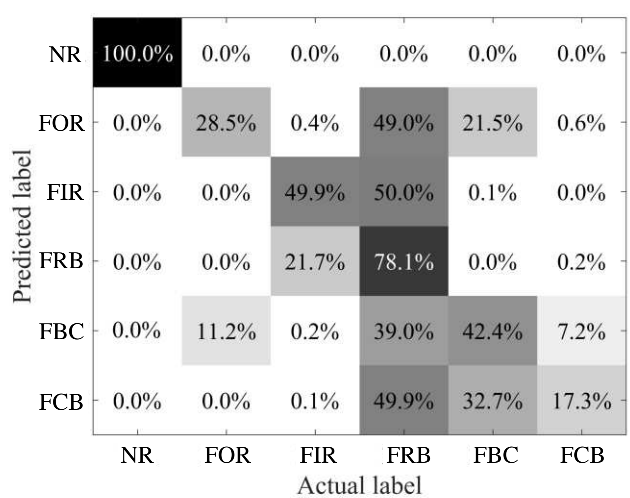
Fig. 9. Classification results of the six tasks on the PT100 platform.

It can be observed intuitively that the average accuracy of the proposed method is about 91.3%, which is better than that of other methods. Specifically, the average classification accuracy of Softmax is about 15% higher than that of SVM, and even reaches 94% at its best. In addition, it is difficult for the k-means and CNN methods to obtain higher accuracy, and the CNN method also needs far more time to train the model. Among the DA methods, JDA, SA, and STL have achieved relatively satisfactory results, while CORAL and EasyTL show the poor performance and even negative transfer. Overall, the proposed method shows excellent performance, but compared with the results for the CWRU dataset, the classification accuracy is generally relatively lower.

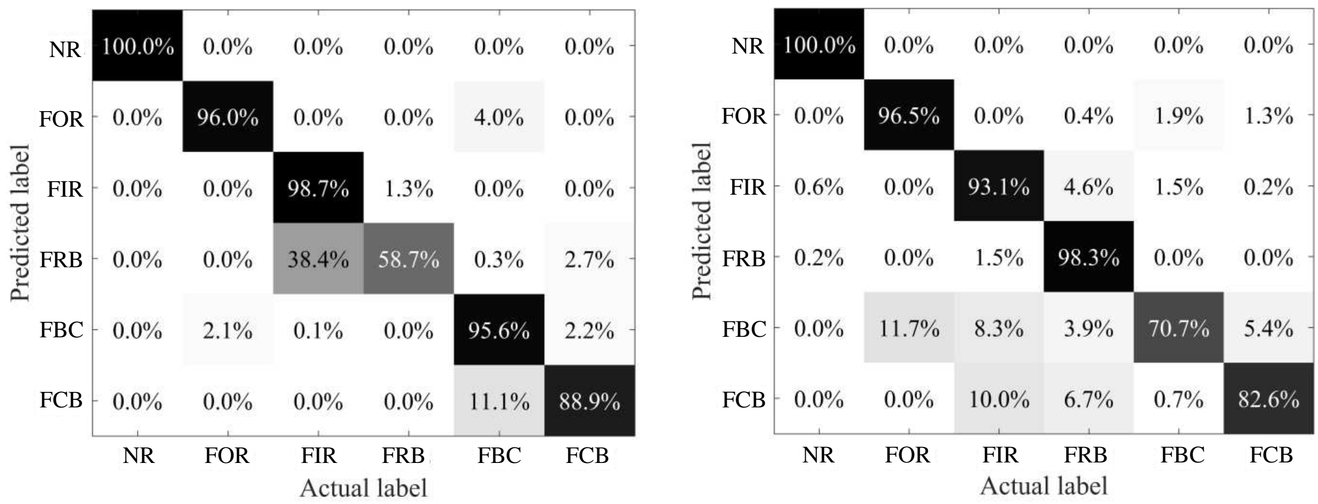
Therefore, in order to understand the causes of the performance degradation, the confusion matrices obtained by the four DA methods (JDA, EasyTL, SA, and TDA) on task PT3 are analyzed. The confusion matrices of the test results are displayed in Fig. 10 to show the prediction accuracy rates of each method on six categories. Unsurprisingly, the four methods correctly identified the normal state of the bearing. For JDA, this method can well deal with three types of faults, but has poor performance on the FIR, FRB, and FBC faults. The reason is that the three types of fault data show large fluctuations and non-stationary behaviors. For EasyTL, its misclassification rate is very high resulting the classification accuracy of FBC is only 17.3%, which again shows that the method has negative transfer in the self-learning process of the model and parameters. The SA method has a good overall performance in the test results, but poor performance on the FRB faults. The proportion of misclassification of FRB as FIR is 38.4%, which is because the two types of faults have some similarities in their basic features. Compared with SA, the proposed method TDA method significantly improves the recognition of FRB faults, but shows relatively poor performance on faults FBC and FCB.



