Expert Systems With Applications 237 (2024) 121585

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-source weighted source-free domain transfer method for rotating machinery fault diagnosis

Qinhe Gao [a], Tong Huang [a], Ke Zhao [b,*], Haidong Shao [c], Bo Jin [d]

[a] National Key Discipline Laboratory of Armament Launch Theory & Technology, Rocket Force University of Engineering, Xi'an 710025, China
[b] Key Laboratory of Road Construction Technology & Equipment Ministry of Education, Chang'an University, Xi'an 710054, China
[c] College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China
[d] Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Coimbra 3030-290, Portugal

ARTICLE INFO

Keywords:
Data privacy
Multi-source weighted source-free domain transfer
Reinforced information maximization strategy
Improved self-training mechanism

ABSTRACT

The mainstream approach to addressing the issues of insufficient historical data and high annotation costs in the domain of rotating machinery is to build transfer learning models based on labeled multi-source data. However, the practical diagnosis of failure cases often relies on data privacy, thereby limiting the widespread application of current multi-source domain transfer approaches for the 'data silos' problem of. In view of the above problem, a multi-source weighted source-free domain transfer approach is designed for rotating machinery fault diagnosis, and the designed scheme can efficiently achieve data privacy and domain transfer. Specifically, the proposed approach achieves knowledge transfer from the source to the target during the training process of the unlabeled target data without accessing the source data. This is accomplished through the utilization of a designed reinforced information maximization strategy and improved self-training mechanism. Additionally, a weighted strategy is devised to automatically apply optimal values to all source domains based on their relevance to the target domain. The proposed framework demonstrates accuracy exceeding 96% across eight cross-domain diagnostic cases in two sets of rotating machinery data, with an average accuracy of 98.26%. These results underscore the exceptional ability of the proposed method to address cross-domain fault diagnosis in rotating machinery while ensuring privacy protection.

1. Introduction

Rotating machinery, such as bearings and gears, serves as crucial components in transmission systems and finds extensive applications in industries related to national defense security and the national economy, including aerospace, weaponry, and manufacturing (Li, Zhong, & Shao, 2022; Liu, Jiang, Liu, Yang, & Sun, 2022; Wu, Jiang, & Zhu, 2023a). The occurrence of faults or performance degradation in rotating machinery can easily lead to a decline in overall equipment performance or even system failure. Therefore, fault diagnosis of rotating machinery has been a focal point of research (Dong, Zhan, & Hu, 2023; Mohajer, Daliri, & Mirzaei, 2022a).

In recent years, traditional way to fault diagnosis has shifted towards intelligent way, and significant success has achieved in deep intelligent diagnosis approaches. These methods primarily leverage complex neural networks to extract hidden information from high-dimensional features, establish correlations between input data and predicted categories, and achieve end-to-end identification (An, Jiang, & Yang, 2022; Hou, Zhang, & Jiang, 2023). However, the application of intelligent diagnosis methods relies on two critical prerequisites: having sufficient labeled data and consistent data distribution (Li, Xu, Li, Yang, & Lei, 2023a). In actual operational conditions of rotating machinery, which are characterized by complexity and variability, the vibration signals of rotating machinery differ significantly under different conditions. This severely limits the practical application of intelligent diagnosis methods.

Transfer learning brings new perspectives to the problem of fault diagnosis in rotating machinery under complex and variable operating conditions. It tackles the challenge by transferring knowledge from different operating condition data to address the issue of identifying unknown operating condition data (Mohajer, Sorouri, & Mirzaei, 2022b; Zhang, Wang, & Li, 2023a). Existing transfer diagnosis approaches can be categorized into single-source diagram and multi-source diagram, with the former being predominant (Zhao, Jia, & Shao, 2023a). However, in practical industrial scenarios, due to the limitations of a single

* Corresponding author.
E-mail addresses: qhgao201@163.com (Q. Gao), zhaoke@chd.edu.cn (K. Zhao), hdshao@hnu.edu.cn (H. Shao), jin.bo@isr.uc.pt (B. Jin).

https://doi.org/10.1016/j.eswa.2023.121585
Received 6 June 2023; Received in revised form 24 August 2023; Accepted 10 September 2023
Available online 14 September 2023
0957-4174/Published by Elsevier Ltd.

operating condition, single-source domain transfer learning methods often suffer from domain mismatch, resulting in negative transfer and inadequate generalization capability (Chen, Liao, & Li, 2023). On the other hand, by leveraging data from multiple operating conditions, multi-source domain transfer approaches can effectively capture knowledge from diverse conditions, extract fault features that are easier to recognize, and thus hold promising application prospects (Xiao, Shao, & Han, 2022).

Although transfer learning methods have achieved exciting results, they require accessing the source data during the training process to achieve the target identification. However, in practical industrial scenarios, data often contains sensitive information belonging to enterprises or companies, limiting the widespread application of transfer learning methods due to data privacy concerns. Recently, federated transfer learning, which combines transfer learning with federated learning, has garnered attention in fault diagnosis as it can achieve both domain transfer and privacy protection (Wang, Yan, & Yu, 2023a). Nevertheless, federated learning involves significant communication between global and local servers, leading to substantial network bandwidth usage and significantly reducing model training efficiency (Liu, Shen, & Gao, 2023). Furthermore, existing federated transfer fault diagnosis methods still have limitations, such as assuming a shared private space to align the source and target domain distributions or requiring prior knowledge of the source or target domain data distributions.

To address these challenges, this paper proposes a novel multi-source weighted source-free domain transfer network (MWSDTN). A pivotal aspect of this study centers on data privacy, where training the target data does not involve accessing source domain data. The inability to access source domain features impedes the applicability of traditional domain adaptation strategies for feature alignment. Drawing inspiration from Ref (Liang et al., 2020), the MSWFSN employs an information maximization strategy to fine-tune the source model for target domain data classification. Recognizing that different source domains contribute distinctively to the identification performance of the target domain, a weighted strategy is introduced, iteratively optimizing the optimal weights for each source model. Furthermore, relying solely on the information maximization strategy may fall short of ensuring precise recognition of target domain features. Consequently, an improved self-training mechanism is developed by amalgamating a self-training mechanism with the weighted strategy to augment the predictive outcomes of the target model for the target domain. In summary, MWSDTN achieves both domain transfer and privacy protection in a simple and efficient manner. The illustration of the multi-source source-free domain transfer method is presented in Fig. 1. The key contributions are concluded as:

1. A sophisticated source-free domain transfer framework is designed for rotating machinery fault diagnosis, achieving both domain transfer and privacy protection.
2. An unsupervised loss is devised to automatically assign weights to each source model based on the difficulty of knowledge transfer from the source domain to the target domain, generating an accurate target identification model.
3. An improved self-training mechanism is developed to enhance the representation learning of the target domain, further improving the identification performance in the target domain.

The remainder of this work is designed as follows. The related works about transfer learning, source-free transfer learning and federated transfer learning are elaborated in Section 2. The detailed description of MWSDTN is illustrated in Section 3. The experimental verification is presented in Section 4. This work is concluded in Section 5.

2. Related works

2.1. Transfer fault diagnosis

In practical scenarios, on one hand, the diverse working environments result in significant variations in vibration data distribution for rotating machinery, making it challenging for deep learning methods to adapt. On the other hand, rotating machinery is typically operated in a healthy state, making it difficult to obtain fault data and data labels. Consequently, deep learning-based intelligent diagnostic methods struggle to address rotating machinery fault problems in real-world conditions. As a branch of machine learning, transfer learning overcomes the limitations of deep learning and promotes the application of intelligent diagnostic methods in practical scenarios.

