EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Bearing Cross-Domain Fault Diagnosis Based on Wavelet Time-Frequency Map and Resnet Transfer Learning in Industrial IoT-Enabled Environments

Wanchuan Wang [1, *]

[1]School of Information Engineering, Yangzhou Polytechnic Institute, China

Abstract

To overcome the difficulties faced by current bearing fault diagnosis methods in industrial applications, such as the difficulty in obtaining sufficient labeled real fault training samples and the subpar performance of cross-domain fault diagnosis due to data distribution deviation, a new bearing cross-domain fault diagnosis method based on wavelet time-frequency maps and Resnet transfer learning (CDF-WTI-RTL) is proposed.In the context of the Industrial Internet of Things (IIoT) and Industry 5.0, real-time condition monitoring and intelligent predictive maintenance have become critical for ensuring manufacturing system resilience and sustainability. First, raw vibration signals are processed using continuous wavelet transform to obtain two-dimensional time-frequency maps. Second, a Resnet network adapted for bearing fault diagnosis is constructed to autonomously mine deep features from source and target domains. A random forest classifier is then employed to build a cross-domain fault pattern recognition classifier. Experimental analysis of training samples using two different bearing fault datasets shows that the accuracy of the CDF-WTI-RTL model reaches 98.38% and 91.5% with unbalanced training samples respectively, notably surpassing the other 5 comparative models. The results demonstrate that the proposed method can be effectively deployed in IIoT architectures for reliable, cross-condition bearing fault diagnosis.

Keywords: Fault diagnosis; time-frequency image; Resnet; transfer learning; domain adaptation; Internet of Things (IIoT)

Received on 15 June 2026, accepted on 16 July 2026, published on 29 July 2026

doi: 10.4108/eetsis.13528

1. Introduction

1.1. Research background

Rolling bearings are commonly found in industrial rotating machinery and are prone to wear or failure. Research into high-precision, highly adaptable intelligent fault diagnosis methods is of significant importance and economic value for ensuring the safe and stable long-term operation of rotating machinery and equipment [1-3].
The rapid advancement of the Industrial Internet of Things (IIoT) has transformed traditional manufacturing by enabling pervasive sensing and edge-cloud collaborative computing[4,5]. In parallel, the emerging Industry 5.0 paradigm emphasizes human-centricity, sustainability, and resilience, where predictive maintenance plays a pivotal role in minimizing unplanned downtime and extending equipment lifecycle [6,7]. Within this context, bearing fault diagnosis systems are expected to not only achieve high accuracy but also exhibit adaptability across diverse operating conditions with minimal labeled data [6].

1.2. Literature review and issue

Although the research findings in the aforementioned literature provide solutions for achieving ideal Bearings fault diagnosis in cross-domain scenarios, challenges remain[8,10]: Firstly, sufficient labeled fault samples are difficult to obtain; Secondly, models perform poorly with imbalanced training samples; Furthermore, complex working conditions cause data distribution discrepancies.
To tackle these problems, this paper introduces a cross-domain fault diagnosis method for bearings using wavelet time-frequency maps and Resnet transfer learning (CDF-WTI-RTL). The method consists of four steps:

*Corresponding author. Email: 252617256@qq.com

①Vibration signal time-frequency map extraction via continuous wavelet transform; ②Deep feature extraction using transfer ResNet; ③Deep feature distribution adaptation based on Manifold Embedded Transfer Joint Match (METJM); ④Random forest classifier training for cross-domain fault diagnosis.

2. Deep Feature Extraction Using Wavelet Time-Frequency Map and Transfer Deep Residual Network

2.1. Deep Residual Network

In recent years, Classical Convolutional Neural Networks (CNN) face gradient issues as network depth increases [9]. The Deep Residual Network (ResNet) [10] adopts a residual learning framework to overcome this. In this paper, ResNet-18 is employed as the deep feature extractor[11] ,thereby improving fault diagnosis accuracy.

2.2. Deep Feature Extraction Based on Transfer Deep Residual Network

In this study, The ResNet-18 network is pre-trained using the ImageNet database. Following model parameter transfer, a transfer ResNet network is constructed. The source domain wavelet time-frequency maps fine-tune the network, resulting in a model adapted to bearing fault data. The source and target domain wavelet time-frequency maps are then input into the model, and the feature data output from the fully connected layer is used to construct the deep feature set for subsequent feature distribution adaptation.

