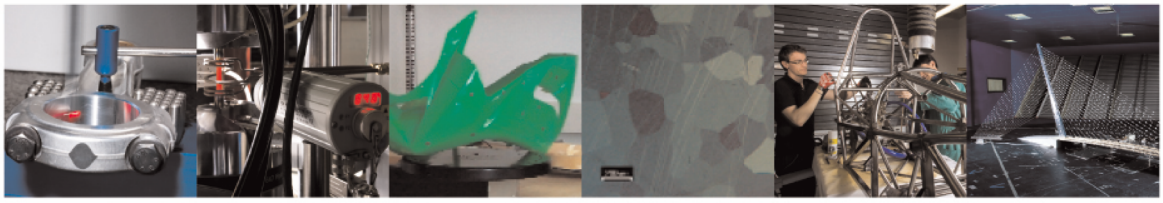


A universal multi-source domain adaptation method with unsupervised clustering for mechanical fault diagnosis under incomplete data

Tian, Jinghui;Han, Dongying;Karimi, Hamid Reza;Zhang, Yu;Shi, Peiming

This is a post-peer-review, pre-copyedit version of an article published in Neural Networks. The final authenticated version is available online at:
https://doi.org/10.1016/j.neunet.2024.106167

A universal multi-source domain adaptation method with unsupervised clustering for mechanical fault diagnosis under incomplete data

Jinghui Tian[a], Dongying Han[a,*], Hamid Reza Karimi[b], Yu Zhang[a], Peiming Shi[c]

[a]*School of Vehicles and Energy, Yanshan University, Qinhuangdao, Hebei 066004, PR China*

[b]*Department of Mechnical Engineering, Politecnico di Milano, 20156 Milan, Italy*

[c]*School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, PR China*

Abstract

Recently, due to the difficulty of collecting condition data covering all mechanical fault types in industrial scenarios, the fault diagnosis problem under incomplete data is receiving increasing attention where no target prior information can be available. The existing open-set or universal domain adaptation (DA) diagnosis methods typically treat private fault samples in the target as a generalized "unknown" fault class, neglecting their inherent structure. This oversight can lead to confusion in latent feature space representations and difficulties in separating unknown samples. Therefore, a universal DA method with unsupervised clustering is developed to explore the intrinsic structure of the target samples for mechanical fault diagnosis, where multi-source information on different working conditions is considered to transfer complementary knowledge. First, a composite clustering metric combining single-domain and cross-domain evaluation is constructed to recognize shared and unknown health classes on source-target domains. Second, to alleviate the intra-class shift while enlarging the inter-class gap, a class-wise DA algorithm is suggested which operates on the basis of maximum mean discrepancy. Finally, an entropy regularization criterion is utilized to facilitate clustering of different health classes. The efficacy of the presented approach in the fault diagnosis issues when monitoring data is inadequate has been verified through extensive experiments on three rotating machinery datasets.

Keywords: Fault diagnosis, domain adaptation (DA), rotating machinery, multi-source information, distribution difference.

1. Introduction

Prognostics health management of mechanical equipment has been widely concerned by the academia and industry because it is an essential part of industrial production such as aerospace, vehicle, and shipbuilding [1]. The rapid advancement of computer hardware and industrial sensors has propelled significant progress in fault diagnosis technology. Various approaches have been employed in this field, spanning from model-based and signal processing-based methods to data-driven ones [2-4]. Data-driven approaches have gained widespread acceptance in fault diagnosis due to their ability to automatically build mapping relationships between mechanical process data

* Corresponding author at: School of Vehicles and Energy, Yanshan University, Qinhuangdao, Hebei 066004, PR China.

E-mail address: hspace@ysu.edu.cn (D. Han).

and health states using intelligent models, without the need for complex physical modeling or specialized diagnostic expertise. [5,6].

During the early stages of the development of data driven-based fault diagnosis technology, it was assumed that a large amount of labeled mechanical process data would be available and that the operating conditions of the machinery would remain stable [7,8]. In this ideal fault diagnosis scenario, a variety of machine learning algorithms such as Naive Bayes, Decision Tree, Support vector machine and Artificial neural network are applied to equipment condition monitoring through offline training [9,10]. The introduction of deep learning models such as Convolutional Neural Network (CNN), Deep Auto-Encoder Network (DAN), and Deep Belief Network (DBN) has been particularly noteworthy. Their robust hierarchical feature extraction capabilities through multiple linear and nonlinear layers make them highly promising tools for data-driven fault diagnosis methods [11,12]. Zhao et al. [13] presented a deep CNN-based model for fault diagnosis of the planet bearings, which employed the synchrosqueezing transform as a signal processing algorithm. Shi et al. [14] designed an adaptive DBN with dynamic learning rate for fault diagnosis of the rotating bearings, in which the optimal structure of DBN was obtained through particle swarm optimization. By constructing a new loss function with maximum correntropy, Shao et al. [15] proposed a DAN-based interpretations learning approach for condition monitoring of electrical locomotive bearing and gearbox. The above-mentioned fault diagnosis issue has been considered to be satisfactorily solved at this time thanks to the availability of high-quality training data [16,17].

Regrettably, in real industrial production, the working environment and operating conditions of mechanical equipment often vary in response to different process requirements. This implies that the data used for training and testing are sourced from distinct distributions, a phenomenon referred to as domain shift [18]. Traditional supervised diagnosis models would drastically deteriorate when there is a domain shift between training data (source domain) and testing data (target domain) [19]. According to Lei et al. [20], the transfer learning theory will make it easier to conduct future research on intelligent fault diagnosis in engineering contexts by allowing users to transfer their knowledge from one or more tasks to related but novel ones. Current articles have extensively used domain adaptation (DA), one of the most well-liked transfer learning approaches, to address the domain shift issue in fault diagnosis. It tries to suppress the distributional differences of transferable features by learning domain-invariant representations from a common feature space with various distributions [21,22]. The DA methods based on difference metrics and adversarial training are extensively adopted. An ensemble deep transfer network for identify rolling bearing faults was presented by Li et al. [23], where the feature adaption layer is constructed through the maximum mean discrepancy (MMD) of multiple different kernels. Qian et al. [24] combined MMD and CORrelation alignment to construct a deep discriminative transfer learning model for cross-machine status monitoring. Chen et al. [25] presented a task-specific features learning network based on adversarial learning to handle large distribution discrepancy in cross-domain fault diagnosis. A residual joint adaptation intelligent transfer fault diagnosis framework with joint MMD and adversarial discriminator was constructed by Jiao et al. [26]. Moreover, because mechanical equipment is typically not permitted to operate with faults, collecting fault data is a rare occurrence. Consequently, some scholars have ventured to build diagnosis models using limited training data [27]. By embedding class-imbalance strategy into the adversarial training process, Kuang et al. [28] developed a transfer learning model with imbalanced data for cross-domain fault detection. Han et al. [29] performed individual DA on the target and source sparse data of the same equipment

conditions to alleviate the unclear data distribution caused by the lack of target data. Li et al. [30] developed a meta-learning few-shot condition monitoring approach with model-agnostic. Deep generative adversarial models are applied in [31] to create fake samples for cross-domain condition identification.

To compensate for the limited state knowledge in a single-source dataset, some scholars transfer diagnostic knowledge from multi-source data collected across various machine-operating scenarios to enhance the diagnostic accuracy of target tasks. Zhu et al. [32] introduced a multi-source domain adaptation approach for fault diagnosis of rotating machinery, which utilizes adversarial strategies to obtain domain-invariant feature representations. Yang et al. [33] proposed a multi-source DA fault diagnosis model, where the anchor adapters are constructed to transfer multi-source knowledge. Chai et al. [34] employed a pre-trained stacked AE to extract multi-source features from various operating conditions and utilized the domain discriminators to align each source with the target domain. Wang et al. [35] proposed a cycle generative adversarial network model to generate complete training data through multi-source mutual learning, and use based on anchor adapters to extract multi-source domain invariant features. Li et al. [36] devised a multi-source domain adaptation model for process fault diagnosis, incorporating feature-level and class-level techniques. This model employs a common feature extractor and a feature selection module to learn global and local features while mitigating negative transfer effects. Collecting labeled data for each specific source domain can be expensive and time-consuming. Multi-source domain adaptation can reduce the need for extensive labeled data in each domain, as it leverages information from related domains to improve the model's performance.

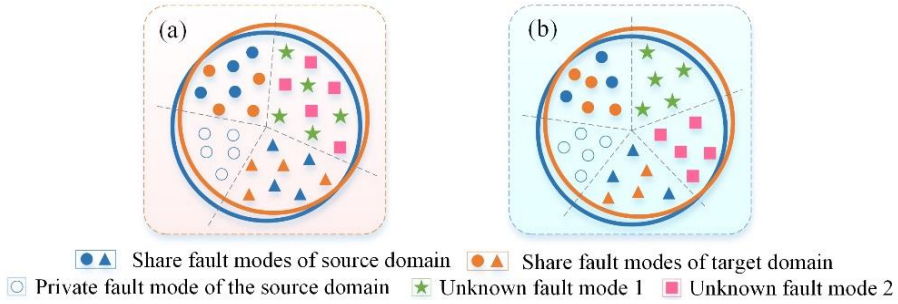


Fig. 1. (a) Previous universal DA method; (b) Proposed universal DA approach.

Despite the fact that the research cited above have made some progress in industrial diagnosis problem, they are always constrained by the consistent requirement of the source and target mechanical state spaces. Since a machine may only have one or two fault modes from the beginning of production to the operation at the end-of-life, very limited fault type data can be collected, so the diagnostic knowledge from the source machine is usually difficult to cover the target machine [37]. Because the label space of the source and target domains is not shared, traditional DA algorithms cannot perform effective distribution matching [38]. There have been some attempts to identify fault modes that are unknown and not visible in the training data. Applying an instance-level weighted mechanism to assist adversarial learning to alleviate negative transfer, Zhang et al. [39] developed an open-set deep learning machinery defect diagnosis solution. Zhu et al. [40] adopted multiple auxiliary classifiers to measure the similarity of each target sample, and applied domain adversarial model to measure domain knowledge to identify unknown and known target instances. A dual adversarial network is suggested by Zhao et al. [41] that uses an auxiliary domain discriminator to provide similarity weights to each target sample for open-set fault diagnosis. Furthermore, since the

collected target domain data has no state labels, no prior knowledge of the label space is available, thus a universal DA fault diagnosis challenge arises, in which private classes in both domains is allowed own [42]. A hybrid weighted deep adversarial learning model was developed in [43] to deal with the fault diagnosis issue of universal DA scenarios, where class-wise and instance-wise weighting strategy was adopted for source classes and target samples, respectively. A self-supervised DA model with neighborhood clustering technique is built by Li et al. [44] for rotating machine fault diagnosis, in which an entropy-based feature match technology was applied to isolate the unknown from known samples.