In recent years, transfer learning is very popular in fault diagnosis. Zhu *et al.* constructed a transfer framework based on multi-kernel maximum mean discrepancy (MK-MMD), and conducted massive experiments to testify its diagnosis performance (Zhu, Chen, & Shen, 2019). Mao *et al.* distilled the relevant information of multiple failure modes, and conducted domain adversarial network for further diagnosis

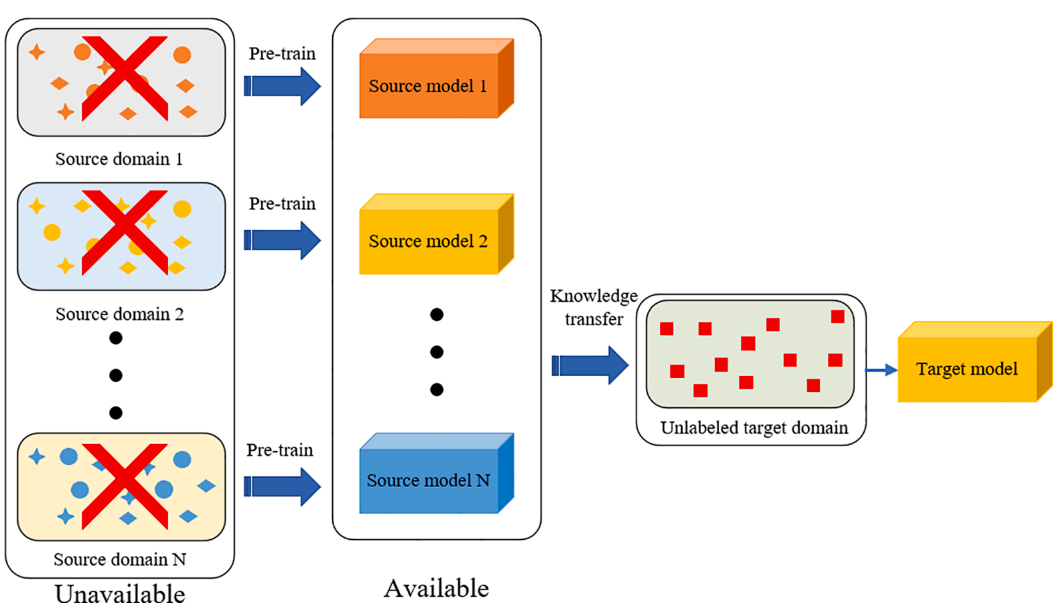

Fig. 1. The presentation of multi-source source-free domain transfer method.

(Mao, Liu, Ding, Safian, & Liang, 2021). Zhao *et al.* constructed the transfer network combined with joint maximum mean discrepancy and domain confusion, which shows enhanced performance in diagnosing rotating machinery issues (Zhao, Jiang, & Wang, 2021). Zhao *et al.* devised an alignment approach that mitigates conditional distribution shift by reducing inter-class center distances and applies adversarial strategies to alleviate marginal distribution discrepancies (Zhao, Liu, & Shen, 2022). Zhang *et al.* employed the local maximum mean difference to diminish inter-domain distribution dissimilarities and utilized K-means to excavate fault knowledge from the target domain data (Zhang, He, & Wang, 2022a). Zhou *et al.* introduced a dynamic adaptive migration framework based on weighted factors to dynamically align conditional and marginal probability distributions (Zhou, Dong, & Zhou, 2021).

To further improve cross-domain diagnostic performance, scholars have turned their attention to multi-source domain transfer learning. Yang *et al.* partitioned the multi-source cross-domain diagnostic problem into individual single-source cross-domain diagnostic problems, followed by weighted ensemble integration of all classifiers, culminating in the final fault diagnostic output (Yang, Kong, & Wang, 2021). Chen *et al.* formulated a weighted transfer model that precisely quantifies the importance of various source domains, validated the model's effectiveness across numerous cross-domain fault diagnosis tasks with bearing and gearbox datasets (Chen, Liao, & Li, 2022a). Tian et al. aligned the domain data using multiple subnetworks and obtained the weighted coefficients of all source classifiers, ultimately outputting the weighted diagnostic results (Tian, Han, & Li, 2022). Cao et al. developed a two-stage multi-source domain transfer method based on dilated convolutional neural networks, demonstrating outstanding diagnostic performance in experimental validation (Cao, Meng, & Sun, 2023). Wu *et al.* employed class-conditional maximum mean discrepancy to diminish conditional probability distribution differences across diverse domains and generated integrated diagnostic results (Wu, Jiang, & Liu, 2023b). Tian *et al.* constructed a multi-source migration model applicable to open sets, quantifying similarities between source domains and the target domain, and employing adversarial strategies for fusion. Moreover, the application of co-supervision strategies further enhanced knowledge transfer effects (Tian, Han, & Karimi, 2023). Although the aforementioned transfer fault diagnosis methods have achieved remarkable results, they require continuous access to the source data as training progress, disregarding the privacy protection of the data. To address this issue, the proposed source-free domain transfer network effectively and simply achieves both domain transfer and privacy protection.

2.2. Federated transfer fault diagnosis

Federated learning, which utilizes distributed model training, has addressed the issue of data privacy protection. Recently, the combination of transfer learning algorithms and federated learning algorithms, known as federated transfer learning methods, have gained significant profile in fault diagnosis. Xia *et al.* proposed an advanced federated learning method by sharing the model information to guarantee the data privacy, and conducted sufficient experiments to verify (Xia, Zheng, & Li, 2022). Wang *et al.* constructed a federated transfer framework to detect the GIS insulation failures (Wang, Yan, & Yang, 2022a). Zhang et al. designed a federated transfer approach to address rotating machinery fault problems, assuming prior knowledge of data distribution (Zhang & Li, 2022a). Zhang *et al.* designed a federated adversarial transfer network to achieve the domain transfer and privacy protection (Zhang & Li, 2022b). Zhao *et al.* combined maximum mean discrepancy and domain discriminator to construct the federated transfer network with multi-sources, and evaluated the diagnosis performance with multiple transfer diagnosis cases (Zhao, Hu, Shao, & Hu, 2023b). However, this approach constructs a global space to align the source and target domain distributions, raising concerns about data privacy. Chen *et al.* reinforced the federated averaging algorithm, and then proposed a federated transfer approach to identify bearings faults (Chen, Li, & Huang, 2022b). Kang *et al.* constructed the transfer ResNet with conditional maximum mean discrepancy, and integrated the transfer ResNet to federated framework. Finally, the federated transfer framework is applied for bearing fault diagnosis with data privacy (Kang, Yang, Sun, Wang, Wang, & Mikulovich, 2022). Wang *et al.* harnessed batch normalized maximum mean discrepancy to obliterate distribution disparities and proposed an advanced aggregation algorithm to attain a global model, thoroughly verifying the method's diagnostic performance across multiple datasets (Wang, Yan, & Yu, 2023b). Lu *et al.* employed encryption algorithms to ensure data privacy within each domain and integrated their migration framework to address cross-domain fault diagnosis under the umbrella of data privacy (Lu, Gao, & Xu, 2022). Li *et al.* accomplished the transfer of source domain knowledge to the target domain through transfer training of model information, combined with target domain adaptation for distribution alignment, and its efficacy is corroborated through comparison with other state-of-the-art algorithms (Li, Zhang, & Li, 2023b). Wang *et al.* introduced adversarial learning strategies to align the reference distribution of the target domain with the source domain, facilitating knowledge migration within a federated framework (Wang, Huang, & Shi, 2022b). Despite showcasing exemplary performance in experiments, this approach may face limitations as actual reference distributions of real-world data are often inaccessible. Although these federated transfer diagnosis methods aim to achieve both domain transfer and data privacy, some methods require prior knowledge of data distribution, limiting their generalization in practical scenarios, and the privacy protection of certain methods remains questionable. Furthermore, the federated framework heavily relies on a central server, making it susceptible to single point failures, and the communication costs of these methods are also high.

2.3. Source-free domain transfer fault diagnosis

The source-free transfer learning algorithm necessitates no sharing of source domain data; instead, it involves the application of the learned source model from the source domain to fine-tune the model in the target domain. During the fine-tuning process, the target domain remains oblivious to the source domain data, and the training of the source model remains isolated from the target domain data as well. In comparison with conventional transfer paradigms, the source-free transfer approach obviates the requirement for sensitive data transmission throughout the entire training phase. This considerable reduction in data transmission curtails the risks associated with potential leaks of source and target domain data, thus effectively upholding data privacy. Furthermore, the source-free transfer learning algorithm boasts efficiency and low network resource costs, underscoring its promising potential for applications in rotating machinery fault diagnosis. Jiao *et al.* introduced a source-free domain transfer diagnosis scheme, which fixes the classifier and utilizes self-training mechanism and label prediction matrix constraints to train a target-domain-specific feature extractor (Jiao et al., n.d.). Zhu *et al.* utilized mean teacher training scheme and information maximization to enhance the generalization of source-free domain transfer diagnosis framework (Zhu, Zeng, Liu, & Adaptation, 2022). Li *et al.* designed an innovative source-free convolutional neural network based on failure mode clustering mechanism and attention mechanism (Li, Yue, & Huang, 2023c). Wang *et al.* conducted nearest centroid filtering to obtain more accurate target labels, and used supervised learning to improve the target model performance. The developed source-free domain transfer approach is demonstrated by electro-mechanical actuator dataset (Wang, Zhang, & Miao, 2023c). Yue *et al.* adopted the information maximization mechanism and self-supervised paradigm to ensure that the trained source model can extract typical fault features from the target domain (Yue, Li, Li, & Adaptation, 2022). Zhang *et al.* developed a clustering method for training a target-domain-specific diagnostic model (Zhang, Jiang, Zhang, & Adaptation, 2022b). Zhang *et al.* established a deep network incorporating Transformer architecture for supervised training in the source domain, following which the well-trained model is dispatched to the target domain. The integration of supervised contrastive learning strategies and entropy loss facilitated precise fine-tuning of the source model and accurate identification within the target domain (Zhang, Ren, & Feng, 2023b).