In IIoT-enabled manufacturing environments, vibration sensors continuously stream high-dimensional time-series data to edge devices or cloud platforms [4,5]. The deployment of transfer ResNet models on such IIoT architectures enables on-device feature extraction, reducing the need for raw data transmission and alleviating network bandwidth pressure [6], aligning with Industry 5.0 predictive maintenance requirements [7].

3. Deep Feature Distribution Adaptation Based on Manifold Embedded Transfer Joint Matching

In deep learning models, extracted deep features exhibit distribution discrepancies that affect cross-domain fault diagnosis accuracy. This paper introduces a domain adaptation approach grounded in Manifold Embedded Transfer Joint Matching (METJM). METJM achieves distribution alignment of deep features between the source and target domains in the manifold space and effectively improves feature separability,which is then used for the next step in cross-domain fault diagnosis model training and testing.

$D_S=\{X_S,Y_S\}$ and $D_T=\{X_T\}$ stand for the labeled source domain feature samples and unlabeled target domain feature samples, respectively. D_S is gathered under specific conditions, whereas D_T is collected under different conditions, $X_S=\{x_i^S\}_{i=1}^{n_S}$ denotes a dataset containing n_S samples, where the i-th sample is labeled as $x_i^S \in \mathbb{R}^{1\times d}$ ($i=1,2,\cdots n_S$). The corresponding class label set of X_S is $Y_S=\{y_i^S\}_{i=1}^{n_S}$, where $y_i^S \in \mathbb{R}^{n_S\times 1}$ is the label of the i-th sample. $X_T=\{x_j^T\}_{j=1}^{n_T}$ denotes a dataset containing n_T samples, where $x_j^T \in \mathbb{R}^{1\times d}$ is the j-th sample, ($j=1,2,\cdots n_T$).

3.1. Transfer Joint Matching (TJM)

Transfer Joint Matching (TJM) [12] is a method that reduces the differences between the source and target domains by jointly matching features and re-weighting cross-domain instances to obtain a new low-dimensional domain-invariant feature space. The objectives of optimizing TJM comprise three components:

(i) Feature Dimensionality Reduction.TJM first uses classical Kernel Principal Component Analysis (KPCA) to reconstruct the input feature set. The goal of optimization is as described below:

$$\max_{\mathrm{A^T A=I}} tr(\mathrm{A^T KHK^T A}) \tag{1}$$

$\mathrm{A} \in \mathbb{R}^{n\times k}$ is the mapping transformation matrix of KPCA, k is the dimensionality of the reduced features, $\mathbf{K}=\phi(\mathbf{X})^{\mathrm{T}}\phi(\mathbf{X}) \in \mathbb{R}^{n\times n}$ is the kernel matrix, ϕ is the kernel mapping function, and $\mathbf{H}=\mathbf{I}\text{-}(1/n)\mathbf{1}$ is the centering matrix.

(ii) Feature Matching.TJM uses the classical Maximum Mean Discrepancy (MMD) method, aiming to minimize MMD as the optimization goal. This part is similar to the classical feature transfer learning method, Transfer Component Analysis (TCA) [13]. The optimization goal is provided as:

$$\min_{\mathbf{A^T KHK^T A=I}} \left\| \frac{1}{n_S}\sum_{i=1}^{n_S}\mathbf{A}^{\mathrm{T}}k_i - \frac{1}{n_T}\sum_{j=n_S+1}^{n_S+n_T}\mathbf{A}^{\mathrm{T}}k_j \right\|_{\mathrm{H}}^2 = \min_{\mathbf{A^T KHK^T A=I}} tr(\mathbf{A}^T\mathbf{KMK}^T\mathbf{A}) \tag{2}$$

where $k_i \in \mathbf{K}$, H is the transformation matrix in the Reproducing Kernel Hilbert Space (RKHS), and $\mathbf{M}$ is the MMD matrix for measuring the differences in marginal probability distributions. The expression of $\mathbf{M}$ is:

$$\mathbf{M}=\begin{cases} 1/n_S^2 & (x_i,x_j \in D_S) \\ 1/n_T^2 & (x_i,x_j \notin D_S) \\ -1/n_S n_T & (otherwise) \end{cases} \tag{3}$$