However, the existing literature generally concentrates on distinguishing shared state classes while treating private samples from target domain as a whole, i.e., unknown fault class, as shown in Fig. 1. In practice, multiple unknown fault modes may exist in the test scenario, which essentially belong to different semantic classes. Simply treating them as a generalization class results in lower representation compactness and blurred decision boundaries. The intrinsic structure of target unlabeled samples is worthy of further exploration, which facilitates the distinction of common and private state classes. Deep learning models can automatically learn fault-related features, and unsupervised clustering algorithms are a suitable tool for exploring unknown data structures. Motivated by these two points, with the complementary information of multi-source domain data for different operating conditions, a universal multi-source DA (UMDA) approach is presented to explore the intrinsic structure of the target domain for mechanical fault diagnosis in this study. The approach employs a convolutional auto-encoder (AE) network to learn multi-source shared features, and unsupervised clustering strategy is introduced to automatically group fault features to facilitate the capture of the intrinsic components of the target representations. The main contributions of this study are threefold:

(1) A universal multi-source DA deep learning framework is constructed for intelligent fault diagnosis of rotating machinery. This method focuses on distinguishing private target samples by exploring the essential structure of multiple unknown fault modes of the target machinery, rather than just treating them as one.

(2) Unsupervised clustering strategy is introduced to group target features, where consistent matching is used to search for shared condition categories in the source-target domains. Additionally, a composite clustering metric is proposed to determine the number of unknown fault classes.

(3) A Class-Wise DA algorithm is introduced to fine-grainedly transfer diagnosis knowledge from multiple sources to target domain, where the intra-class gap of the cross-domain is narrowed and the inter-class shift is enlarged.

The following sections make up the remainder of this paper. The problem setup and the basic theories are presented in Section 2. Section 3 introduces the developed universal multi-source DA fault diagnosis approach in detail. The experimental cases and analysis results are given in Section 4. The concludes are given in Section 5 finally.

2. Preliminary

2.1. Problem setup

In the fault diagnosis setting of universal multi-source DA, it is assumed that M source domains $D_{sj} = \{x_i^{sj}, y_i^{sj}\}_{i=1}^{n_j}$, $j = 1, 2, \cdots, M$ with n_j labeled samples sampled from P_{sj} and one target domain $D_t = \{x_i^t\}_{i=1}^{n_t}$ with n_t unlabeled samples sampled from P_t can be collected

under different mechanical working situations, where $P_{si} \neq P_{sj}\ (i \neq j)$ and $P_{sj} \neq P_t$. Let Y^{sj} denote health condition label set of source domain D_{sj}, and Y^t denotes health condition label set of target domain D_t. The unified label set of all source domains is represented by $Y^s = \cup_j Y^{sj}$, while the shared label set of all source and target domains is represented by $Y = Y^s \cap Y^t$. $\overline{Y^s} = Y^s \backslash Y^t$ represents the union of all sources private label set that is not observable in the target domain. $\overline{Y^t}$=$Y^t \backslash Y$ represents the private label set of target domain. Noted that not only domain shift exists but category gaps may also exist between any two domains. The commonness between $\cup_j D_{sj}$ and D_t may be calculated as $\xi = \frac{|Y|}{|Y^s \cup Y^t|}$[45]. This study seeks to build a general intelligent diagnosis model that can transfer the shared health state knowledge from multiple sources to the target domain while exploring potential unknown faults.

2.2. Convolutional AE

Similar to traditional AE, convolutional AE (CAE) has a basic encoder and decoder structure and follow an unsupervised training paradigm. The distinction is that the encoder of CAE uses conventional convolutional and pooling layers to extract discriminative features and decrease the dimensionality of the input, as illustrated in Fig. 2. In contrast to encoders, decoders aim to reconstruct the target data. Deconvolution and up-sampling layers are included in the decoder part to recover data dimensions [46]. The procedure for encoding and decoding may be formulated as,

$$h = f_{en}(x * W^1 + b^1) \quad (1)$$

$$\tilde{x} = f_{de}(h * W^2 + b^2) \quad (2)$$

where W^1 and b^1 represent the weight and bias of the encoder, respectively. W^2 and b^2 denote the weight and bias of the decoder, respectively. f_{en} and f_{de} denote the nonlinear activation functions in the CAE, respectively. x, h and $\tilde{x}$ indicate input data, the output features of the encoder and the rebuild data, respectively.

The CAE model is optimized by reconstruction loss, and the output map of the encoder can be considered as a typical feature of the raw data. These representative representations can be applied to data structure analysis and downstream pattern recognition tasks [47]. To facilitate the exploration of unlabeled target data structures in this study, CAE is employed as the mechanical condition feature extractor. Simultaneously, both unsupervised data analysis and supervised fault identification are performed during the training process.

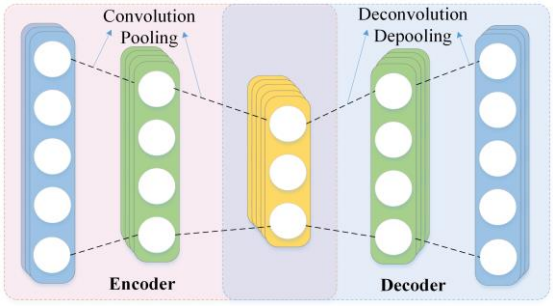


Fig. 2. The network structure of convolutional auto-encoder.

2.3. Composite Clustering Metric

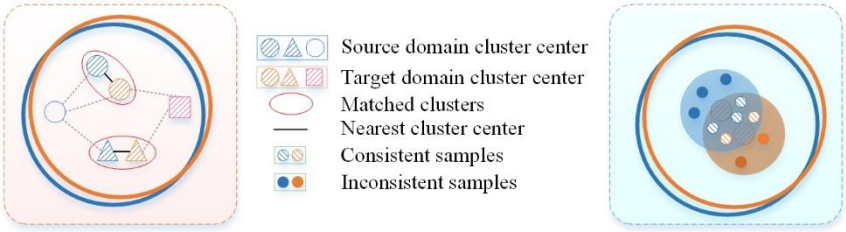
In the universal DA tasks, identifying the shared classes in source-target domains is a crucial step. The previous methods generally separate the shared and private classes through measuring the

transferability of each target instance based on domain similarity or uncertainty. However, neural networks may have high predictions for outliers, which makes quantitative indicators less discriminative due to the overconfidence of neural networks [48]. Therefore, to avoid unreliable measurements and exploit the latent structure of unlabeled target data, k-means unsupervised clustering algorithm was applied to match shared fault classes in this study.

For the target essential representations learned by the feature extractor, the k-means algorithm is performed for unsupervised clustering. For each source domain, clustering is performed based on its true states label. The cluster center of each cluster can be expressed as the mean of the unit embedding of the samples within that cluster.

$$v^c = \frac{1}{m_c} \sum_{x_i \in D_c^*} \frac{f(x_i)}{\|f(x_i)\|} \tag{3}$$

where D_c^* represents the cth clustering set of a certain source or target domain, and v^c is the cluster center of the cth clustering features. $f(x_i)$ represents the feature vector of sample x_i. The samples number of this clustering is represented by m_c.



(a) consistent matching (b) consistent sample selection

Fig. 3. The consistency of cluster matching.

In each pair of source-target domains, to identify shared classes, a logical approach is considered that the cluster centers of the same status classes should be closer to each other compared to different status classes or private classes. Consequently, two clusters with the closest cluster centers in different domains are chosen as a shared status pair. Conversely, clusters that do not have a matching counterpart are regarded as private status classes, as illustrated in Fig. 3(a). This matching approach is referred to as "consistent matching" in this study.

However, as a prerequisite for unsupervised clustering, the number of target clusters needs to be determined in advance. Although various clustering evaluation indicators have been presented to evaluate the number of clusters, they only assess the clustering results from one aspect [49]. Therefore, considering both single-domain scenarios and cross-domain knowledge, a composite clustering evaluation metric is proposed in this study by combining Silhouette coefficient and Normalized Mutual Information (NMI) [50]. Silhouette coefficient (SC) is a popular statistic for assessing the quality of clustering algorithms. It measures how similar an object is to its own cluster compared to other clusters. In other words, it assesses the compactness and separation between clusters in a clustering result. A higher SC value indicates an improved clustering result.

$$\begin{cases} \mathrm{SC} = \frac{1}{n} \sum_{i=1}^{n} s(i) \\ s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \end{cases} \tag{4}$$

where the average distance between a sample and all other samples in the same cluster is represented by $a(i)$. $b(i)$ represents the smallest average distance between a sample and all samples in a different cluster, where the sample does not belong. $s(i)$ denotes the silhouette coefficient of a single sample, and the global silhouette coefficient SC is the average value of the silhouette

coefficients of all samples. More information about the metric coefficients is available in [51].

However, SC is employed to assess how well single-domain clustering approaches perform and cannot judge the matching degree of cross-domain clustering. Therefore, a cross-domain consistency metric NMI is introduced to assess the consensus degree of the clustering between source and target domains. NMI evaluates the mutual information between the two clusters of a dataset in order to determine how similar two clusters of the dataset are to one another. Through the normalized average entropy of two clusters, NMI calculation formula is given by.

$$\begin{cases} \mathrm{NMI}(K,C) = \frac{2*I(K;C)}{H(K)+H(C)} \\ I(K;C) = H(K) - H(K|C) \end{cases} \tag{5}$$

where K and C denote the cluster of the source and target domain, respectively. $I(K;C)$ indicates the mutual information between two domains. $H(\cdot) = -\sum_{i=1}^{|Y|} p(i)\log p(i)$ denotes cross entropy in shared classes, in which $p(i) = \frac{m_i}{\sum_{i=1}^{|Y|} m_i}$ and m_i indicates the samples number of ith shared class in source or target domain. $H(Y|C) = -\sum_{c=1}^{|Y|} p(c)\sum_{k=1}^{|Y|} p(k|c)\log p(k|c)$, in which $p(k|c)$indicates the probability that the sample of the cth shared target class is closest to the center of the kth shared source class.

Specifically, after all matching pairs are determined by consensus matching, the NMI between matching pairs of source-target shared classes can be calculated. For a pair of matched source cluster $\{v_i^k\}_{i=1}^{m_k}$and target cluster $\{v_i^c\}_{i=1}^{m_c}$ whose clustering centers $\{v_k^s,\ v_c^t\}$ are closest to each other, the proportion of samples holding the cross-domain corresponding cluster label is counted. For example, for each sample v_i^c in the target domain cluster, its cosine similarity with all source domain cluster centers $\{v_1^s,\cdots,v_{|Y|}^s\}$ is calculated. If v_i^c most similar to its matching source domain cluster center v_k^s, it is recorded as a consistent sample, as depicted in Fig. 3(b). $p(k|c)$ can be formulated as,

$$p(k|c) = \frac{1}{m_c}\sum_{i=1}^{m_c} \mathbf{1}\{\arg\max_k \left(\frac{\langle v_i^c, v_k^s\rangle}{\|v_i^c\|\|v_k^s\|}\right) = k\} \tag{6}$$

where $\mathbf{1}$ is an indicator function to judge whether v_i^c is consistent with a matching cluster in source domain, so that $H(Y|C)$ and $\mathrm{NMI}(K,C)$ can be calculated.