Seeing that existing source-free domain transfer approaches focus primarily on the paradigm from single source to the target. Motivated by these methods, a novel multi-source weighted source-free domain transfer network is constructed for rotating machinery fault diagnosis.

3. Methodology

3.1. Problem definition

The data measured from different conditions of rotating machinery equipment form individual domains, and there exist significant differences in data distributions among these domains. Let $D_S^n = \{(x_{s_n}^i, y_{s_n}^i)\}_{i=1}^{K_n}$ represents the n_{th} labeled source domain, K_n denotes the corresponding sample quantity in the source domain, and $\{P_{S_n}(x,y)\}_{n=1}^{N}$ defines the probability distribution of the source domain. Let $D_T = \{(x_T^i)\}_{i=1}^{K}$ represents the unknown target domain, K means the corresponding sample quantity in the target domain, and $\{P_T(x,y)\}$ defines the probability distribution of the target domain. It is assumed that the health status of all source and target domains is consistent, and $(P_{S_i} \neq P_{S_j} \neq P_T, i \neq j, i, j \in N)$. In contrast to conventional multi-source transfer learning methods, the proposed approach, unable to access source data D_S. Specifically, the multi-source source-free transfer model can only access a series of source models θ_S trained based on all source domains $D_S = \{(x_s, y_s)\}$. Therefore, the goal is using θ_S and D_T to obtain a trained target model θ_T that can precisely predict the corresponding labels of $D_T = \{(x_T^i)\}_{i=1}^{K}$.

3.2. Source model pretraining

The proposed framework achieves domain-specific feature extraction for the target domain by pretraining the source models and adapting them to multiple source domains, as shown in Fig. 2. All source models share the same structure, where θ_S includes a feature extractor $\varphi_S:\chi\rightarrow\mathbb{R}^f$ and a classifier $\nu_S:\mathbb{R}^f\rightarrow\mathbb{R}^c$, with f represents the dimension of the extracted features and c denotes the total number of failure modes, $\theta_S = \varphi_S \circ \nu_S$.

The source model $\theta_S^n:\chi\rightarrow\mathbb{R}^c$ is trained using the source domain $D_S = \{(x_s^i, y_s^i)\}_{i=1}^{K}$, optimizing the following objective:

$$\mathscr{L}_{cls} = -\mathbb{E}_{x_s \in X_S} \sum_{c=1}^{C} z_c \log\sigma_c(\theta_S(x_s)) \tag{1}$$

where $\sigma_c(a) = \frac{\exp(a_c)}{\sum_{i=1}^{C}\exp(a_i)}$ represents a C-dimensional vector in the softmax output, representing the one-of-K encoding, with the correct category having a value of "1" and the other categories having a value of "0".

The introduction of label smoothing strategy enhances the training effectiveness of the source models and improves the adaptability of the source and target domain distributions. The optimization objective is represented as:

$$\mathscr{L}_{cls}^{ls} = -\mathbb{E}_{x_s \in X_S} \sum_{c=1}^{C} z_c^{ls} \log\sigma_c(\theta_S(x_s)) \tag{2}$$

where $z_c^{ls} = (1-\alpha)z_c + \alpha/C$ represents the smoothed labels, and α is the smoothing coefficient set to 0.1.

3.3. Multi-source domain adaptation

This study addresses the realm of multi-source domain adaptation, wherein variations in the efficacy of source domains on target domain recognition are acknowledged. To account for these discrepancies, an

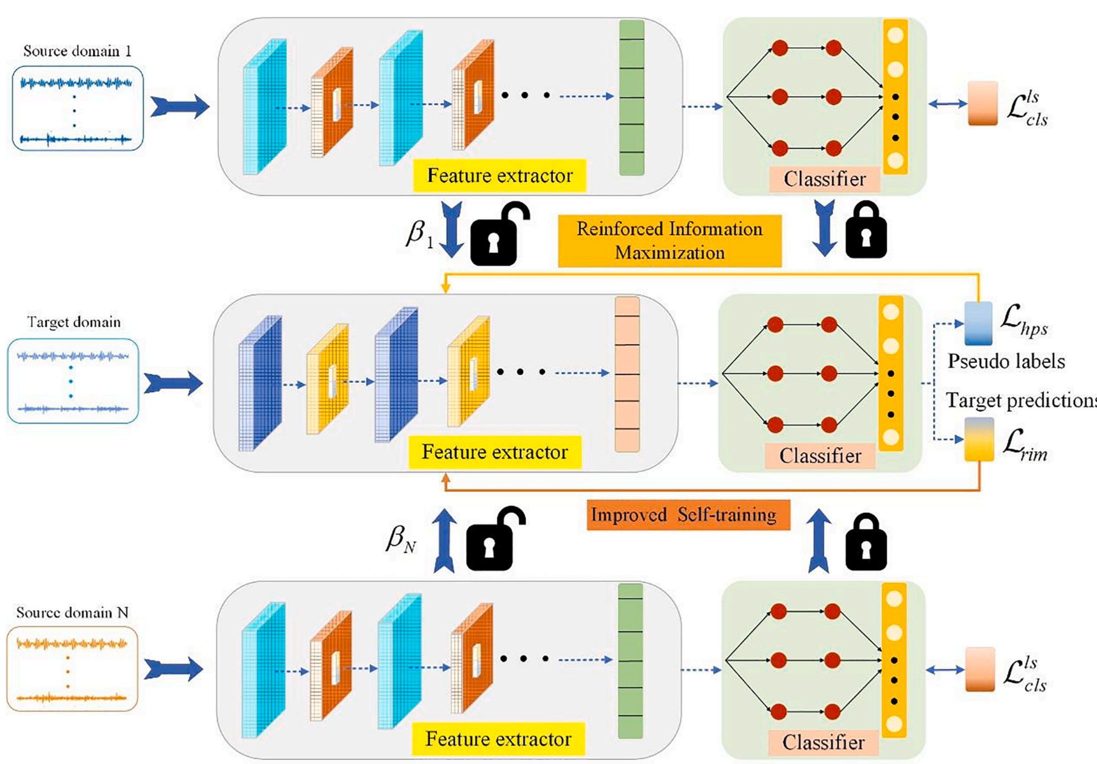

Fig. 2. The illustration of MWSDTN. Source models are trained with all source domains, and then the source models with corresponding weights are utilized to construct the target model, where the target model is trained with target domain.

approach involving the allocation of a set of weights to all source domains, denoted as $\{\beta_n\}_{n=1}^{N}$, and $\sum_{n=1}^{N}\beta_n = 1$. These weights encapsulate the probability mass functions across source domains. In instances where the source domain exhibits a higher propensity for easy transferability to the target domain, the corresponding weight is proportionally augmented, and vice versa. The intricate workings of the reinforced information maximization and improved self-training strategies, developed in conjunction with the weighted approach, are expounded upon as follows.

3.3.1. *Reinforced information maximization*

For transfer learning algorithms, precisely aligning the feature distributions is crucial for achieving accurate classification in the unknown target domain. However, this study concentrates upon the cross-domain fault diagnosis issue under data privacy constraints and refuses to access source data, making it impossible to obtain the source distribution. Considering that if the target data is mapped to features similar to the source data, the classifier of the trained source model would produce similar outputs for the target data as it does for the source data, indicating that the output for the target domain data is close to one-hot encoding (as it accurately identifies the category, the vector should be close to one-hot; otherwise, the probabilities for each category would be similar). Therefore, by fixing the source classifier's parameters, the proposed approach fine-tunes the parameters of the source feature extractor by maximizing the information between the intermediate features and the classifier's output. Specifically, the information maximization strategy is introduced to minimize the differences across domains, including conditional entropy and category diversity. The conditional entropy term is denoted as:

$$\mathscr{L}_{cet} = -\mathbb{E}_{x_t\in D_T}\left[\sum_{c=1}^{C}\sigma_c(\theta_T(x_t))\log\sigma_c(\theta_T(x_t))\right] \tag{3}$$

where $\theta_T(x_t) = \sum_{n=1}^{N}\beta_n\theta_S^n(x_t)$, β_n means the weights of the source model θ_S^n, and $\sum_{n=1}^{N}\beta_n = 1$.