(iii) Instance Re-weighting.TJM integrates the instance re-weighting process with feature matching to further improve domain adaptation. TJM introduces an -

norm structured sparse regularization term into the mapping transformation matrix A. Its expression is:

$$\|A_S\|_{2,1} + \|A_T\|_F^2 \tag{4}$$

Where A_S and A_T are the mapping transformation matrices associated with the instances from the source and target domains. By applying the $l_{2,1}$--norm regularization term to the source domain instance samples, TJM re-weights the source domain instances relevant to the target domain. By minimizing Equation (4) and maximizing Equation (1), the source domain instances relevant (or irrelevant) to the target domain instances are adaptively re-weighted with greater (or lesser) importance in the new representation, further enhancing domain adaptation performance.

Based on the three optimization objectives of TJM, the overall optimization problem is defined as follows:

$$\min_{\mathbf{A}^T\mathbf{KHK}^T\mathbf{A}=\mathbf{I}} tr(\mathbf{A}^T\mathbf{KMK}^T\mathbf{A}) + \lambda(\|A_S\|_{2,1} + \|A_T\|_F^2) \tag{5}$$

λ serves as a regularization parameter to balance feature matching and instance re-weighting.

3.2. Improved Transfer Joint Matching with Integrated Neighborhood Preserving Embedding and Feature Separability Enhancement

TJM still has certain limitations: ①Feature data distribution alignment is carried out exclusively in the initial feature space, neglecting potential feature distortion in the high-dimensional space. ②Feature distribution alignment considers only the marginal probability distribution, ignoring the conditional probability distribution among various domain feature datasets. ③It does not focus on improving the discriminative performance of feature samples during domain adaptation and fails to fully leverage the neighborhood relationships among feature samples. To tackle these constraints, this section proposes a new improved Transfer Joint Matching method that integrates Neighborhood Preserving Embedding (NPE), adaptive alignment of feature probability distributions, and class separability enhancement. The optimization objectives of this method include four parts:

(1) Manifold Subspace Learning Based on Neighborhood Preserving Embedding(NPE). To overcome the feature distortion issues encountered in adapting feature data distributions within the initial high-dimensional space of feature, this study employs Neighborhood Preserving Embedding (NPE) to learn a manifold subspace that maintains data neighborhood relationships from the initial high-dimensional space of feature. The operational process is shown in Figure 1 [14].

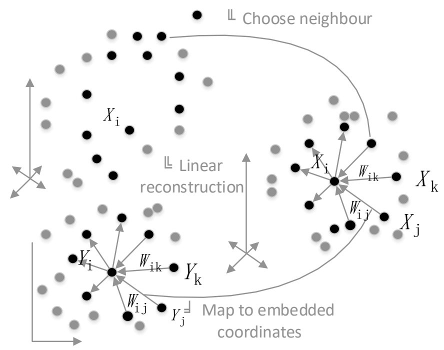

Figure 1. Schematic Diagram of NPE Algorithm Process

(i)Selecting Neighbors and Constructing a Neighborhood Graph G: Utilizing the K-nearest neighbors (KNN) method, the data sample point x_i is associated with the i-th node. The distance between x_i and its adjacent node x_j is calculated based on the Euclidean distance. If the distance falls within the neighborhood range, a connection is created between x_i and x_j, forming the neighborhood graph G.

(ii)Linear Reconstruction and Calculating Reconstruction Weights Matrix **W**. The elements w_{ij} of the weight matrix represent the weight of the edge from node i to node j. The reconstruction loss function of NPE is:

$$\varphi(W) = \sum_i \left\| x_i - \sum_j w_{ij} x_j \right\|^2 \tag{6}$$

The weight matrix W is obtained by minimizing the reconstruction loss function as:

$$\begin{cases} \varphi(W) = \min(\sum_i \left\| x_i - \sum_j w_{ij} x_j \right\|^2) \\ \sum_i w_{ij} = 1, j = 1, 2, \cdots, N \end{cases} \tag{7}$$

By substituting $y_i = O^T x_i$ into loss function, the solution can further be transformed into:

$$P(A) = \min(tr(O^T XZX^T O)) \tag{8}$$

where **O** represents the feature matrix, and $Z = (I - W)^T (I - W)$, $I = diag(1, \cdots, 1)$.