Similarly, the proportion of consistent samples is counted for each sample v_i^k in the source cluster, and $H(C|Y)$ and $\mathrm{NMI}(C,K)$ can be obtained. The mean of both $\mathrm{NMI} = \frac{1}{2}(\mathrm{NMI}(K,C) + \mathrm{NMI}(C,K))$ is used as a measure of the matching of the source-target clusters.

In accordance with the analysis above, SC assesses how similar each sample is to other clusters, and NMI evaluates the consensus across clusters in the source and target domains. Combining the two clustering indexes from different perspectives, a composite clustering metric is proposed to comprehensively determine the number of target feature clusters,

$$\mathrm{w} = \frac{\mathrm{SC+NMI}}{2} \tag{7}$$

It is worth noting that to select the optimal number of target clusters c, this paper employs a method that the number of target clusters increases uniformly from 2 to $2|Y^s|$ to cluster the target features multiple times. The composite clustering metric value is recorded in each target clustering, and the optimal number of clusters c for this iteration is determined through the maximum

composite clustering metric. Also, to increase search efficiency, the search is terminated when the optimal cluster number c value remains constant for a certain number of epochs.

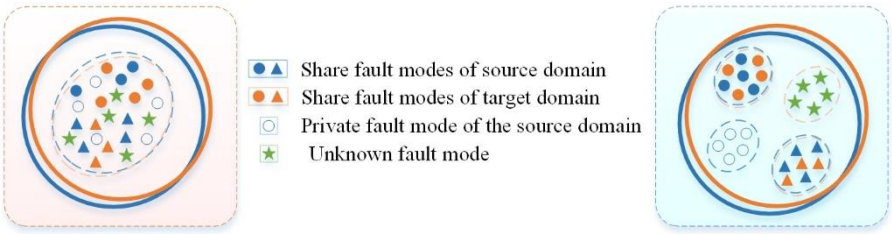
2.4. Class-Wise domain adaptation

MMD, a non-parametric function, is frequently used cross-domain fault diagnosis scenario since it can calculate the separation between any two distinct but related distributions. The MMD formula is defined as follows,

$$D_{\mathcal{H}}(P_s, P_t) \triangleq \left\| \mathbf{E}_{x\sim P_s}[\phi(x^s)] - \mathbf{E}_{x\sim P_t}[\phi(x^t)] \right\|_{\mathcal{H}}^2 \tag{8}$$

where $\mathcal{H}$ denotes the reproducing kernel Hilbert space (RKHS), and $\phi(\cdot)$ indicates the feature mapping function. x^s and x^t are the instances in P_s and P_t, respectively. $\mathbf{E}[\cdot]$ represents the expectation function. By constructing RHKS with characteristic kernels $\boldsymbol{K}$, the MMD between any two different distributions can be estimated as follows,

$$D_{\mathcal{H}}(P_s, P_t) = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\boldsymbol{K}(x^s, x^t) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{K}(x^s, x^t) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{K}(x^s, x^t) \tag{9}$$



(a) Global distribution matching (b) class-level distribution matching

Fig. 4. Schematic diagram of distribution matching.

While MMD has proven effective in DA, it is typically employed for aligning global marginal distributions, which may lead to mismatch when private classes are present, as shown in Fig. 4(a). According to the cluster matching relationship of the source-target domains, while considering the intra-class and inter-class distribution differences, an MMD-based Class-Wise DA algorithm is proposed to transfer status information at a fine-grained level, as shown in Fig. 4(b). This algorithm can be formulated as follows:

$$D_{cw}(P_s, P_t) = \frac{1}{|Y|}\sum_{k=1}^{|Y|} D_{\mathcal{H}}(P_s^k, P_t^k) - \frac{1}{|Y^s \cup Y^t|}\frac{1}{|Y^s \cup Y^t|-1}\sum_{k=1}^{|Y^s \cup Y^t|}\sum_{\substack{k'=1 \\ k'\neq k}}^{|Y^s \cup Y^t|} D_{\mathcal{H}}(P_s^k, P_t^{k'}) \tag{10}$$

Note that the target clusters that match the source domain clusters are given pseudo-labels with the source cluster labels. Furthermore, to minimize negative transfer in the early stage of the model, only consistent samples are selected for Class-Wise DA training in this study.

3. The proposed method

In industrial scenarios, the challenge of fault diagnosis is exacerbated by the difficulty in collecting comprehensive condition data covering all mechanical fault types. The paper proposed UMDA to solves this fault diagnosis problem in scenarios where fault type data is missing. The approach utilizes unsupervised learning to explore the internal structure of target features, thereby mitigating the issue of feature overlap among multiple unknown classes. It employs composite clustering metric to match shared state classes, accurately identifying multiple unknown fault classes. The network architecture and methodology of UMDA are presented in detail next.

3.1. *Network framework of UMDA*

The developed UMDA network architecture is divided into two parts, as illustrated in Fig. 5, the feature extractor and the health status classifier modules. It is trained using a combination of supervised and unsupervised learning.

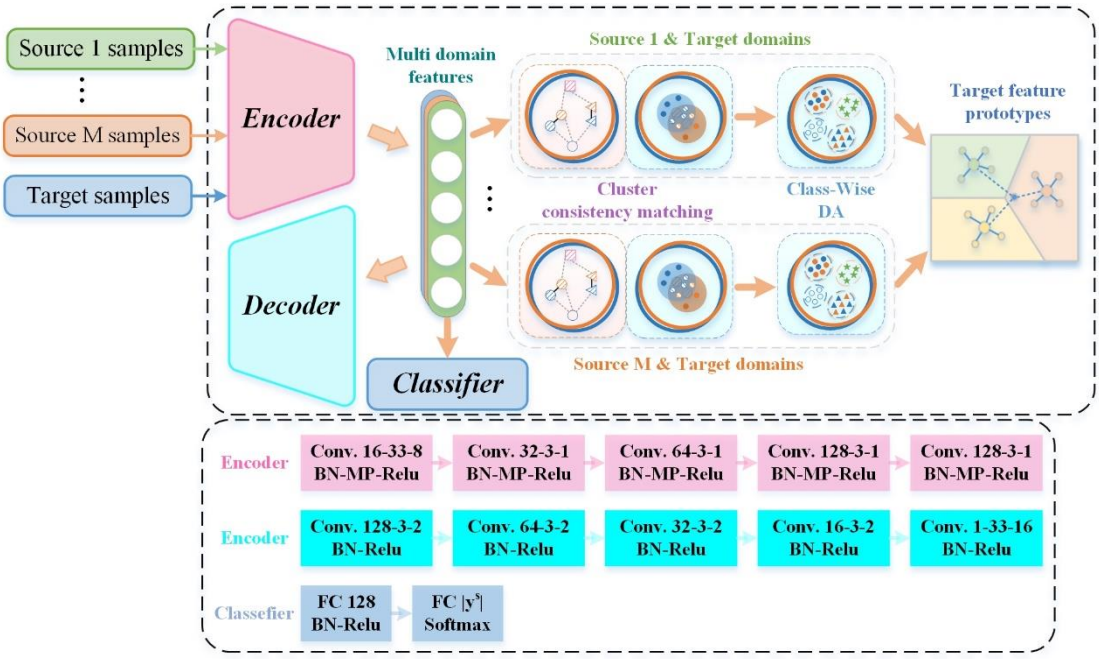


Fig. 5. The overall framework of the proposed UMDA.

The feature extractor is acted by a CAE to mine the intrinsic information of the input data. To extract identifiable features directly from the raw vibration status signal, the CAE is stacked with one-dimensional convolutional layers. It is noting that the first convolutional layer of the encoder uses a larger filter with a size of 32 to extract global features from a longer time series, and other convolutional layers use small filter with a size of 3 to extract local features. The number of filters in each layer of the encoder is 16, 32, 64, 128 and 128, respectively. Additionally, max-pooling operations are applied in each layer. In the decoder of CAE, the deconvolution and up-sampling layers corresponding to the encoder are applied to rebuild the raw data so that the extracted representations retain the essential information. Through unsupervised learning of CAE, the intrinsic spatial structure relationship of multiple source and target domain features can be effectively explored. Simultaneously, the classifier is used to link the representation space and the condition space to identify the mechanical health status of multiple source domains with finite state label spaces. It comprises of two fully connected layers, and the last layer is attached with a Softmax function to transform the status features to the fault probability distribution.

After obtaining representative features from the feature extractor, the consistent matching is performed to match shared health classes separately in each source-target domain pair. The unmatched states classes in the target domain are automatically treated as different unknown classes. The composite clustering metric is utilized to comprehensively evaluate the quality of clustering in the target domain from both single-domain and cross-domain perspectives. The objective is to determine the optimal number of clusters for the target features. Once the optimal number of clusters for the target features is determined, Class-Wise DA is conducted between the matched cluster pairs in multiple source-target domains. This process facilitates the precise transfer of diagnostic

knowledge.

The specific network structure parameters are also listed in Fig.5. Furthermore, rectified linear units are applied as activation functions in all layers, with the exception of the last layer of the decoder, which applies sigmoid as activation function. The batch normalization is used at all network layers to speed up model convergence.

3.2. Model optimization

In this study, the proposed UMDA uses CAE as the feature extractor, it is trained in an unsupervised manner, so samples from multiple source and target domains are rebuilt in the training phase. Through reconstruction learning of multi-domain samples, multi-domain representative features with essential structures can be learned. The reconstruction loss function is determined using the mean square error,

$$L_{ae} = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{n_j}\sum_{i=1}^{n_j}\left|x_i^{sj} - \tilde{x}_i^{sj}\right|^2 + \frac{1}{n_t}\sum_{i=1}^{n_t}\left|x_i^{t} - \tilde{x}_i^{t}\right|^2 \tag{11}$$

where $\tilde{x}_i^{sj}$ and $\tilde{x}_i^{t}$ represent the reconstructed data of x_i^{sj} and x_i^{t}, respectively.