From Eq. (3), it can be observed that when the source model is more amenable to transfer, the corresponding conditional entropy value becomes smaller, and minimizing Eq. (3) increases the weight of that source model. Furthermore, considering that minimizing the conditional entropy term alone may lead to a concentration of prediction results in a single category, category diversity is introduced to further enhance the prediction effectiveness for the target domain.

$$\mathscr{L}_{cdt} = \sum_{c=1}^{C}\overline{q}_c\log\overline{q}_c \tag{4}$$

where $\overline{q} = \mathbb{E}_{x_t\in D_T}[\sigma(\theta_T(x_t))]$ denotes the mean outputs of the whole target data. The final reinforced information maximization loss can be calculated as:

$$\mathscr{L}_{rim} = \mathscr{L}_{cet} + \mathscr{L}_{cdt} \tag{5}$$

3.3.2. *Improved self-training*

Although the proposed reinforced information maximization strategy improves the robustness and diversity of target predictions, misclassifications of target predict labels are still inevitable. Therefore, an improved self-training mechanism is proposed to enhance the target model's predictions for the target domain. Specifically, the centroids of all categories in the target dataset are obtained as follows:

$$\lambda_{c_n}^{(0)} = \frac{\sum_{x_t\in D_T}\sigma_c(\widehat{\theta}_S^n(x_t))\widehat{\nu}_S^n(x_t)}{\sum_{x_t\in D_T}\sigma_c(\widehat{\theta}_S^n(x_t))} \tag{6}$$

where $\lambda_{c_n}^{(t)}$ represents the cluster centroid of category c at iteration t, which is generated by the n_{th} source model. $\widehat{\theta}_S^n$ means the n_{th} source model of the previous iteration, and $\widehat{\nu}_S^n$ means the n_{th} source classifier of the previous iteration. After obtaining the centroids for each category from all source models on the target dataset, they are aggregated according to the following equation to obtain the final centroids of the target dataset.

$$\lambda_c^{(0)} = \sum_{n=1}^{N}\beta_n\lambda_{c_n}^{(0)} \tag{7}$$

Inferred from the nearest cluster centroid theory, the initial pseudo-labels of all target samples can be calculated by:

$$\widehat{y}_t^{(0)} = \underset{c}{\operatorname{argmin}}\left\|\widehat{\theta}_T(x_t) - \lambda_c^{(0)}\right\|_2^2 \tag{8}$$

The centroids of each category in the target dataset and the pseudo-labels corresponding to target samples are updated through iterative training, denoted as follows:

$$\lambda_{c_n}^{(1)} = \frac{\sum_{x_t\in D_T}\mathbf{1}\left\{\widehat{y}_t^{(0)} = c\right\}\widehat{\nu}_S^n(x_T)}{\sum_{x_t\in D_T}\mathbf{1}\left\{\widehat{y}_t^{(0)} = c\right\}} \tag{9}$$

$$\lambda_c^{(1)} = \sum_{n=1}^{N}\beta_n\lambda_{c_n}^{(1)} \tag{10}$$

$$\widehat{y}_t^{(1)} = \underset{c}{\operatorname{argmin}}\left\|\widehat{\theta}_T(x_t) - \lambda_c^{(1)}\right\|_2^2 \tag{11}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function with the parameter being 1 for true outputs. The cross-entropy loss function of the pseudo-labels is calculated by:

$$\mathscr{L}_{hps} = -\mathbb{E}_{x_t\in D_T}\sum_{c=1}^{C}\mathbf{1}\{\widehat{y}_t = c\}\log\sigma_c(\theta_T(x_t)) \tag{12}$$

Algorithm 1
Pseudo-code of MWSDTN

Input: Pre-trained source models set $\{\theta_S^n\}_{n=1}^{N}$; Weights set to be trained $\{\beta_n\}_{n=1}^{N}$; Target dataset$D_T = \{(x_T^i)\}_{i=1}^{K}$; Bach size B; Training epoch T.
Output: Optimal target model θ_T^*
Initialization: Frozen classifiers $\{\nu_S^n\}_{n=1}^{N}$ and set initial weights $\{\beta_n\}_{n=1}^{N}$.
for t=1: T **do**
1. Obtain the pseudo labels of D_t according to **Eq. (6)**.
2. Obtain the mean outputs of the whole target data $\overline{q}$ according to **Eq. (2)**.
for b=1: B **do**
1. Small batches of samples are randomly selected from the target set and sent to all source models.
2. The losses for each component are computed according to **Eq. (3)**, **Eq. (4)**, and **Eq. (12)**.
3. The total loss is computed according to **Eq. (13)**.
4. The parameters $\{\varphi_S^n\}_{n=1}^{N}$ and $\{\beta_n\}_{n=1}^{N}$ are updated according to **Eq. (14)**.
5. The values of β_n are ensured to be positive according to $\beta_n = 1/(1+e^{-\beta_n})$.
6. β_n is normalized according to $\beta_n = \beta_n/\sum_{n=1}^{N}\beta_n$.
End for
End for
Return Optimal target model θ_T^*

3.4. *Training objective*

Combining the discussion in Section 3.3, the overall objective of target domain training is denoted as Eq. (13), and the solving problem is illustrated in Eq. (14).

$$\mathscr{L}_{total} = \mathscr{L}_{rim} + \mu\mathscr{L}_{hps} \tag{13}$$

$$\begin{aligned}&\text{minimize } \mathscr{L}_{total}, \{\varphi_S^n\}_{n=1}^N, \{\beta_n\}_{n=1}^N \\ &\text{subject to } \beta_n \geqslant 0, \forall n \in \{1,2,\cdots,N\}, \sum_{n=1}^{N}\beta_n = 1\end{aligned} \tag{14}$$

After satisfying the stopping condition of the iterations, the optimal sets $\{\varphi_S^{n*}\}_{n=1}^N$ and $\{\beta_n^*\}_{n=1}^N$ are obtained, leading to a well-trained target model. The proposed algorithm's flowchart is depicted in **Algorithm 1**.

4. Experiments

4.1. Dataset illustration

Dataset 1: A specially designed bearing fault diagnosis test rig is used for vibration signal acquisition in this experiment, as shown in Fig. 3 (Jia, Wang, & Zhang, 2022). The experimental setup consisted of a motor, rotor, bearing seat, shaft coupling, gear box and brake. Three-axis vibration accelerometers are installed on the bearing seat surface, and the vibration signals are measured at 25.6 kHz. The rolling bearing type is chosen as NU205EM, and 10 different health conditions are designed, including normal condition and three types of inner race faults (0.2 mm, 0.4 mm, and 0.6 mm), three types of outer race faults (0.2 mm, 0.4 mm, and 0.6 mm), and three types of roller faults (0.2 mm, 0.4 mm, and 0.6 mm). The data are collected for each health condition under different loads of 0 N, 20 N, 40 N, and 60 N. Each health condition has 200 samples, with each sample consisting of 1568 data points, resulting in a total of 2000 samples. The time-domain samples are processed using FFT transformation to obtain the corresponding frequency-domain samples. The datasets from the four operating conditions are named T1, T2, T3, and T4, and four sets of multi-source transfer diagnosis cases are established.

Dataset 2: The Paderborn University rolling bearing public dataset is introduced to further validate the diagnostic performance, as shown in Fig. 4 (Lessmeier et al., 2016). This dataset consists of five bearing health conditions: healthy bearing, inner race faults (minor and severe), and outer race faults (minor and severe). The fault conditions are obtained through artificial damage (electric spark, drilling, engraving) and accelerated life tests. The tested bearing type is chosen as 6203. Data are collected on four conditions, corresponding to rotation speed, torque, and load of 1500 rpm/0.1 Nm/1000 N (T5), 1500 rpm/0.7 Nm/400 N (T6), 1500 rpm/0.7 Nm/1000 N (T7), and 900 rpm/0.7 Nm/1000 N (T8). Each health condition has 300 samples, with each sample consisting of 1568 data points, resulting in a total of 1500 samples. The frequency-domain samples are obtained by applying FFT transformation to the time-domain samples. Four sets of multi-source transfer diagnosis cases are established. Fig. 5.

4.2. Detailed settings of MWSDTN

Based on previous experience in network construction, the structural parameters of the proposed MWSDTN framework are designed as presented in Table 1. 'Conv' represents a 2D convolutional layer, 'Pool' represents a max pooling layer, 'BN' represents a batch normalization layer, and ReLU represents the corresponding activation function. 1–64 indicates that the input channel is 1, and the output channel is 64. The kernel size is 5x1, and the pooling size is 2x1.