(iii) Mapping to Embedding Coordinates. To further solve the mapping matrix **O**, the optimization issue presented in Equation (8) can be converted into a generalized eigenvector problem for minimizing eigenvalues:

$$XZX^T a = \lambda XX^T a \tag{9}$$

(2) Adaptive Alignment of Feature Probability Distributions. An improved feature matching method adaptively balances marginal and conditional probability distributions using an A-distance metric, based on Equation (2):

$$\min_{\mathbf{A}^T\mathbf{KHK}^T\mathbf{A}=\mathbf{I}} \mu tr(\mathbf{A}^T\mathbf{KMK}^T\mathbf{A})+(1\text{-}\mu)\sum_{c=1}^{C} tr(\mathbf{A}^T\mathbf{KM}_c\mathbf{K}^T\mathbf{A}) \tag{10}$$

Where μ is an adaptive factor that adjusts the balance between $tr(\mathbf{A}^T\mathbf{KMK}^T\mathbf{A})$ and $\sum_{c=1}^{C} tr(\mathbf{A}^T\mathbf{KM}_c\mathbf{K}^T\mathbf{A})$, $\mu \in [0,1]$. The adaptive factor is determined using the A-distance metric for optimal alignment. $\sum_{c=1}^{C} tr(\mathbf{A}^T\mathbf{KM}_c\mathbf{K}^T\mathbf{A})$ represents the conditional probability distribution MMD between the source and target domain feature samples. Let C be the overall number of classes, $c \in [1, C]$; then:

$$\mathbf{M}_c = \begin{cases} \dfrac{1}{n_S^{(c)} n_S^{(c)}} & (x_i, x_j \in D_S^{(c)}) \\ \dfrac{1}{n_T^{(c)} n_T^{(c)}} & (x_i, x_j \in D_T^{(c)}) \\ \dfrac{-1}{n_S^{(c)} n_T^{(c)}} & \begin{cases} x_i \in D_S^{(c)}, x_j \in D_T^{(c)} \\ x_i \in D_T^{(c)}, x_j \in D_S^{(c)} \end{cases} \\ 0 & (otherwise) \end{cases} \tag{11}$$

Where $n_S^{(c)}$ and $n_T^{(c)}$ are the numbers of source and target domain feature samples in class c, and $D_S^{(c)}$ and $D_T^{(c)}$ are the feature collections of class c in both the source and target domains.

(3) Instance Re-weighting. This part retains the instance re-weighting objective from TJM to further enhance domain adaptation performance.

(4) Enhancement of Class Separability. To improve the discriminative performance of feature samples, which is often overlooked in domain adaptation methods, and to leverage neighborhood relationships among feature samples [15], an optimization objective for enhancing class separability is introduced:

$$\min_{\mathbf{A}^T\mathbf{KHK}^T\mathbf{A}=\mathbf{I}} tr(\mathbf{A}^T\mathbf{S}_w\mathbf{A})) \tag{12}$$

Here $\mathbf{S}_w$ represents the local scatter matrix within each class, defined as:

$$\mathbf{S}_w = \frac{1}{2}\sum_{i,j=1}^{n} p_{ij}^{w}(x_i - x_j)(x_i - x_j)^T \tag{13}$$

p_{ij}^{w} is a weight factor given by:

$$p_{ij}^{w} = \begin{cases} A_{ij}/n_c, & y_i = y_j = c \\ 0, & y_i \neq y_j \end{cases} \tag{14}$$

Where n signifies the total number of feature samples, whereas n_c represents the number of samples in class c. A_{ij} is a tuning parameter. The closer x_i and x_j are, the larger A_{ij} is, which reduces the distance between intra-class samples. This improves the separability of different class feature samples, thereby enhancing the fault mode recognition accuracy.

Based on the above four optimization objectives, the improved TJM optimization objective is defined as:

$$\begin{aligned} &\min_{\mathbf{A}^T\mathbf{KHK}^T\mathbf{A}=\mathbf{I}} \mu tr(\mathbf{A}^T\mathbf{KMK}^T\mathbf{A})+(1\text{-}\mu)\sum_{c=1}^{C} tr(\mathbf{A}^T\mathbf{KM}_c\mathbf{K}^T\mathbf{A}) \\ &+tr(\mathbf{A}^T\mathbf{S}_w\mathbf{A}))+\lambda(\|A_S\|_{2,1}+\|A_T\|_F^2) \end{aligned} \tag{15}$$

Where λ are regularization parameters that balance feature matching, instance re-weighting, and class separability.