At the same time, supervised training is essential for feature extractors to learn recognizable features. Therefore, the classifier receive multiple source domain representations output from feature extractors for training with corresponding state labels. Generally, the loss of supervised training is calculated by the cross entropy function.

$$L_{cls} = \frac{1}{M}\sum_{j=1}^{M}\left(-\frac{1}{n_j}\sum_{i=1}^{n_j}\sum_{k=1}^{|Y^s|}\mathbf{I}(y_i^{sj}=k)\log\tilde{y}_{i,k}^{sj}\right) \tag{12}$$

where $\tilde{y}_{i,k}^{sj}$ indicates that the classifier outputs the predicted probability that the sample belongs to the kth state class. The optimization on Equation 12 is comparable to aligning the distributions in the shared label sets of any two source domains and separating the private fault classes of each source domain.

During each training epoch, after the target features are optimally clustered according to the consistent matching and composite clustering metric, Class-Wise DA is then performed with consistent samples to facilitate more accurate cluster matching and diagnosis knowledge transfer. The consistent matching cluster pairs between multiple source-target is simultaneously optimized by the following Class-Wise DA loss,

$$L_{cw} = \frac{1}{M}\sum_{j=1}^{M} D_{cw}(X_{cons}^{sj}, X_{cons}^{t}) \tag{13}$$

where X_{cons}^{t} and X_{cons}^{sj} denote the consistent samples in target domain D_t and source domain D_{sj}, respectively. Since the distributions of shared health classes in any two source domains are aligned in Equation 12, the distributions of all shared classes in the source-target domains are well matched in Equation 13.

Furthermore, to further improve the discriminability of target clusters, an entropy regularization criterion is applied to encourage low-density cluster boundaries. Specifically, this criterion can be implemented by the following formula,

$$L_{ent} = -\frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{c=1}^{C} y_{i,c}^{t}\log\frac{\exp((v_i^t)^T\cdot v_c^t/\tau)}{\sum_{c=1}^{C}\exp((v_i^t)^T\cdot v_c^t/\tau)} \tag{14}$$

where y_i^t is the one-hot vector of the corresponding cluster label of the target sample. v_i^t is the feature vector of the target sample, v_c^t denotes the cth class feature cluster center stored in the last

epoch. τ is a temperature parameter, with an empirical value of 0.1.

The proposed UMDA training program is broken into two phases: pre-training and formal training. UMDA is optimized through the classification loss L_{cls} and the reconstruction loss L_{ae} in the pre-training stage. The pre-training process provides initialization network parameters for subsequent clustering matching and Class-Wise DA. In the formal training phase, Class-Wise DA loss and entropy regularization loss are also added to the training of the network. It is possible to define the total objective loss function of the suggested UMDA as follows:

$$L_{total} = L_{ae} + L_{cls} + \lambda L_{cw} + \gamma L_{ent} \tag{15}$$

where λ and γ are the trade-off parameters of the corresponding loss item. λ is set to 0.1 for all consistent samples. $\gamma = e^{-\omega \times \frac{i}{N}}$ is set through a ramp-up function, i and N represent current and global iteration numbers, respectively. ω is fixed at 3.

By minimizing the value of each of the aforementioned loss functions, the feature extractor and classifier parameters of the proposed UMDA model may be simultaneously optimized. The optimization issue can be stated as follows:

$$\begin{cases} \hat{\theta}_F = \arg\{\min_{\theta_F} L_{ae}(\theta_F), \min_{\theta_F} L_{cls}(\theta_F, \hat{\theta}_C), \min_{\theta_F} L_{cw}(\theta_F), \min_{\theta_F} L_{ent}(\theta_F)\} \\ \hat{\theta}_C = \arg\{\min_{\theta_C} L_{cls}(\hat{\theta}_F, \theta_C)\} \end{cases} \tag{16}$$

where θ_F and θ_C indicate the parameters of the feature extractor and classifier, respectively. $\hat{\theta}_F$ and $\hat{\theta}_C$ are the optimal values of θ_F and θ_C, respectively.

By applying the stochastic gradient descent algorithm, the proposed UMDA network parameters may be simply updated concurrently in each iteration.

$$\begin{cases} \theta_F \leftarrow \theta_F - \eta(\frac{\partial L_{ae}}{\partial \theta_F} + \frac{\partial L_{cls}}{\partial \theta_F} + \lambda \frac{\partial L_{cw}}{\partial \theta_F} + \gamma \frac{\partial L_{ent}}{\partial \theta_F}) \\ \theta_C \leftarrow \theta_C - \eta \frac{\partial L_{cls}}{\partial \theta_C} \end{cases} \tag{17}$$

where η is the learning rate of the network model, which decays with the factor of $(1 + \alpha \frac{i}{N})^{-\beta}$. α and β are set to 10 and 0.75, respectively. It is set at 0.001 for the initial learning rate η_0.

Algorithm 1: Proposed Methodology

Input: Labeled source domains $\{D_{sj}\}_{j=1}^{M}$; unlabeled target domain D_t; Tradeoff parameters λ and γ; Mini-batch size m.

1. Initialize UMDA model parameters θ_F and θ_C.
2. **for** each pre-training iteration **do**
3. Randomly draw m samples from each source and target domain.
4. Calculate the reconstruction and classification loss with Equation (11) and (12).
5. Update parameters of θ_F and θ_C with Equation (17).
6. **end for**
7. **for** each formal training iteration **do**
8. Randomly draw m samples from each source and target domain.
9. Compute the total loss with Equation (15).
10. Update parameters of θ_F and θ_C with Equation (17).
11. **end for**

Output: Optimized model parameters θ_F and θ_C.

In the last iteration of the training process, all target cluster centers $[v_1^t, v_2^t, \cdots, v_C^t]$ are stored

as a prototype bank. In the test scenario, only the feature extractor of UMDA is performed, and each target sample can be assigned a label with the nearest prototype. Consequently, not only known state classes that have appeared in multiple source domains can be identified, but also multiple unknown fault classes in target domain can be detected. Algorithm 1 presents the training process for the suggested UMDA.

4. Experimental study

A universal multi-source DA approach is presented in this study for the exploration of target unknown faults. To evaluate the efficiency of the suggested method, extensive experiments are implemented for various fault diagnosis tasks based on vibration data collected on three rotating machinery test benches. Fig. 6 depicts overall diagnostic procedure of the suggested method.

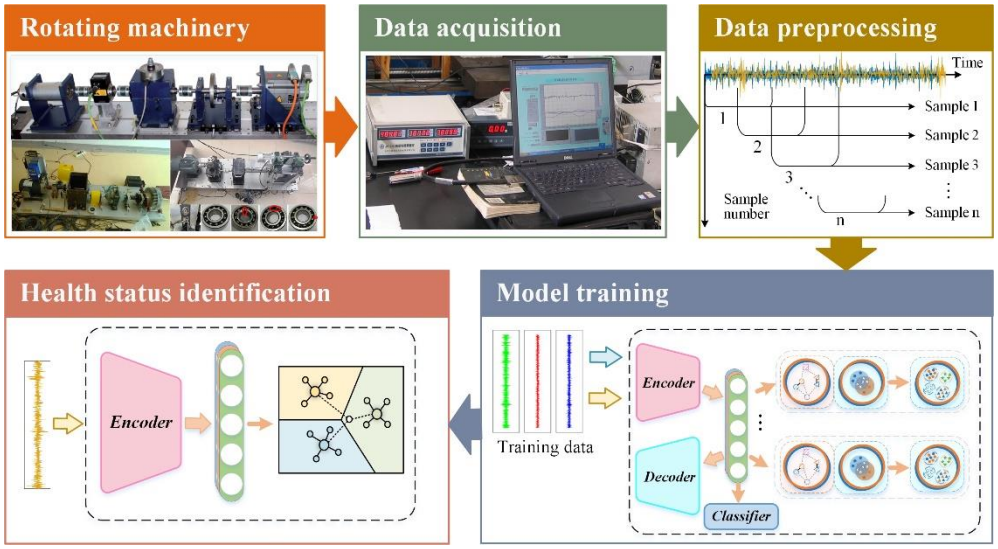


Fig. 6. The monitoring and diagnosis process of the proposed method.

4.1. Dataset descriptions

1) KAT bearing dataset. The KAT bearing dataset is a vibration dataset of 6023 type ball bearings publicly provided by the KAT data center in Paderborn University [52]. This test bench mainly includes five mechanical components: drive motor, load motor, test module, torque measurement shaft and flywheel, as shown in Fig. 7(a). A piezoelectric acceleration sensor with a sample frequency of 64kHz is mounted on the bearing housing to collect vibration signals. 32 bearing signals in different status were collected, including 6 normal states, 12 artificial damages, and 14 actual damages caused by accelerated life tests. By changing the torque of the load motor and the speed of the driving motor, the experimental scenarios under four different working conditions were simulated. To more restore the actual industrial scene, the data of one normal bearing and 9 damaged bearings of accelerated life test were selected for experiment in this study. The related information of the bearing status is shown in Table 1.

2) Reduction Gearbox (RG) dataset. The reduction gearbox dataset was obtained from a mechanical bench consisting of a planetary gearbox, a fixed-shaft gearbox and an AC variable frequency motor, and Fig. 7(b) shows the experimental setup. The data was sampled at 5120 Hz while the motor speed was held constant at 880 rpm. The control current of the magnetic powder brake is adjusted to 0.05 A, 0.1 A and 0.2 A in turn to apply various loads. Different fault modes were simulated by sequentially replacing the large or small gears in the fixed-shaft gearbox, while

the planetary gearbox remained in a healthy state. The health status of the fixed-shaft gearbox is listed in Table 2 and included single-point faults in the large gear, small gear, and compound faults of both.

3) Rolling Mill (RM) bearing dataset. As seen in Fig. 7(c), a direction-changing gearbox, a reduction gearbox, a drive motor and a four-high rolling mill are included in the rolling mill bench. The experiment simulated four different mill conditions, including three different faults in upper work roll bearing and a normal status, and three different bearing failures were introduced by artificial damage. An accelerometer with a sample frequency of 10240 Hz was positioned close to the upper work roll bearing to obtain condition data. The experiment was conducted at driving motor speeds of 1800 rpm, 1200 rpm, and 600 rpm. Table 2 provides details on the health condition of the RM datasets.

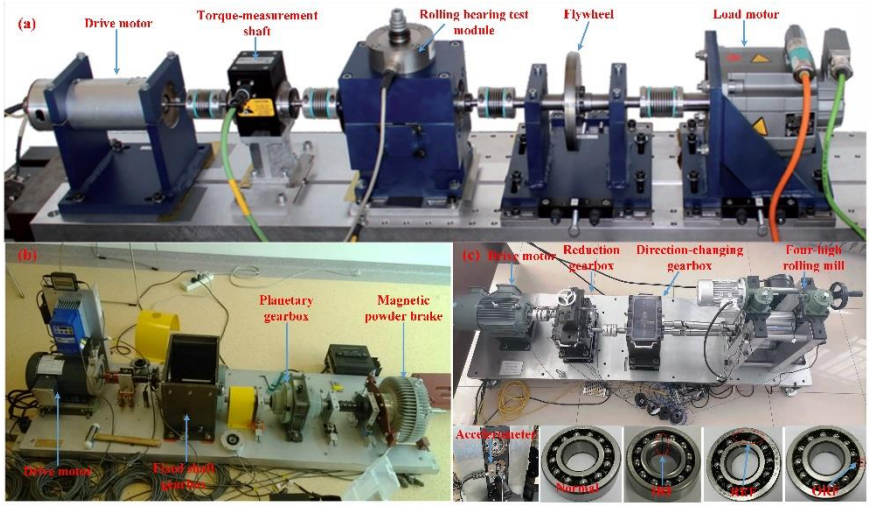


Fig. 7. (a) KAT bearing test rig, (b) Reduction Gearbox test bench, (c) Rolling Mill test rig.