Both the source model pre-training and target domain adaptation stages employed the mini-batch SGD algorithm, with a momentum of 0.9 and weight decay of 0.001. In the source model training stage, the learning rate is set to 0.01, the number of epochs is set to 20, and the batch size is set to 32. In the target domain adaptation stage, the corresponding parameters are set as 0.001, 20 and 64. For all experiments, parameter μ is fixed at 0.3.

4.3. Comparison methods

In this study, several advanced transfer learning approaches are introduced to compare with MWSDTN, including conventional transfer approaches, source-free domain transfer approaches, and federated transfer approach. In this study, the classic transfer learning methods MAD and DATN, widely applied in the fault diagnosis domain and commonly used as benchmarks for assessing algorithmic diagnostic performance, are selected as comparative approaches. Additionally, Source hypothesis transfer (SHOT), a prominent source-free domain transfer method, has been thoroughly validated in the detection outcomes across three target detection datasets within multiple transfer learning methodologies, which validates its performance. Furthermore, SHOT's knowledge transfer concept has informed the design of MWSDTN; hence, SHOT is chosen for comparison. Source-free adaptation diagnosis (SFAD), having established itself as a verified source-free domain transfer approach within the fault diagnosis domain, is introduced to more comprehensively illustrate MWSDTN's diagnostic capabilities. Given the research focus on multi-source cross-domain fault diagnosis with a concern for data privacy, and since federated learning algorithms are designed to address data privacy, the well-performing Federated Knowledge Alignment (FedKA) is chosen as a comparative measure to evaluate the strengths and weaknesses of the proposed MWSDTN framework. As for conditional weighting transfer Wasserstein auto-encoder (CWTWAE)it is a multi-source cross-domain fault diagnosis method adept at precise cross-domain fault diagnosis while assigning weights to source domains based on their similarity to the target domain. Consequently, CWTWAE is selected to assess MWSDTN's cross-domain diagnostic performance. The parameters for each method are as follows.

MAD: MK-MMD is popular in transfer learning to transfer knowledge from source to target by aligning their marginal probability distribution. Therefore, a transfer diagnostic model based on the network in Table 1 is constructed, where MK-MMD is added at the FC1 layer. For parameter optimization, the SGD algorithm is employed with a learning rate set at 0.001, and the number of iterations is fixed at 200. The coefficient is $2/(1+\exp(-10\times p))-1$, while p progressively transitioned from 0 to 1.

DATN: Adversarial strategies are commonly used in transfer learning to distill domain-invariant feature representations across domains. Therefore, a domain adversarial transfer network (DATN) is built by

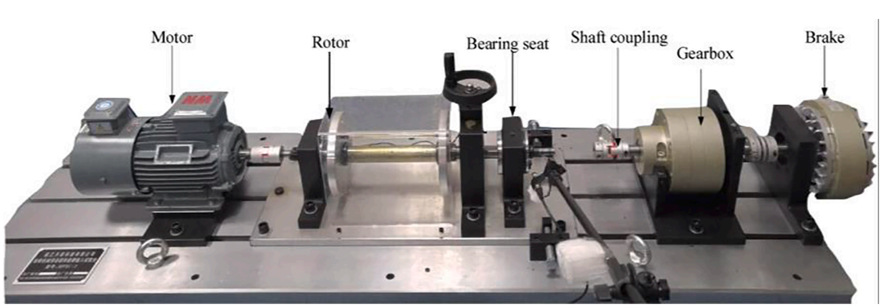

Fig. 3. Photo of experimental equipment.

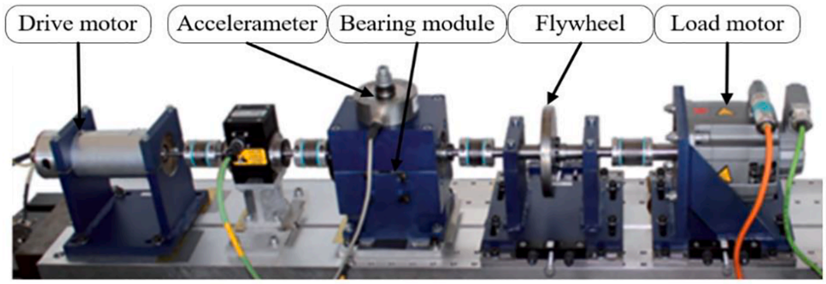

Fig. 4. The rolling bearing experimental platform.

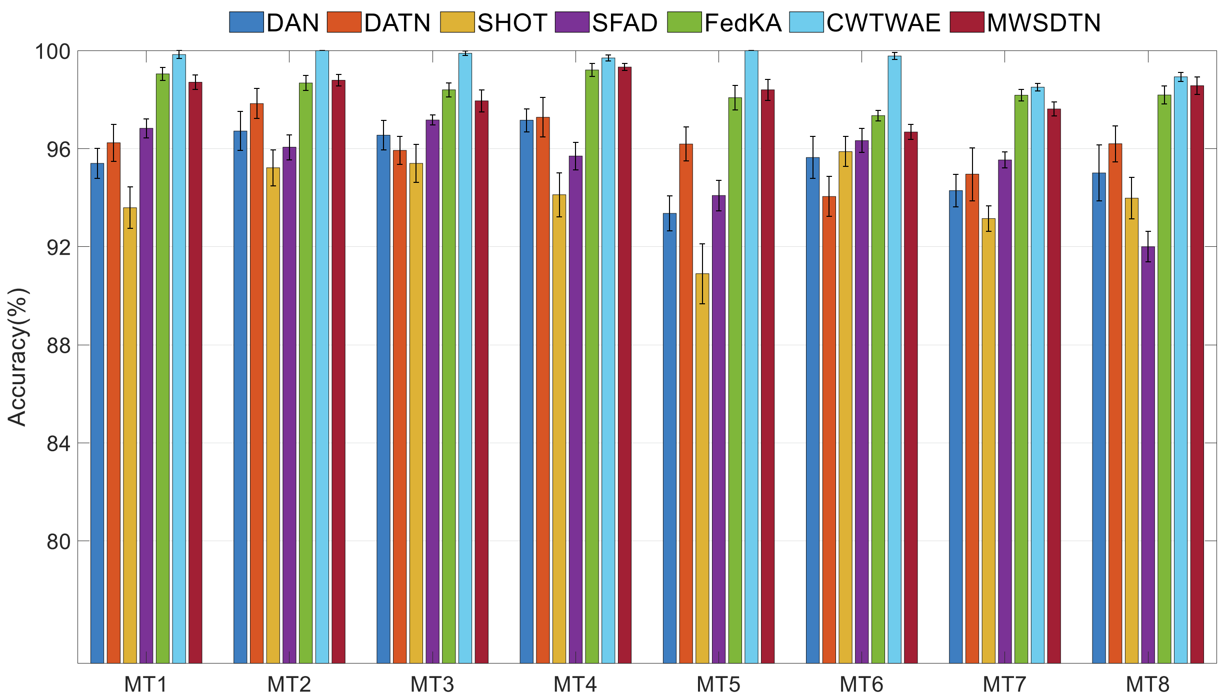

Fig. 5. The diagnosis results of all approaches.

Table 1
The structural parameters of MWSDTN.

Module	Layer	Parameter information	Output dimension
Feature extractor	Input	/	1 × 28 × 28
	Conv1	1–64, 5 × 1, BN, ReLU	64 × 24 × 24
	Pool1	2 × 1	64 × 12 × 12
	Conv2	64–50, 5 × 1, BN, Dropout, ReLU	50 × 8 × 8
	Pool2	2 × 1	50 × 4 × 4
Classifier	FC0	/	800
	FC1	ReLU	150
	Output	Softmax	C

combining the network in Table 1 with a domain discriminator which is constructed by fully connected layers. For parameter optimization, the SGD algorithm is employed with a learning rate set at 0.001, and the number of iterations is fixed at 200. The coefficient is $2/(1+\exp(-10\times p))-1$, while p progressively transitioned from 0 to 1.

SHOT: SHOT is a representative source-free domain transfer method that is introduced in this study for comparison (Liang et al., 2020). To achieve fair comparison, the network structure of SHOT is also set according to Table 1. Both the pre-training phase of the source model and the target domain adaptation phase were executed using the mini-batch SGD algorithm. The momentum is set to 0.9, while weight decay is configured at 0.001. Throughout the source model training and the target domain adaptation stages, the learning rate is established at 0.001, with an epoch iteration count of 15 and a batch size of 64. Across all experiments, $\mu = 0.3$.