(a)



(b)



(c) (d)

Fig. 10. Confusion matrices (%) of the testing results. (a) JDA; (b) EasyTL; (c) SA; (d) TDA.

Note that in comparison with the CWRU dataset, the PT100 dataset includes two new types of faults: FBC and FCB. The inclusion of the two new types of faults makes the associated classification and fault diagnosis tasks be more realistic, and meanwhile make the tasks more challenging.

In general, compared with the compared methods, the proposed TDA method can achieve better performance on the PT100 dataset.

V. Conclusion

To better solve the fault diagnosis problem of rolling bearings in modern industry, a novel TDA method was proposed for reducing the effects of data distortion and domain shift in the diagnosis process under variable conditions. In this proposed model, the third-order tensor model of the original data was constructed by comprehensively using the time domain signals, the frequency domain signals, and the Hilbert marginal spectrum. With the proposed tensor representation method, the important structural information of bearing data was fully explored and integrated to improve the fault diagnosis ability. The source and target domains were then simultaneously aligned to the shared subspace by using the tensor decomposition under orthogonal constraints, together with joint-domain alignments based tensor invariant subspace learning. Finally, the classifier was built to identify the fault categories. The main advantage of the proposed TDA method is that it can effectively learn the transferable features, significantly reduce the cross-domain differences, and at the same time strengthen the identifiable information contained in the original data. The proposed method has been applied to two real datasets, and its performance is compared with nine methods. Experimental results confirm the superiority of the proposed TDA method to the nine compared methods. As a domain adaption technique, the proposed TDA model is trained on a single source domain first and then it is applied to a single target domain. Therefore, the proposed method is limited to applications where there is a single source domain and single target domain. In addition, this method consumes more training time than other simple domain adaptation methods. Therefore, future work will focus on extending the method so it can effectively learn knowledge in multiple domains and improving its computing efficiency in practical industrial applications.

CRediT authorship contribution statement

Zhao-Hua Liu: Conceptualization, Methodology, Funding acquisition, Resources, Supervision. **Liang Chen:** Methodology, Software, Validation, Writing - Original Draft, Investigation. **Hua-Liang Wei:** Writing - Review & Editing. **Fa-Ming Wu:** Validation, Funding acquisition, Project administration, Data Curation. **Lei Chen:** Visualization, Project administration. **Ya-Nan Chen:** Validation, Funding acquisition, Formal analysis.

Declaration of interest

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

Acknowledgments

This work was supported by the National Key Research and Development Project of China [grant number 2019YFE0105300]; and the National Natural Science Foundation of China [grant number 61972443 and Grant number 62103143].

References

[1] F. Ferracuti, A. Freddi, A. Monteriù, and L. Romeo, "Fault diagnosis of rotating machinery based on wasserstein distance and feature selection," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 1997-2007, Jul. 2022.

[2] L. Zuo, F. Xu, C. Zhang, T. Xiahou, and Y. Liu, "A multi-layer spiking neural network-based approach to bearing fault diagnosis," *Reliability Engineering and System Safety*, vol. 225, Art no. 108561, 2022, doi: 10.1016/j.ress.2022.108561.

[3] Y. Xu, X. Yan, B. Sun, and Z. Liu, "Global contextual residual convolutional neural networks for motor fault diagnosis under variable-speed conditions," *Reliability Engineering and System Safety*, vol. 225, Art no. 18618, 2022, doi: 10.1016/j.ress.2022.108618.

[4] L. Wen, L. Gao, and X. Li, "A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis, *IEEE Trans. Syst. Man Cybern. Syst*., vol. 49, no. 1, pp. 136-144, Jan. 2019

[5] J. Li, R. Huang, G. He, Y. Liao, Z. Wang and W. Li, "A two-stage transfer adversarial network for intelligent fault diagnosis of rotating machinery with multiple new faults," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1591-1601, June 2021, doi: 10.1109/TMECH.2020.3025615.

[6] L. Chen, Q. Li, C. Shen, J. Zhu, D. Wang, and M. Xia, "Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1790-1800, Mar. 2022.