Table 1

Description of the health conditions of the KAT dataset

Dataset	Label	Fault mode	Location	Combination	severity
KAT	0	Normal	/	/	/
	1	Plastic deform.: Indentations	OR	S	1
	2	Fatigue: Pitting	OR	R	2
	3	Fatigue: Pitting	OR	S	1
	4	Fatigue: Pitting	IR+OR	M	2
	5	Fatigue: Pitting	IR+OR	M	3
	6	Fatigue: Pitting	IR	M	1
	7	Fatigue: Pitting	IR	S	3
	8	Fatigue: Pitting	IR	S	2
	9	Fatigue: Pitting	IR	S	1
OR: Outer ring; IR: Inner ring; S: Single damage; R: Repetitive damage; M: Multiple damage					

4.2. Compared approaches

1) Basic: First, as a baseline for comparison, the standard deep learning model is introduced, and only conventional supervised learning techniques are employed within this framework.

2) MK-MMD [53]: The multi-kernel MMD (MK-MMD) is a common distance metric-based

closed-set DA approach, where MMD is applied in high-level interpretations to reduce cross-domain marginal distribution.

Table 2
Description of the health conditions of the RG and RM datasets

Dataset	Label	Fault mode	Fault Location	Dataset	Label	Fault Location
RG	0	Normal	/	RM	0	Normal
	1	Pitting	Large gear		1	Inner raceway
	2	tooth broken	Large gear		2	rolling element
	3	Pitting; wear	Large gear; Small gear		3	outer raceway
	4	tooth broken; wear	Large gear; Small gear			
	5	wear	Small gear			

Table 3
The working condition description of the source data

Dataset	Source name	Operating condition	Dataset	Source name	Operating condition
KAT	$\mathcal{K}_1$	Rotational Speed(rpm): 1500 Load Torque(Nm): 0.7 Radial force(N): 1000	RG	$\mathcal{G}_1$	Control current(A): 0.05
	$\mathcal{K}_2$	Rotational Speed(rpm): 900 Load Torque(Nm): 0.7 Radial force(N): 1000		$\mathcal{G}_2$	Control current(A): 0.1
	$\mathcal{K}_3$	Rotational Speed(rpm): 1500 Load Torque(Nm): 0.1 Radial force(N): 1000		$\mathcal{G}_3$	Control current(A): 0.2
	$\mathcal{K}_4$	Rotational Speed(rpm): 1500 Load Torque(Nm): 0.7 Radial force(N): 400			
RM	$\mathcal{M}_1$	Rotational Speed(rpm): 600			
	$\mathcal{M}_2$	Rotational Speed(rpm): 1200			
	$\mathcal{M}_3$	Rotational Speed(rpm): 1800			

3) SAN [54]: The selective adversarial network (SAN) is a partial DA approach that uses multiple domain discriminators to select outlier source classes to avoid negative transfer.

4) OSBP [55]: Open-Set DA by Backpropagation (OSBP) is a classical approach in addressing the open-set DA problem. It enables the generator to isolate unknown samples from known classes by a new adversarial learning technique.

5) UAN [56] Universal Adaptation Network (UAN) is a pioneering method in tackling the universal DA challenge, which assesses sample-level portability to identify shared label sets and private label sets to each domain.

It should be noted that for fair comparison, all methods adopt similar network structures and hyper-parameters to the proposed model, and source domain samples from multiple sources are integrated for model training.

4.3. Experimental setup

In this experimental setup, the vibration signal acquired on the three test benches in each operating state is considered as a source, as shown in Table 3. In each diagnosis task for the universal DA diagnosis issue, one source is randomly selected as the target domain, and the remaining

multiple sources are used as source domain. 300 samples from each health status signal were intercepted using a sliding window with a size 2048. The training set is composed of a random sample of 70% of the data from each source, while the testing set is composed of the remaining 30%. Before being fed into the network, all samples are normalized such that they follow the standardized normal distribution.

Table 4

Information of the diagnosis tasks on the KAT dataset

Tasks	Source class	Target class	Commonness	Problem
A_1	$\mathcal{K}_1$(0,1,3,4,5)&$\mathcal{K}_2$(0,2,5,6,8)&$\mathcal{K}_3$(0,4,7,8,9)	$\mathcal{K}_4$(All)	1.0	Closed-set DA
A_2	$\mathcal{K}_1$(0,1,2,4,5)&$\mathcal{K}_2$(0,3,5,6,7,8)&$\mathcal{K}_3$(0,4,7,8,9)	$\mathcal{K}_4$(0,1,2,3,4,6,7,8)	4/5	Partial DA
A_3	$\mathcal{K}_1$(0,1,2,4)&$\mathcal{K}_2$(0,5,6,7,8)&$\mathcal{K}_4$(0,4,5,7,8)	$\mathcal{K}_3$(All)	4/5	Open-set DA
A_4	$\mathcal{K}_1$(0,1,2,4)&$\mathcal{K}_2$(0,1,3,5)&$\mathcal{K}_4$(0,6,7,8)	$\mathcal{K}_3$(0,1,2,3,4,6,7,8,9)	4/5	Universal DA
A_5	$\mathcal{K}_1$(0,1,3,4,5)&$\mathcal{K}_3$(0,2,4,5,8)&$\mathcal{K}_4$(0,3,7,8)	$\mathcal{K}_2$(0,1,2,3,4,6,7,9)	3/5	Universal DA
A_6	$\mathcal{K}_1$(0,1,2,5,7)&$\mathcal{K}_3$(0,2,4,7,9)&$\mathcal{K}_4$(0,4,5,8)	$\mathcal{K}_2$(0,2,3,5,6,7,8,9)	3/5	Universal DA
A_7	$\mathcal{K}_2$(0,1,2,4,7)&$\mathcal{K}_3$(0,1,4,5,7)&$\mathcal{K}_4$(0,1,4,7,8)	$\mathcal{K}_1$(0,1,3,4,6,7,9)	2/5	Universal DA
A_8	$\mathcal{K}_2$(0,1,3,6,9)&$\mathcal{K}_3$(0,3,4,6,9)&$\mathcal{K}_4$(0,3,6,7,9)	$\mathcal{K}_1$(0,2,3,5,6,8,9)	2/5	Universal DA

Table 5

Information of the diagnosis tasks on the RG dataset

Tasks	Source class	Target class	Commonness	Problem
B_1	$\mathcal{G}_1$(0,1,2,4)&$\mathcal{G}_2$(0,3,5)	$\mathcal{G}_3$(All)	1.0	Closed-set DA
B_2	$\mathcal{G}_1$(0,2,4)&$\mathcal{G}_2$(0,1,3,5)	$\mathcal{G}_3$(0,2,4,5)	2/3	Partial DA
B_3	$\mathcal{G}_1$(0,3,5)&$\mathcal{G}_3$(0,4,5)	$\mathcal{G}_2$(All)	2/3	Open-set DA
B_4	$\mathcal{G}_1$(0,1,4)&$\mathcal{G}_3$(0,3,5)	$\mathcal{G}_2$(0,2,4,5)	1/2	Universal DA
B_5	$\mathcal{G}_2$(0,3,4)&$\mathcal{G}_3$(0,3,5)	$\mathcal{G}_1$(0,1,2,4,5)	1/2	Universal DA
B_6	$\mathcal{G}_2$(0,1,4)&$\mathcal{G}_3$(0,4,5)	$\mathcal{G}_1$(0,2,3,5)	1/3	Universal DA

Table 6

Information of the diagnosis tasks on the RM dataset

Tasks	Source class	Target class	Commonness	Problem
C_1	$\mathcal{M}_1$(0,1,3)&$\mathcal{M}_2$(0,2,3)	$\mathcal{M}_3$(All)	1.0	Closed-set DA
C_2	$\mathcal{M}_1$(0,1,2)&$\mathcal{M}_2$(0,2,3)	$\mathcal{M}_3$(0,1,2)	3/4	Partial DA
C_3	$\mathcal{M}_1$(0,1)&$\mathcal{M}_3$(0,1,2)	$\mathcal{M}_2$(All)	3/4	Open-set DA
C_4	$\mathcal{M}_1$(0,1)&$\mathcal{M}_3$(0,3)	$\mathcal{M}_2$(0,1,2)	1/2	Universal DA
C_5	$\mathcal{M}_2$(0,1)&$\mathcal{M}_3$(0,2)	$\mathcal{M}_1$(0,2,3)	1/2	Universal DA
C_6	$\mathcal{M}_2$(0,1)&$\mathcal{M}_3$(0,2)	$\mathcal{M}_1$(0,3)	1/4	Universal DA

Table 7

The relevant hyper-parameters of the proposed UMDA

Parameter	Value	Parameter	Value
Pre-training epochs	10	λ	0.1
Formal training epochs	20	γ	$e^{-\omega\times\frac{i}{N}}$
Batch size	32	η	$\eta_0(1+\alpha\frac{i}{N})^{-\beta}$

To comprehensively assess the capability of the developed model in various diagnosis tasks, the closed-set, partial, open-set and universal DA diagnosis tasks are designed based on three rotating machinery datasets, which are presented in Tables 4-6. The relevant hyper-parameters of the proposed UMDA are shown in Table 7, and they were established by the validation work using the KAT dataset. The stated diagnosis results were taken as the average of ten repeated trials to decrease randomness in this investigation.

4.4. Results display and analysis

This section displays and analyzes the diagnostic results of several methods on the KAT, RG, and RM datasets for a variety of diagnosis tasks with different degrees of commonness. The average diagnostic accuracy, H-score, confusion matrix and feature visualization were used to demonstrate the diagnosis performance. According to reference [57], the formula for calculating the H-score is defined as follows,

$$\text{H-score} = 2\frac{a_Y \cdot a_{\overline{Yt}}}{a_Y + a_{\overline{Yt}}} \tag{18}$$

where a_Y and $a_{\overline{Yt}}$ represent the diagnostic accuracy for known shared classes and target unknown private classes, respectively.