SFAD: To thoroughly validate the diagnostic performance of MWSDTN, SFAD, an advanced source-free domain transfer framework in fault diagnosis, is adopted to compare with the proposed method (Jiao et al., n.d). Similarly, to ensure fairness in the experiments, the network structure of SFAD is also set according to Table 1. The parameter optimization is conducted via the mini-batch SGD algorithm, with a momentum of 0.9 and weight decay of 0.0005. The initial value of the learning rate η_0 is 0.01, subsequently updated according to $\eta = \eta_0/(1+10\times\lambda)^{0.75}$, where λ incrementally transitioned from 0 to 1. The batch size is set at 32, with iteration counts of 10 and 100 for the two respective stages.

FedKA: FedKA as a federated transfer approach, can ensure data privacy when solving multi-source transfer problems (Sun, Chong, & Ochiai, 2023). Therefore, FedKA is used as a benchmark to compare with our proposed method MWSDTN. For parameter optimization, the Adam algorithm is employed with a learning rate set at0.0003, and the number of iterations is fixed at 100. The coefficient is $\lambda = 2/1 + \exp(-5q) - 1$, while q progressively transitioned from 0 to 1.

CWTWAE: CWTWAE as a multi-source domain transfer method has demonstrated excellent diagnostic performance in three rolling bearing datasets (Zhao et al., 2023). Hence, CWTWAE is used for comparison with MWSDTN. For parameter optimization, the SGD algorithm is employed with a learning rate set at 0.005, and the number of iterations is fixed at 600. The coefficient is $2/(1 + \exp(-10 \times p)) - 1$, while p progressively transitioned from 0 to 1. More parameters can be seen in Ref (Zhao et al., 2023a).

4.4. Comparison results

All methods are validated on eight multi-source cross-domain fault diagnosis cases, Dataset 1 and Dataset 2. To accurately reflect the diagnostic performance and robustness of all methods, the average value and standard deviation of ten runs are adopted as the final diagnostic results, as shown in Table 2. Furthermore, all methods are implemented using the PyTorch framework and accelerated using a NVIDIA GPU, running on the Windows system. The computational efficiency of each method is evaluated by calculating their average running times on the eight multi-source transfer diagnosis cases. "MT4″ represents the transfer diagnosis task from source domains T1, T2, and T3 to the target domain T4, where the diagnostic results of single-source transfer learning methods are the best result of three single-source cross-domain diagnosis tasks: "T1 → T4," "T2 → T4," and "T3 → T4." "PP" indicates whether the method has privacy protection capability.

From Table 2, it can be observed that DAN, DATN, and CWTWAE neglect data privacy, while SHOT, SFAD, FedKA, and MWSDTN ensure data privacy. DAN, DATN, SHOT, and SFAD are single-source domain fault diagnosis methods, while FedKA, CWTWAE, and MWSDTN are multi-source domain fault diagnosis methods. The average accuracies of DAN, DATN, SHOT, and SFAD are 95.52%, 96.09%, 94.03%, and 95.47%, respectively. This means that DATN has the best diagnostic performance, DAN and SFAD have similar diagnostic performance, and SHOT has the lowest diagnostic performance. These single-source transfer learning methods show significant variations in diagnostic accuracy across different cross-domain diagnosis tasks. For instance, DAN achieves an average diagnostic accuracy of 97.16% on MT4 but only 93.36% on MT5. DATN achieves an average diagnostic accuracy of 97.84% on MT2 but only 94.95% on MT6. SHOT achieves an average diagnostic accuracy of 95.88% on MT6 but only 90.90% on MT5. SFAD achieves an average diagnostic accuracy of 97.17% on MT3 but only 92.00% on MT8. The maximum and minimum diagnostic accuracy differences for DAN, DATN, SHOT, and SFAD are 3.80%, 2.89%, 4.98%, and 5.17%, respectively. Moreover, these single-source transfer learning methods exhibit poor robustness. The average accuracies of FedKA, CWTWAE, and MWSDTN are 98.39%, 99.58%, and 98.26%, respectively. The comparison results indicate that CWTWAE has the best diagnostic performance, followed by FedKA, and the diagnostic performance of the proposed framework, MWSDTN, is slightly lower than that of FedKA. Compared to single-source transfer learning methods, multi-source transfer learning methods show smaller variations in diagnostic accuracy across different cross-domain tasks and better robustness. Regarding the runtime analysis of all methods, source-free transfer learning methods have longer runtimes compared to conventional transfer approaches. The runtime of multi-source transfer approach is shorter than the time required to run single-source transfer learning methods three times. Due to the complex communication settings, FedKA has a much longer runtime than other methods.

In conclusion, based on the comprehensive comparison of various advanced transfer learning methods, the diagnostic performance of MWSDTN has been thoroughly validated. The conclusions drawn from this study are as follows: (1) Under the premise of data privacy, source-free transfer learning methods often achieve inferior recognition results compared to advanced conventional transfer learning methods during the distribution adaptation phase, as they are unable to access the source domain data. (2) Compared to single-source transfer pattern, multi-source transfer pattern undoubtedly has brighter prospects, achieving higher diagnostic accuracy, faster training speed, and stronger diagnostic robustness. (3) Although the diagnostic accuracy of MWSDTN is 1.32% lower than that of CWTWAE, the value of a slight performance improvement is much lower than the value of maintaining data privacy. (4) The diagnostic accuracy of MWSDTN is only 0.13% lower than that of the advanced federated transfer learning method, FedKA, but the runtime is significantly shorter, indicating that the source-free domain transfer approach undoubtedly has a brighter application prospect compared to the federated transfer approach. In summary, the diagnostic performance of the proposed method has been fully demonstrated by comparing it with various advanced transfer learning methods.

4.5. Ablation study

In this section, ablation studies are conducted to assess the importance of the conditional entropy strategy, category diversity strategy, and improved self-training mechanism in the proposed MWSDTN framework. Specifically, the following comparative methods were employed: (1) using only the conditional entropy strategy (MDTN-CE), (2) using the reinforced information maximization strategy (MDTN-RIM), and (3) using only the improved self-training strategy (MDTN-IST). The average results of ten runs for each method were taken as the final results, as shown in Table 3 and Fig. 6. Fig. 7.

Known from Table 3 that the average accuracy rates of MDTN-CE, MDTN-RIM, MDTN-IST, and MWSDTN across eight cross-domain fault diagnosis tasks are 95.77%, 97.05%, 97.34%, and 98.26%, respectively. The comparative results indicate that MWSDTN exhibits the best diagnostic performance, followed by MDTN-IST, MDTN-RIM, and MDTN-CE. Furthermore, MWSDTN demonstrates significantly better robustness compared to the other three methods. The experimental results demonstrate that applying the reinforced information maximization strategy and the improved self-training mechanism can maximize the diagnostic performance of the source-free domain transfer model.

Table 2
The detailed diagnosis results of all experiments.

	DAN	DATN	SHOT	SFAD	FedKA	CWTWAE	MWSDTN
PP	×	×	√	√	√	×	√
MS	×	×	×	×	√	√	√
MT1	95.40±0.62	96.24±0.75	93.59±0.58	96.83±0.39	99.05±0.26	**99.84±0.16**	98.71±0.30
MT2	96.72±0.79	97.84±0.61	95.22±0.73	96.06±0.51	98.68±0.30	**100.00±0.00**	98.79±0.24
MT3	96.55±0.60	95.93±0.58	95.40±0.77	97.17±0.20	98.40±0.29	**99.89±0.09**	97.95±0.45
MT4	97.16±0.47	97.28±0.80	94.12±0.90	95.70±0.56	99.21±0.26	**99.70±0.13**	99.33±0.14
MT5	93.36±0.71	96.19±0.69	90.90±1.22	94.09±0.62	98.08±0.49	**100.00±0.00**	98.40±0.43
MT6	95.64±0.85	94.05±0.82	95.88±0.61	96.33±0.49	97.35±0.22	**99.78±0.15**	96.68±0.30
MT7	94.29±0.66	94.96±1.08	93.15±0.52	95.54±0.33	98.18±0.24	**98.51±0.15**	97.62±0.28
MT8	95.01±1.14	96.20±0.73	93.98±0.84	92.00±0.62	98.19±0.37	**98.93±0.18**	98.57±0.36
Avg	95.52	96.09	94.03	95.47	98.39	**99.58**	98.26
Time(s)	**145*3**	189*3	459*3	482*3	2006	764	861

Table 3
Composition strategy ablation diagnosis results.