[7] H. O. A. Ahmed and A. K. Nandi, "Three-stage hybrid fault diagnosis for rolling bearings with compressively sampled data and subspace learning techniques," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 7, pp. 5516-5524, Jul. 2019.

[8] H. Karamti, M. M. A. Lashin, F. M. Alrowais, and A. M. Mahmoud, "A new deep stacked architecture for multi-fault machinery identification with imbalanced samples," *IEEE Access*, vol. 9, pp. 58838-58851, 2021.

[9] Z. K. Abdul, A. K. Al-Talabani, and D. O. Ramadan, "A hybrid temporal feature for gear fault diagnosis using the long short term memory," *IEEE Sensors J.*, vol. 20, no. 23, pp. 14444-14452, 1 Dec. 2020.

[10] F. Pacheco, A. Drimus, L. Duggen, M. Cerrada, D. Cabrer, and R.-V. Sanchez, "Deep ensemble-based classifier for transfer learning in rotating machinery fault diagnosis," *IEEE Access*, vol. 10, pp. 29778-29787, 2022.

[11] Z. H. Lei, G. R. Wen, S. Z. Dong, X. Huang, H. X. Zhou, Z. F. Zhang, and X. F. Chen, "An intelligent fault diagnosis method based on domain adaptation and its application for bearings under polytropic working conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-14, 2021, doi: 10.1109/TIM.2020.3041105.

[12] B. Rezaeianjouybari, and Y. Shang, "A novel deep multi-source domain adaptation framework for bearing fault diagnosis based on feature-level and task-specific distribution alignment," *Measurement*, vol. 178, Art. no. 109359, 2021.

[13] M. Azamfar, X. Li, and J. Lee, "Intelligent ball screw fault diagnosis using a deep domain adaptation methodology," *Mechanism and Machine Theory*, vol. 151, 103932, 2020.

[14] A. Ainapure, X. Li, J. Singh, Q. Yang, and J. Lee, "Enhancing intelligent cross-domain fault diagnosis performance on rotating machines with noisy health labels," *Procedia Manufacturing*, vol. 48, pp. 940-946, 2020.

[15] A. K. Sharma and N. K. Verma, "Quick learning mechanism with cross-domain adaptation for intelligent fault diagnosis," I*EEE Transactions on Artificial Intelligence*, vol. 3, no. 3, pp. 381-390, Jun. 2022.

[16] S. Schwendemann, Z. Amjad, and A. Sikora, "Bearing fault diagnosis with intermediate domain based Layered Maximum Mean Discrepancy: A new transfer learning approach," *Engineering Applications of Artificial Intelligence*, vol. 105, Art. no. 104415, 2021.

[17] C. Raab, M. Röder, and F.-M. Schleif, "Domain adversarial tangent subspace alignment for explainable domain adaptation," *Neurocomputing*, vol. 506, pp. 418-429, 2022.

[18] H. Lu, L. Zhang, Z. G. Cao, W. Wei, K. Xian, C. H. Shen, and A. van den Hengel, "When unsupervised domain adaptation meets tensor representations," in *Proc. IEEE Conf. Comput. Vis.*, Oct. 2017, pp. 599-608, doi: 10.1109/ICCV.2017.72.

[19] C. F. Hu, and Y. X. Wang, "Multidimensional denoising of rotating machine based on tensor factorization." *Mech. Syst. Signal Process.*, vol. 122, pp. 273-289, 2019.

[20] L. Luo, L. Xie and H. Su, "Deep learning with tensor factorization layers for sequential fault diagnosis and industrial process monitoring," *IEEE Access*, vol. 8, pp. 105494-105506, 2020, doi: 10.1109/ACCESS.2020. 3000004.

[21] Z. Y. He, H. D. Shao, J. S. Cheng, Y. Yang, and J. W. Xiang, "Kernel flexible and displaceable convex hull based tensor machine for gearbox fault intelligent diagnosis with multi-source signals," *Measur.*, vol. 163, pp. 1-10, 2020, doi: https://doi.org/10.1016/j.measurement.2020. 107965.

[22] H. S. Zhao, and W. Zhang, "Fault diagnosis method for rolling bearings based on segment tensor rank-(Lr, Lr, 1) decomposition," *Mech. Syst. Signal Process.*, vol. 132, pp. 762-775, 2019.

[23] S. Mohammadi, D. F. Gleich, T. G. Kolda and A. Grama, "Triangular alignment (TAME): A tensor-based approach for higher-order network alignment," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 14, no. 6, pp. 1446-1458, 2017, doi: 10.1109/TCBB.2016.2595583.