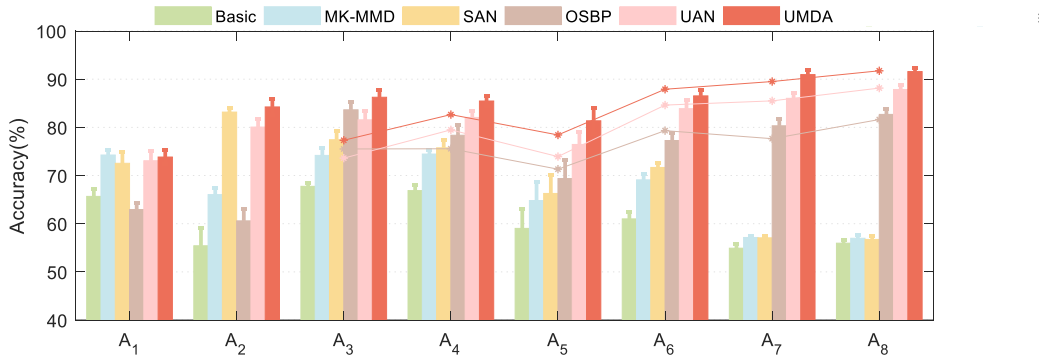


Fig. 8. Diagnostic accuracy and H-score on the KAT dataset with different methods (The broken lines represents the H-score.)

Table 8
Average diagnostic accuracy and H-score of different methods on the KAT dataset

methods	Tasks	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	Average
Basic	Accuracy	65.67	55.42	67.76	66.87	59.03	61.00	54.92	55.96	60.83
MK-MMD	Accuracy	**74.26**	66.04	74.18	74.46	64.81	69.10	57.14	56.98	67.12
SAN	Accuracy	72.54	83.17	77.43	75.75	66.28	71.68	57.07	56.74	70.08
OSBP	Accuracy	62.94	60.59	83.64	78.36	69.37	77.27	80.32	82.69	74.40
	H-score	/	/	75.51	75.56	71.32	79.32	77.65	81.62	76.83
UAN	Accuracy	73.07	80.06	81.56	82.06	76.45	83.88	86.03	87.86	81.37
	H-score	/	/	73.61	79.48	73.93	84.62	85.50	88.15	80.88
UMDA	Accuracy	73.84	**84.24**	**86.25**	**85.47**	**81.36**	**86.53**	**90.95**	**91.58**	85.03
	H-score	/	/	**77.32**	**82.65**	**78.42**	**87.91**	**89.51**	**91.74**	84.59

4.4.1 Diagnosis results discussion

The average accuracy and H-score in various diagnosis tasks on KAT bearing dataset are reported in Fig. 8 and the detailed are displayed in Table 8. The results demonstrate that the presented approach exhibits superior diagnosis performance compared to other methods across

different diagnostic tasks with an average test accuracy of 85.03%. The Basic method, which solely relies on the traditional supervised learning paradigm, exhibits low diagnostic accuracy across all tasks due to disparities in the distribution of multi-source features. The incorporation of distributional distance metrics has led to a slight enhancement of MK-MMD over the Basic method in several diagnostic tasks. Nonetheless, MK-MMD is unable to handle non-closed set fault diagnosis tasks (A_2-A_8) due to the absence of a private class handling mechanism. The SAN and OSBP methods are tailor-made to address partial set and open set DA problems, respectively. They achieved impressive diagnostic performance in task A_2 and A_3, respectively. However, when it comes to universal set diagnostic tasks, SAN is unable to distinguish unknown classes, and OSBP struggles with handling private classes from the source domain. UAN is a typical approach to address the universal DA problem and demonstrated competitive diagnostic performance in different tasks. However, UAN may generate high-confidence predictions for target outliers under source supervision. Additionally, it treats all private fault samples as a single "unknown" fault category, overlooking the internal structure of private fault features. Therefore, there is a need to further improve the diagnostic performance of UAN. The proposed UMDA method utilizes a cluster matching strategy for identifying shared classes. This approach effectively leverages the inherent feature structure of the target samples, leading to increased reliability in comparison to the similarity weight measurement strategy. Compared with UAN, its average diagnosis accuracy has increased by 3.66%, and its H-score has increased by 3.71% on the KAT dataset task. This result confirms the effectiveness and superiority of the proposed UMDA in universal fault diagnosis tasks.

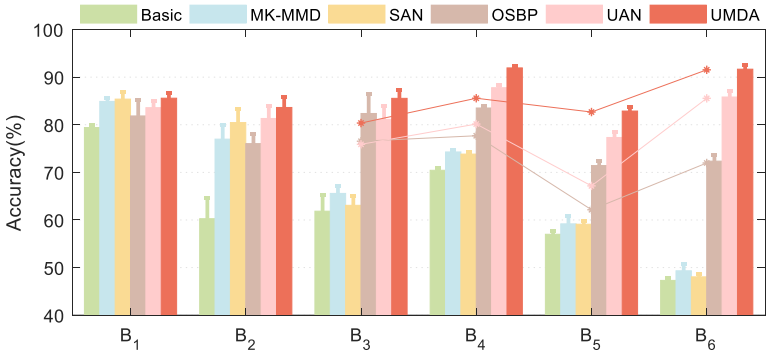


Fig. 9. Diagnostic accuracy and H-score on the RG dataset with different methods (The broken lines represents the H-score.)

Table 9
Average diagnostic accuracy and H-score of different methods on the RG dataset

methods	Task	B_1	B_2	B_3	B_4	B_5	B_6	Average
Basic	Accuracy	79.41	60.27	61.85	70.44	57.00	47.31	62.71
MK-MMD	Accuracy	84.90	76.98	65.58	74.29	59.19	49.33	68.38
SAN	Accuracy	85.39	80.46	63.08	73.84	59.11	48.06	68.32
OSBP	Accuracy	81.85	76.04	82.37	83.52	71.43	72.36	77.93
	H-score	/	/	76.54	77.69	62.15	71.97	72.09
UAN	Accuracy	83.59	81.32	81.13	87.78	77.33	85.83	82.83
	H-score	/	/	75.89	80.15	67.16	85.55	77.19
UMDA	Accuracy	**85.58**	**83.63**	**85.55**	**91.94**	**82.88**	**91.67**	**86.88**
	H-score	/	/	**80.36**	**85.57**	**82.67**	**91.54**	**85.04**

The results of various diagnostic tasks on the RG dataset is presented in Fig. 9, while Table 9

provides more specific information on diagnostic accuracy and H-score. Notably, UMDA exhibits superior diagnostic ability compared to other methods and its performance is comparable to that observed in the CWRU dataset. The proposed method demonstrates superior diagnostic accuracy in both B_2 and B_3 tasks, even surpassing the performance of SAN and OSBP, which were specifically designed for these tasks. Furthermore, as the number of private classes increases, the advantage of UMDA in diagnostic performance becomes more apparent. Compared to UAN, the average diagnostic accuracy of UMDA across multiple tasks increases by 4.05%, while the H-score shows an improvement of 7.85%. The results of the diagnosis task for different methods on the RG dataset are displayed in Fig. 10 and Table 10. The comparison reveals that the diagnosis performance of the proposed UMMDA method is superior to that of several other methods across various universal diagnosis tasks. In the diagnosis task C_1 and C_2, all methods exhibit satisfactory diagnostic accuracy. This can be attributed to the presence of diagnostic knowledge from the source domain that can cover the target domain. This also implies that the distribution of features across different sources in this dataset is relatively minimal differences. The Basic, MK-MMD and SAN methods without target private class detection mechanism perform poorly on diagnostic tasks C_3-C_5. The inability of OSBP to handle source-domain private classes results in suboptimal diagnostic performance on diagnostic tasks C_4-C_6. The proposed UMDA achieves the average diagnostic accuracy of 93.20% and an H-score of 89.42% on the RM dataset, surpassing the most competitive UAN by 2.75% and 5.39%, respectively. The efficacy and superiority of the proposed method has been established through its successful application to universal DA diagnosis problems.

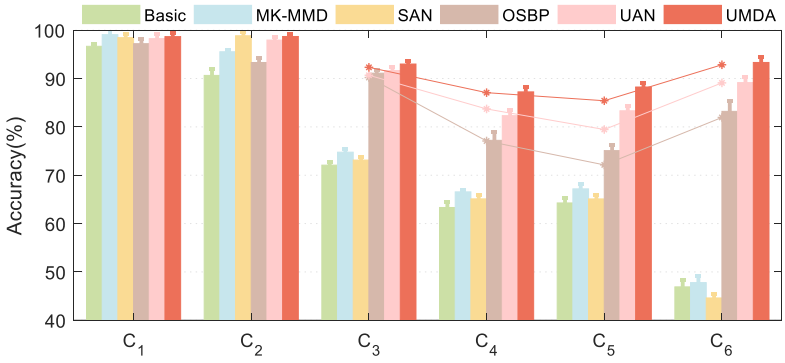


Fig. 10. Diagnostic accuracy and H-score on the RM dataset with different methods (The broken lines represents the H-score.)

Table 10
Average diagnostic accuracy and H-score of different methods on the RM dataset

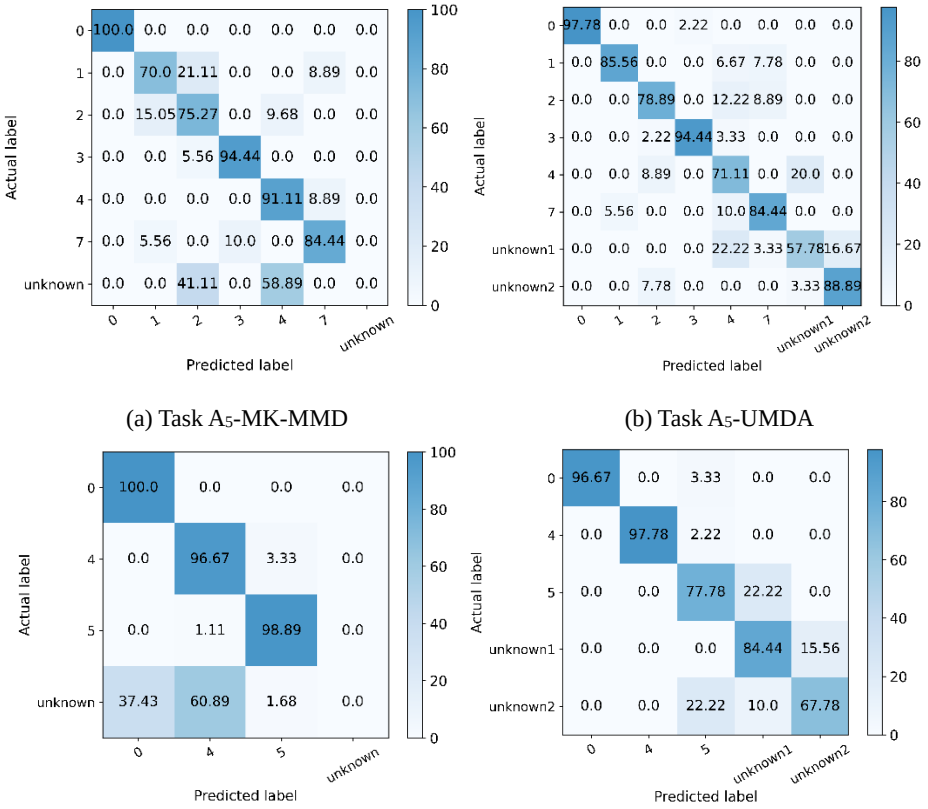
methods	Task	C_1	C_2	C_3	C_4	C_5	C_6	Average
Basic	Accuracy	96.67	90.65	72.09	63.33	64.27	46.91	72.32
MK-MMD	Accuracy	**99.08**	95.54	74.77	66.56	67.18	47.79	75.15
SAN	Accuracy	98.46	**98.85**	73.14	65.11	65.10	44.62	74.21
OSBP	Accuracy	97.21	93.32	91.11	77.22	75.09	83.21	86.19
	H-score	/	/	90.21	77.08	72.14	81.95	80.35
UAN	Accuracy	98.25	97.96	91.65	82.32	83.33	89.16	90.45
	H-score	/	/	90.68	83.71	79.46	89.10	85.74
UMDA	Accuracy	98.67	98.70	**93.00**	**87.24**	**88.25**	**93.33**	**93.20**
	H-score	/	/	**92.32**	**87.08**	**85.41**	**92.85**	**89.42**