	MDTN-CE	MDTN-RIM	MDTN-IST	MWSDTN
MT1	95.70±0.63	96.94±0.27	97.35±0.40	**98.71±0.30**
MT2	96.17±0.50	97.50±0.34	96.84±0.54	**98.79±0.24**
MT3	94.46±0.72	96.09±0.45	96.52±0.31	**97.95±0.45**
MT4	98.65±0.32	98.35±0.21	99.26±0.16	**99.33±0.14**
MT5	96.35±0.57	97.14±0.47	97.70±0.38	**98.40±0.43**
MT6	94.89±0.44	95.18±0.30	96.02±0.55	**96.68±0.30**
MT7	94.24±0.81	97.49±0.53	96.88±0.33	**97.62±0.28**
MT8	95.67±0.61	97.68±0.48	98.11±0.27	**98.57±0.36**
Avg	95.77	97.05	97.34	**98.26**

Additionally, the improvement in the performance of the source-free domain transfer model solely using the improved self-training mechanism is superior to that of the reinforced information maximization strategy.

5. Conclusion

This study presents a novel multi-source weighted source-free domain transfer method for fault diagnosis of rotating machinery. The proposed scheme effectively addresses challenges related to data privacy and domain transfer caused by limited labeled historical data. Specifically, our method eliminates the need for accessing source data during the training process of the unlabeled target domain. Instead, it employs a reinforced information maximization strategy and an improved self-training mechanism to facilitate knowledge transfer from the source to

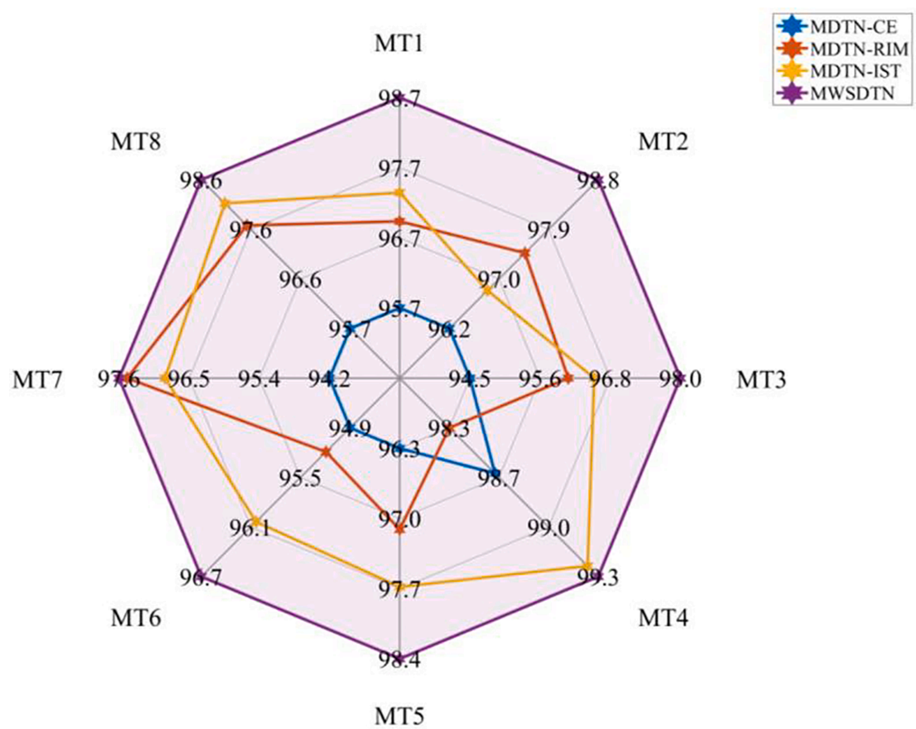

Fig. 6. The comparison results of all methods.

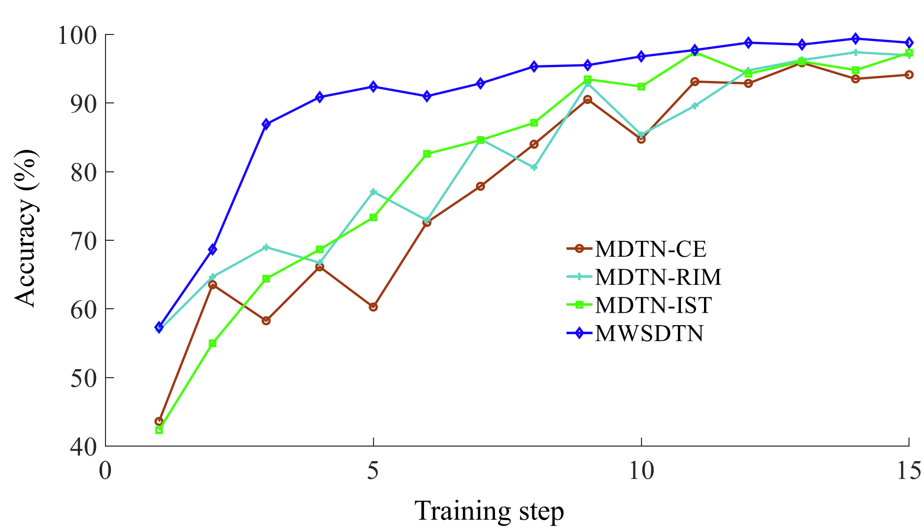

Fig. 7. Variation of test results for all methods in MT1.

the target. Moreover, a weighted strategy is introduced, automatically applying optimal values to all source domains based on their relevance to the target domain. Extensive experiments are conducted to evaluate the diagnostic performance of MWSDTN framework, comparing it with various advanced transfer learning methods. The results demonstrate the effectiveness and competitiveness of MWSDTN.

In the source-free domain transfer scenario, the quality of the initial source model becomes crucial due to the unavailability of source data. Additionally, real-world data, with its complex composition, presents even greater challenges for domain adaptation. To address these issues, future research endeavors will primarily center on exploring advanced strategies for enhancing the source model. This includes the application of cutting-edge strategies such as Nuclear-Norm Maximization to amplify the discriminative and diverse facets of target model predictions. The development of more robust self-training strategies, such as employing soft-label techniques to refine the accuracy of pseudo-labels in target domain data, is also a key focus. The amalgamation of continual learning algorithms will enable the source model to progressively adapt to novel information from the target domain over time, thus augmenting the model's generalization capacity. Furthermore, it is important to note that this study confines its attention solely to fault diagnosis in the context of cross-operational conditions, with uniform fault patterns across domains, which presents inherent limitations. In the future, efforts will be directed toward expanding the diagnostic scenarios of source-free transfer models, encompassing contexts such as cross-domain diagnosis between distinct devices and scenarios involving disparate fault patterns across domains.

CRediT authorship contribution statement

Qinhe Gao: Validation, Formal analysis. **Tong Huang:** Validation, Formal analysis. **Ke Zhao:** Conceptualization, Methodology, Supervision, Project administration. **Haidong Shao:** Funding acquisition, Formal analysis. **Bo Jin:** Formal analysis.

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.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (No. 51905160), the Natural Science Fund for Excellent Young Scholars of Hunan Province (No. 2021JJ20017) and the Fundamental Research Funds for the Central Universities, CHD (No. 300102253102).

References

Li, W., Zhong, X., Shao, H., et al. (2022). Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework[J]. *Advanced Engineering Informatics, 52*, Article 101552.

Wu, Z., Jiang, H., Zhu, H., et al. (2023a). A knowledge dynamic matching unit-guided multi-source domain adaptation network with attention mechanism for rolling bearing fault diagnosis[J]. *Mechanical Systems and Signal Processing, 189*, Article 110098.

Liu, Y., Jiang, H., Liu, C., Yang, W., & Sun, W. (2022). Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowledge-Based Systems, 252*, Article 109439.

Mohajer, A., Daliri, M. S., Mirzaei, A., et al. (2022a). Heterogeneous computational resource allocation for NOMA: Toward green mobile edge-computing systems[J]. *IEEE Transactions on Services Computing, 16*(2), 1225–1238.

Dong, S., Zhan, J., Hu, W., et al. (2023). Energy-efficient hierarchical resource allocation in uplink-downlink decoupled NOMA HetNets [J]. *IEEE Transactions on Network and Service Management*.

Hou, W., Zhang, C., Jiang, Y., et al. (2023). A new bearing fault diagnosis method via simulation data driving transfer learning without target fault data. *Measurement, 215*, Article 112879.

An, Z., Jiang, X., Yang, R., et al. (2022). Actively imaginative data augmentation for machinery diagnosis under large speed fluctuation conditions[J]. *IEEE Transactions on Industrial Informatics*.

Li, X., Xu, Y., Li, N., Yang, B., & Lei, Y. (2023a). Remaining Useful Life Prediction With Partial Sensor Malfunctions Using Deep Adversarial Networks. *IEEE/CAA Journal of Automatica Sinica, 10*(1), 121–134. https://doi.org/10.1109/JAS.2022.10593

Zhang, Z., Wang, J., Li, S., et al. (2023a). Fast nonlinear blind deconvolution for rotating machinery fault diagnosis [J]. *Mechanical Systems and Signal Processing, 187*, Article 109918.