[24] G. M. Gao and Y. F. Gu, "Tensorized principal component alignment: a unified framework for multimodal high-resolution images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 46-61, Jan. 2019, doi: 10.1109/TGRS.2018.2852066.

[25] C. C. Jia, M. Shao and Y. Fu, "Sparse canonical temporal alignment with deep tensor decomposition for action recognition," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 738-750, Feb. 2017, doi: 10.1109/TIP.2016.2621664.

[26] W. Wang and M. Zhang, "Tensor deep learning model for heterogeneous data fusion in Internet of Things," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 4, no. 1, pp. 32-41, Feb. 2020, doi: 10.1109/ TETCI.2018.2876568.

[27] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *Society for Industrial and Applied Mathematics review*, vol. 51, no. 3, pp. 455-500, Jan. 2009, doi: 10.1137/07070111X.

[28] H. Lu, C. Shen, Z. Cao, Y. Xiao and A. van den Hengel, "An embarrassingly simple approach to visual domain adaptation," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3403-3417, Jul. 2018, doi: 10.1109/TIP.2018.2819503.

[29] Q. Kang, S. Y. Yao, M. C. Zhou, K. Zhang and A. Abusorrah, "Enhanced subspace distribution matching for fast visual domain adaptation," *IEEE Trans. Comput. Soc. Syst.,* vol. 7, no. 4, pp. 1047-1057, Aug. 2020, doi: 10.1109/TCSS.2020.3001517.

[30] B. Schölkopf; J. Platt; and T. Hofmann, "Analysis of representations for domain adaptation," in Advances in *Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, MIT Press, 2007, pp.137-144.

[31] T. Yoon, J. Lee and W. Lee, "Joint transfer of model knowledge and fairness over domains using wasserstein distance," *IEEE Access*, vol. 8, pp. 123783-123798, 2020, doi: 10.1109/ACCESS.2020.3005987.

[32] S. J. Pan, I. W. Tsang, J. T. Kwok and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199-210, Feb. 2011, doi: 10.1109/TNN.2010.2091281.

[33] Y. G. Lei, F. Jia, J. Lin, S. B. Xing and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.,* vol. 63, no. 5, pp. 3137-3147, May 2016, doi: 10.1109/TIE.2016.2519325.

[34] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network". *Computer Science*, vol. 14, no. 7, pp. 38-39, 2015, doi: 10.4140/TCP.n.2015.249.

[35] Z. Wen, W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, pp. 397–434, 2013, doi: 10.1007/s10107-012-0584-1.

[36] K. Loparo. Case Western Reserve University Bearing Data Center, (2013), [online]. Available: http://csegroups.case.edu/bearingdatacenter/ pages/12k-drive-end-bearing-fault-data.

[37] N. R. Dreher, I. O. de Almeida, G. C. Storti, G. B. Daniel, and i. H. Machado, “Feature analysis by k-means clustering for damage assessment in rotating machinery with rolling bearings,” *J. Braz. Soc. Mech. Sci. Eng.*, vol. 44, Art. no. 330, 2022, doi: 10.1007/s40430-022-03637-1.

[38] R. M. Souza, E. G. S. Nascimento, U. A. Miranda, W. J. D. Silva, and H. A. Lepikson, “Deep learning for diagnosis and classification of faults in industrial rotating machinery,” *Computers & Industrial Engineering*, vol. 153, Art. no. 107060, 2021, doi: 10.1016/j.cie.2020.107060.

[39] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” *in Proc. Amer. Assoc. Artif. Intell. Conf.,* Dec. 2016, pp. 2058–2065.

[40] M. S, Long, J. M. Wang, G. G. Ding, J. G. Sun and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Conf. Comput. Vis.*, Dec. 2013, pp. 2200-2207, doi: 10.1109/ICCV.2013.274.

[41] B. Fernando, A. Habrard, M. Sebban and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Conf. Comput. Vis.*, Dec. 2013, pp. 2960-2967, doi: 10.1109/ICCV.2013.368.

[42] J. D. Wang, Y. Q. Chen, H. Yu, M. Y. Huang and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," in *Proc. Int. Conf. Multimed. Expo.*, Jul. 2019, pp. 1210-1215, doi: 10.1109/ICME.2019. 00211.

[43] J. D. Wang, Y. Q. Chen, L. Hu, X. H. Peng and P. S. Yu, "Stratified transfer learning for cross-domain activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2018, pp. 1-10, doi: 10.1109/ PERCOM.2018.8444572.

[44] V. D. M. Laurens and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res*., vol. 9, no. 2605, pp. 2579-2605, 2008.