4.4.2 Visualizations analysis and performance investigations

In this section, multiple visualization results from UMDA and the typical DA method MK-

MMD on diagnostic tasks A_5 and B_5 are presented. These visualizations offer deeper insights into the diagnostic capabilities of the proposed approach.

First, the confusion matrices for the two diagnostic approaches are presented in detail in Fig. 11. It is clear that the MK-MMD struggles with identifying unknown fault modes, as illustrated in Fig. 11(a) and Fig. 11(c), it consistently misclassifies unknown fault instances into other fault classes. The proposed UMDA utilizes an unsupervised clustering strategy to match shared classes and clusters private samples based on their internal structures. The effectiveness of UMDA is evident from Fig. 11(b) and Fig. 11(d), where it successfully identifies shared target samples and clusters private samples of unknown targets, thereby enhancing the accuracy of fault diagnosis.

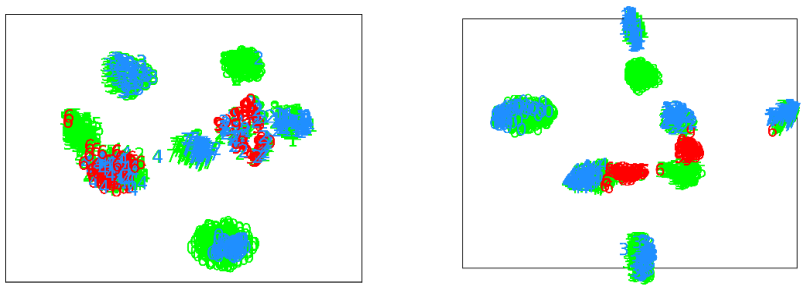


(a) Task A_5-MK-MMD (b) Task A_5-UMDA

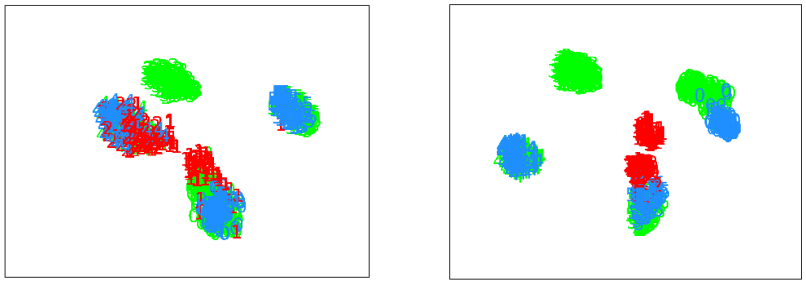
(c) Task B_5-MK-MMD (d) Task B_5-UMDA

Fig. 11. Confusion matrix of the MK-MMD and the proposed approach in the task A_5 and B_5.

Afterwards, the t-SNE technique [58] is utilized to generate a visualization of the high-level representations acquired by the feature extractor. The corresponding feature cluster visualizations are displayed in Fig. 12. Obviously, MK-MMD cannot detect private class features, and the negative transfer of fault knowledge exists in this model, which diminishing the diagnostic performance of the model. The proposed UMDA performs unsupervised clustering of target features, and uses the shared matching strategy to effectively identify shared classes and private classes. As shown in Fig. 12(b) and Fig. 12(d), different categories of fault features can be effectively separated by UMDA.



(a) Task A_5-MK-MMD (b) Task A_5-UMDA



(c) Task B_5-MK-MMD (d) Task B_5-UMDA

Fig. 12. Feature visualization in tasks A_5 and B_5 by MK-MMD and the proposed method. The green values denote source instances. The blue and red values mean target shared-class instances and target outliers, respectively.

Furthermore, to showcase the efficacy of the proposed approach in identifying common categories, the number of matched shared classes $|Y|$ and the composite clustering metric W are observed in diagnosis tasks A_5 and B_5. As shown in Fig. 13, the number of matched shared classes tends to be smaller at the beginning of training, and the value of the composite clustering metric is also smaller. However, as the training progresses, the number of matched shared classes gradually tends to the correct number of shared classes, and the value of the composite clustering metric gradually increases. Eventually, these two metrics stabilize near the best quality.

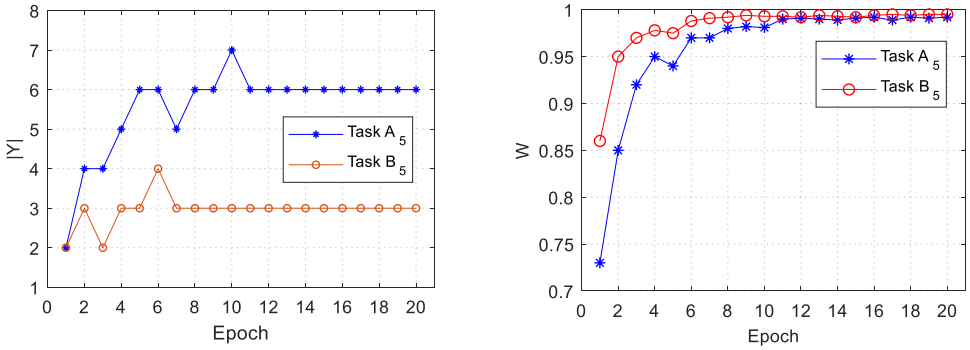


Fig. 13. The evolution of the number of matching shared classes $|Y|$ and the composite clustering metric W as training progresses.

4.4.3 Ablation Studies

In this section, each loss item of the proposed method is used individually in model training to evaluate the effectiveness of both Class-Wise DA and entropy regularization criterion. As shown in Fig. 14 and Table 11, in multiple universal DA fault diagnosis tasks, using both loss terms together achieve the highest diagnosis accuracy. In contrast, employing either L_{cw} or L_{ent} in isolation leads to suboptimal diagnostic accuracy, and applying L_{cw} alone leads to better diagnostic performance than L_{ent} alone.

The primary goal of Class-Wise DA is to mitigate the distribution gap between shared classes in the source-target domains, facilitating the more effective knowledge transfer from the source domain training to the target domain. This contributes to enhancing the model's generalization capability on the target domain. Entropy regularization introduces the concept of information entropy, encouraging the model to provide smoother and more consistent predictions for each sample. The effectiveness of entropy regularization is contingent upon the uncertainty in the label distribution within the target domain. In this study, the distribution difference between domains is the main challenge, and Class-Wise DA can significantly improve model performance when there is a significant difference between source-target domains. The label distribution of the target domain is relatively balanced, and the impact of entropy regularization is relatively small. Therefore, applying L_{cw} alone has higher diagnostic accuracy than L_{ent}. The combined use of the two can play their complementary roles to achieve the best diagnostic performance.

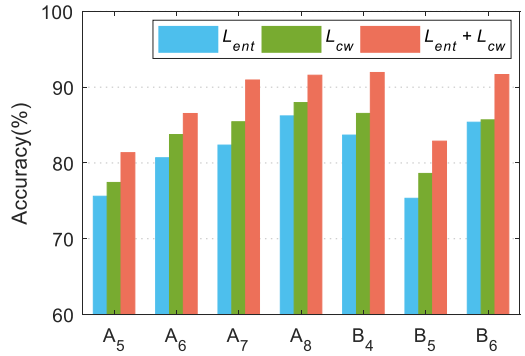


Fig. 14. Diagnostic accuracy for different loss terms of the proposed UMDA.

Table 11 Diagnostic accuracy of different loss terms.

Task	A_5	A_6	A_7	A_8	B_4	B_5	B_6
L_{ent}	75.61	80.69	82.36	86.21	83.68	75.34	85.38
L_{cw}	77.43	83.75	85.44	87.97	86.54	78.62	85.69
$L_{cw} + L_{ent}$	**81.36**	**86.53**	**90.95**	**91.58**	**91.94**	**82.88**	**91.67**

4.4.4 Hyper-parameter sensitivity analysis

This section separately explores the impact of the two trade-off parameters λ and γ used in this study on the performance of the model. As depicted in Fig. 15, the diagnostic accuracy curves for various λ values in tasks A_5 and B_5 exhibit a convex shape, with the highest accuracy achieved when $\lambda = 0.1$. Furthermore, the accuracy reaches 79% in a relatively wide range, which showed good model robustness. Also, the stepwise increasing γ used in this study was compared with various constant γ values on diagnostic tasks A_5 and B_5. The best diagnostic results are obtained when $\gamma = e^{-\omega \times \frac{i}{N}}$,which verifies the effectiveness of the progressively increasing design.

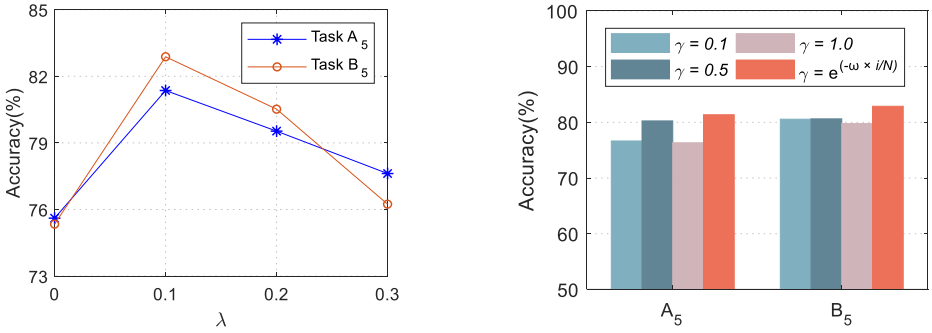


Fig. 15. Effect of different parameters on model performance.