Mohajer, A., Sorouri, F., Mirzaei, A., et al. (2022b). Energy-aware hierarchical resource management and backhaul traffic optimization in heterogeneous cellular networks [J]. *IEEE Systems Journal, 16*(4), 5188–5199.

Zhao, K., Jia, F., & Shao, H. (2023a). A novel conditional weighting transfer Wasserstein auto-encoder for rolling bearing fault diagnosis with multi-source domains. *Knowledge-Based Systems, 262*, Article 110203.

Chen, Z., Liao, Y., Li, J., et al. (2023). A Multi-Source Weighted Deep Transfer Network for Open-Set Fault Diagnosis of Rotary Machinery. *IEEE Transactions on Cybernetics, 53*(3), 1982–1993. https://doi.org/10.1109/TCYB.2022.3195355

Xiao, Y., Shao, H., Han, S., et al. (2022). Novel joint transfer network for unsupervised bearing fault diagnosis from simulation domain to experimental domain[J]. *IEEE-ASME Transactions on Mechatronics*. https://doi.org/10.1109/TMECH.2022.3177174

Wang, R., Yan, F., Yu, L., et al. (2023a). A federated transfer learning method with low-quality knowledge filtering and dynamic model aggregation for rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing, 198*, Article 110413.

Liu, G., Shen, W., Gao, L., et al. (2023). Active federated transfer algorithm based on broad learning for fault diagnosis. *Measurement, 208*, Article 112452.

Liang J, Hu D, Feng J. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation[C]//International Conference on Machine Learning. PMLR, 2020: 6028-6039.

Zhu, J., Chen, N., & Shen, C. (2019). A New Deep Transfer Learning Method for Bearing Fault Diagnosis Under Different Working Conditions. *IEEE Sensors Journal, 20*(15), 8394–8402.

Mao, W., Liu, Y., Ding, L., Safian, A., & Liang, X. (2021). A New Structured Domain Adversarial Neural Network for Transfer Fault Diagnosis of Rolling Bearings Under Different Working Conditions. *IEEE Transactions on Instrumentation and Measurement, 70*, 1–13.

Zhao, K., Jiang, H., Wang, K., et al. (2021). Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis. *Knowledge-Based Systems, 222*, Article 106974.

Zhao, C., Liu, G., & Shen, W. (2022). A balanced and weighted alignment network for partial transfer fault diagnosis[J]. *ISA transactions, 130*, 449–462.

Zhang, X., He, L., Wang, X., et al. (2022a). Transfer fault diagnosis based on local maximum mean difference and K-means[J]. *Computers & Industrial Engineering, 172*, Article 108568.

Zhou, Y., Dong, Y., Zhou, H., et al. (2021). Deep dynamic adaptive transfer network for rolling bearing fault diagnosis with considering cross-machine instance[J]. *IEEE Transactions on Instrumentation and Measurement, 70*, 1–11.

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

Chen, Z., Liao, Y., Li, J., et al. (2022a). A multi-source weighted deep transfer network for open-set fault diagnosis of rotary machinery[J]. *IEEE Transactions on Cybernetics, 53*(3), 1982–1993.

Tian, J., Han, D., Li, M., et al. (2022). *A multi-source information transfer learning method with subdomain adaptation for cross-domain fault diagnosis, 243*, Article 108466.

Cao, W., Meng, Z., Sun, D., et al. (2023). A two-stage domain alignment method for multi-source domain fault diagnosis. *Measurement, 214*, Article 112818.

Wu, Z., Jiang, H., Liu, S., et al. (2023b). Conditional distribution-guided adversarial transfer learning network with multi-source domains for rolling bearing fault diagnosis[J]. *Advanced Engineering Informatics, 56*, Article 101993.

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

Xia, L., Zheng, P., Li, J., et al. (2022). Privacy-preserving gradient boosting tree: Vertical federated learning for collaborative bearing fault diagnosis. *IET Collaborative Intelligent Manufacturing, 4*(3), 208–219.

Wang, Y., Yan, J., Yang, Z., et al. (2022a). A Novel Federated Transfer Learning Framework for Intelligent Diagnosis of Insulation Defects in Gas-Insulated Switchgear. *IEEE Transactions on Instrumentation and Measurement, 71*, 1–11.

Zhang, W., & Li, X. (2022a). Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions. *Structural Health Monitoring, 21*(4), 1329–1344.

Zhang, W., & Li, X. (2022b). Federated Transfer Learning for Intelligent Fault Diagnostics Using Deep Adversarial Networks With Data Privacy. *IEEE/ASME Transactions on Mechatronics, 27*(1), 430–439.

Zhao, K., Hu, J., Shao, H., & Hu, J. (2023b). Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy. *Reliability Engineering & System Safety, 236*, Article 109246.

Chen, J., Li, J., Huang, R., et al. (2022b). Federated Transfer Learning for Bearing Fault Diagnosis With Discrepancy-Based Weighted Federated Averaging. *IEEE Transactions on Instrumentation and Measurement, 71*, 1–11.

S. Kang, J. Yang, Y. Sun, Y. Wang, Q. Wang and V. I. Mikulovich, “Fault Diagnosis Method of Rolling Bearings Under Different Working Conditions Based on Federated Feature Transfer Learning,” 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Harbin, China, 2022, pp. 1-6, doi: 10.1109/ICSMD57530.2022.10058221.

Wang, R., Yan, F., Yu, L., et al. (2023b). A federated transfer learning method with low-quality knowledge filtering and dynamic model aggregation for rolling bearing fault diagnosis[J]. *Mechanical Systems and Signal Processing, 198*, Article 110413.

Lu, S., Gao, Z., Xu, Q., et al. (2022). Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication[J]. *IEEE Transactions on Industrial Informatics, 18*(12), 9101–9111.

Li, X., Zhang, C., Li, X., et al. (2023b). Federated transfer learning in fault diagnosis under data privacy with target self-adaptation[J]. *Journal of Manufacturing Systems, 68*, 523–535.

Wang, R., Huang, W., Shi, M., et al. (2022b). Federated adversarial domain generalization network: A novel machinery fault diagnosis method with data privacy [J]. *Knowledge-Based Systems, 256*, Article 109880.

J. Jiao, H. Li, T. Zhang, *et al.* Source-Free Adaptation Diagnosis for Rotating Machinery, IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2022.3231414.

M. Zhu, X. Zeng, J. Liu, *et al.* Source-free Unsupervised Domain Adaptation for Privacy-Preserving Intelligent Fault Diagnosis 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Harbin, China, 2022, pp. 1-6, doi: 10.1109/ICSMD57530.2022.10058437.

Li, J., Yue, K., Huang, R., et al. (2023c). Unsupervised Continual Source-Free Network for Fault Diagnosis of Machines Under Multiple Diagnostic Domains. *IEEE Sensors Journal, 23*(8), 8292–8303.

Wang, J., Zhang, H., & Miao, Q. (2023c). Source free unsupervised domain adaptation for electro-mechanical actuator fault diagnosis. *Chinese Journal of Aeronautics, 36*(4), 252–267.

K. Yue, W. Li, J. Li, *et al.* Source-free Partial Domain Adaptation in Rotating Machinery Fault Diagnosis using Deep Hypothesis Domain Adaptation Network, 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Harbin, China. 2022. 1-6. doi: 10.1109/ICSMD57530.2022.10058429.

J. Zhang, Z. Jiang, L. Zhang. Source-Free Domain Adaptation for Rotating Machinery Cross-Domain Fault Diagnosis with Neighborhood Reciprocity Clustering, 2022 China Automation Congress (CAC), Xiamen, China. 2022. 5333-5340. doi: 10.1109/CAC57257.2022.10055182.

Zhang, Y., Ren, Z., Feng, K., et al. (2023b). Universal source-free domain adaptation method for cross-domain fault diagnosis of machines[J]. *Mechanical Systems and Signal Processing, 191*, Article 110159.

Jia, M., Wang, J., Zhang, Z., et al. (2022). A novel method for diagnosing bearing transfer faults based on a maximum mean discrepancies guided domain-adversarial mechanism. *Measurement Science and Technology, 33*, Article 015109.

Lessmeier C. Kimotho J. 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. European Conference of the Prognostics and Health Management Society, Bilbao (Spain), 2016. [Online]. Available: mb.uni-paderborn.de/kat/datacenter.

Sun Y, Chong N, Ochiai H. Feature distribution matching for federated domain generalization[C]//Asian Conference on Machine Learning. PMLR. 2023. 942-957.