5. Conclusion

In this study, a universal multi-source DA approach based on unsupervised clustering is developed for rotating machinery fault diagnosis when no target prior information is available. First, the unsupervised clustering algorithm is applied to capture the intrinsic structure of the target features. Then, a single-domain and cross-domain composite clustering metric is proposed to identify shared and unknown health classes in the source-target domain. Second, a class-wise DA algorithm is introduced to mitigate the intra-class shift, while simultaneously increasing the inter-class gap. Lastly, an entropy regularization criterion is incorporated to facilitate clustering of various health classes. The study conducted extensive experiments on three datasets from rotating machinery benches for comparison and analysis. The experimental results indicated that this approach can precisely identify shared target samples and efficiently detect private target samples, and surpasses the other DA methods on various universal set fault diagnosis tasks.

While the presented method has shown promising results in universal set fault diagnosis tasks, current study mainly focuses on diagnosing faults under various working conditions of a single equipment. In real-world engineering applications, data may be collected from various devices of the same type or different types. This can result in significant distribution shift among multi-source data and increasing the difficulty of diagnosis. Therefore, it is essential to study and develop fault diagnosis techniques for scenarios with large distribution difference among multi-source data.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgments

The studies were funded by the National Natural Science Foundation of China (Grant number 61973262), Natural Science Foundation of Hebei Province, China (Grant number E2020203147), the central government guides local science and technology development fund project (Grant numbers 216Z2102G and 216Z4301G), the scholarship from the China Scholarship Council (Grant

number CSC202208130095) and the Horizon Marie Skłodowska-Curie Actions program (101073037).

References

[1] Nuñez D L, Borsato M. OntoProg: An ontology-based model for implementing Prognostics Health Management in mechanical machines[J]. Advanced Engineering Informatics, 2018, 38: 746-759.

[2] Jia F, Lei Y, Lin J, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data[J]. Mechanical systems and signal processing, 2016, 72: 303-315.

[3] Tian J, Han D, Karimi H R, et al. Deep learning-based open set multi-source domain adaptation with complementary transferability metric for mechanical fault diagnosis. Neural Networks, 2023, 162: 69-82.

[4] Li T, Sbarufatti C, Cadini F, et al. Particle filter- based hybrid damage prognosis considering measurement bias[J]. Structural Control and Health Monitoring, 2022, 29(4): e2914.

[5] Yang D, Karimi H R, Pawelczyk M. A new intelligent fault diagnosis framework for rotating machinery based on deep transfer reinforcement learning. Control Engineering Practice, 2023, 134: 105475.

[6] Yang D, Karimi H R, Sun K. Residual wide-kernel deep convolutional auto-encoder for intelligent rotating machinery fault diagnosis with limited samples[J]. Neural Networks, 2021, 141: 133-144.

[7] Chen Z, Gryllias K, Li W. Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network[J]. IEEE Transactions on Industrial Informatics, 2019, 16(1): 339-349.

[8] Lei Y, Jia F, Lin J, et al. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5): 3137-3147.

[9] Xu X, Cao D, Zhou Y, et al. Application of neural network algorithm in fault diagnosis of mechanical intelligence[J]. Mechanical Systems and Signal Processing, 2020, 141: 106625.

[10] Manikandan S, Duraivelu K. Fault diagnosis of various rotating equipment using machine learning approaches–A review[J]. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 2021, 235(2): 629-642.

[11] Tian J, Han D, Li M, et al. A multi-source information transfer learning method with subdomain adaptation for cross-domain fault diagnosis[J]. Knowledge-Based Systems, 2022, 243: 108466.

[12] Zhang Y, Han D, Tian J, et al. Domain adaptation meta-learning network with discard-supplement module for few-shot cross-domain rotating machinery fault diagnosis[J]. Knowledge-Based Systems, 2023, 268: 110484.

[13] Zhao D, Wang T, Chu F. Deep convolutional neural network based planet bearing fault classification[J]. Computers in Industry, 2019, 107: 59-66.

[14] Shi P, Xue P, Liu A, et al. A novel rotating machinery fault diagnosis method based on adaptive deep belief network structure and dynamic learning rate under variable working conditions[J]. IEEE Access, 2021, 9: 44569-44579.

[15] Shao H, Jiang H, Zhao H, et al. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis[J]. Mechanical Systems and Signal Processing, 2017, 95: 187-204.

[16] Li X, Zhang W, Ma H, et al. Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks[J]. Neural Networks, 2020, 129: 313-322.

[17] Li W, Chen Z, He G. A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery[J]. IEEE Transactions on Industrial Informatics, 2020, 17(3): 1753-1762.

[18] Li C, Zhang S, Qin Y, et al. A systematic review of deep transfer learning for machinery fault diagnosis[J]. Neurocomputing, 2020, 407: 121-135.

[19] Wang Q, Michau G, Fink O. Domain adaptive transfer learning for fault diagnosis[C]//2019 Prognostics and

System Health Management Conference (PHM-Paris). IEEE, 2019: 279-285.

[20] Lei Y, Yang B, Jiang X, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap[J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.

[21] Sharma A K, Verma N K. Quick learning mechanism with cross-domain adaptation for intelligent fault diagnosis[J]. IEEE Transactions on Artificial Intelligence, 2021, 3(3): 381-390.

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

[23] Li X, Jiang H, Wang R, et al. Rolling bearing fault diagnosis using optimal ensemble deep transfer network[J]. Knowledge-Based Systems, 2021, 213: 106695.

[24] Qian Q, Qin Y, Luo J, et al. Deep discriminative transfer learning network for cross-machine fault diagnosis[J]. Mechanical Systems and Signal Processing, 2023, 186: 109884.

[25] Chen Z, He G, Li J, et al. Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(11): 8702-8712.

[26] Jiao J, Zhao M, Lin J, et al. Residual joint adaptation adversarial network for intelligent transfer fault diagnosis[J]. Mechanical Systems and Signal Processing, 2020, 145: 106962.

[27] Zhang T, Chen J, Li F, et al. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions[J]. ISA transactions, 2022, 119: 152-171.

[28] Kuang J, Xu G, Tao T, et al. Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 71: 1-11.

[29] Han T, Liu C, Wu R, et al. Deep transfer learning with limited data for machinery fault diagnosis[J]. Applied Soft Computing, 2021, 103: 107150.

[30] Li C, Li S, Zhang A, et al. Meta-learning for few-shot bearing fault diagnosis under complex working conditions[J]. Neurocomputing, 2021, 439: 197-211.

[31] Li X, Zhang W, Ding Q. Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks[J]. IEEE Transactions on Industrial Electronics, 2018, 66(7): 5525-5534.

[32] Zhu J, Chen N, Shen C. A new multiple source domain adaptation fault diagnosis method between different rotating machines[J]. IEEE Transactions on Industrial Informatics, 2020, 17(7): 4788-4797.

[33] Yang S, Kong X, Wang Q, et al. A multi-source ensemble domain adaptation method for rotary machine fault diagnosis[J]. Measurement, 2021, 186: 110213.

[34] Chai Z, Zhao C. Deep transfer learning based multisource adaptation fault diagnosis network for industrial processes[J]. IFAC-PapersOnLine, 2021, 54(3): 49-54.

[35] Wang Q, Xu Y, Yang S, et al. A domain adaptation method for bearing fault diagnosis using multiple incomplete source data[J]. Journal of Intelligent Manufacturing, 2023: 1-15.

[36] Li S, Yu J. A Multisource Domain Adaptation Network for Process Fault Diagnosis Under Different Working Conditions[J]. IEEE Transactions on Industrial Electronics, 2022, 70(6): 6272-6283.

[37] Yang B, Xu S, Lei Y, et al. Multi-source transfer learning network to complement knowledge for intelligent diagnosis of machines with unseen faults[J]. Mechanical Systems and Signal Processing, 2022, 162: 108095.

[38] Yu X, Zhao Z, Zhang X, et al. Deep-learning-based open set fault diagnosis by extreme value theory[J]. IEEE Transactions on Industrial Informatics, 2021, 18(1): 185-196.

[39] Zhang W, Li X, Ma H, et al. Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning[J]. IEEE Transactions on Industrial Informatics, 2021, 17(11): 7445-7455.

[40] Zhu J, Huang C G, Shen C, et al. Cross-Domain Open-Set Machinery Fault Diagnosis Based on Adversarial Network With Multiple Auxiliary Classifiers[J]. IEEE Transactions on Industrial Informatics, 2021, 18(11): 8077-8086.

[41] Zhao C, Shen W. Dual adversarial network for cross-domain open set fault diagnosis[J]. Reliability Engineering & System Safety, 2022, 221: 108358.

[42] Yan Z, Liu G, Wang J, et al. A new universal domain adaptive method for diagnosing unknown bearing faults[J]. Entropy, 2021, 23(8): 1052.

[43] Zhang W, Li X, Ma H, et al. Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 7957-7967.

[44] Li J, Huang R, Chen J, et al. Deep Self-Supervised Domain Adaptation Network for Fault Diagnosis of Rotating Machine With Unlabeled Data[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-9.

[45] You K, Long M, Cao Z, et al. Universal domain adaptation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2720-2729.

[46] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]//Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer Berlin Heidelberg, 2011: 52-59.

[47] Qian Q, Qin Y, Wang Y, et al. A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis[J]. Measurement, 2021, 178: 109352.

[48] Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 427-436.

[49] Maulik U, Bandyopadhyay S. Performance evaluation of some clustering algorithms and validity indices[J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(12): 1650-1654.

[50] McDaid A F, Greene D, Hurley N. Normalized mutual information to evaluate overlapping community finding algorithms[J]. arXiv preprint arXiv:1110.2515, 2011.

[51] Palacio-Niño J O, Berzal F. Evaluation metrics for unsupervised learning algorithms[J]. arXiv preprint arXiv:1905.05667, 2019.

[52] Lessmeier C, Kimotho J K, Zimmer D, et al. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification[C]//PHM Society European Conference. 2016, 3(1).

[53] Li X, Zhang W, Ding Q, et al. Multi-layer domain adaptation method for rolling bearing fault diagnosis[J]. Signal processing, 2019, 157: 180-197.

[54] Cao Z, Long M, Wang J, et al. Partial transfer learning with selective adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2724-2732.

[55] Saito K, Yamamoto S, Ushiku Y, et al. Open set domain adaptation by backpropagation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 153-168.

[56] You K, Long M, Cao Z, et al. Universal domain adaptation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2720-2729.

[57] Fu B, Cao Z, Long M, et al. Learning to detect open classes for universal domain adaptation[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16. Springer International Publishing, 2020: 567-583.

[58] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008, 9(11).