Steps for Improved TJM Process are as follow:

①Perform manifold subspace learning on the source domain feature data $\mathbf{D}_S$ using neighborhood preserving embedding. Given the manifold subspace dimensionality parameter, obtain the new source domain feature subset $\mathbf{Z}_S$ and mapping matrix $\mathbf{O}$.

②Use the mapping matrix $\mathbf{O}$ obtained in Step 1 to alter the target domain's feature data $\mathbf{D}_T$, obtaining a novel target domain feature subset $\mathbf{Z}_T$.

③Based on the transformed $\mathbf{Z}_S$ and $\mathbf{Z}_T$, compute the MMD matrices $\mathbf{M}_0$ and $\mathbf{M}_c$ using Equations (3) and (11).

④Solve the optimization problem in Equation (15) to obtain the mapping transformation matrix $\mathbf{A}$. Use $\mathbf{A}$ to transform $\mathbf{Z}_S$ and $\mathbf{Z}_T$, obtaining the domain-adapted source and target domain feature sets $\mathbf{Z}'_S$ and $\mathbf{Z}'_T$.

Finally, the labeled $\mathbf{Z}'_S$ is employed in training a random forest classifier, and the unlabeled $\mathbf{Z}'_T$ is used for fault mode recognition and classification, yielding the fault diagnosis results.

From an IIoT perspective, the proposed METJM-based domain adaptation module can be integrated as a middleware service in the cloud layer, processing features extracted from distributed edge nodes [5]. This architecture supports centralized model updating and facilitates knowledge sharing across different production lines or even factories, thereby enhancing the generalization capability of fault diagnosis models in heterogeneous industrial environments [4,6].

4. Experimental Validation

Experiments were conducted using the Case Western Reserve University (CWRU) bearing fault dataset [3,8] . Comparative models were constructed based on classical feature transfer learning methods [13], including Transfer Component Analysis (TCA), Joint Distribution Adaptation (JDA), Balanced Distribution Adaptation (BDA), Geodesic Flow Kernel (GFK), and Transfer Joint Matching (TJM). A comparison was made between these models and the proposed method.

4.1. Experiment Analysis Based on the CWRU Bearing Fault Dataset

4.1.1. Experimental Data and Task Setup

This section validates the proposed method utilizing the bearing fault dataset of CWRU. The dataset contains bearing vibration signals collected under four motor speeds and ten states, labeled as 1-10, which is described in Table 1. For each category of bearing vibration data, 300 samples were randomly selected. Within this collection, 270 samples were allocated for training, and the leftover 30 were designated for testing. The 270 training samples were used to fine-tune the transfer ResNet network. With this dataset as the foundation, two types of tasks for cross-domain fault diagnosis were created: fault diagnosis under

balanced training samples and unbalanced training samples, which is displayed in Tables 2 and 3.
In this experimental analysis, unbalanced training sample tasks were designed to simulate real-world industrial scenarios with insufficient labeled fault samples. This served to confirm the practical potential of the proposed method for industrial applications. As shown in Table 2, the quantity of training samples for each fault category (categories 1-10) in the unbalanced training data task were: 260, 250, 240, 230, 220, 210, 200, 190, 180, and 170, respectively. Thus, the total number of training samples was 2150.

Table 1. Introduction to the CWRU Bearing Fault Dataset

Bearing Condition	Defect Size (cm)	Number of Training/Testing Samples at Different Motor Speeds				Class Label
		1797r/min	1772 r/min	1750 r/min	1730 r/min	
Normal	0	270/30	270/30	270/30	270/30	1
Inner Race Fault	0.01778	270/30	270/30	270/30	270/30	2
	0.03556	270/30	270/30	270/30	270/30	3
	0.05334	270/30	270/30	270/30	270/30	4
Outer Race Fault	0.01778	270/30	270/30	270/30	270/30	5
	0.03556	270/30	270/30	270/30	270/30	6
	0.05334	270/30	270/30	270/30	270/30	7
Ball Fault	0.01778	270/30	270/30	270/30	270/30	8
	0.03556	270/30	270/30	270/30	270/30	9
	0.05334	270/30	270/30	270/30	270/30	10

Table 2. Cross-Domain Fault Diagnosis Tasks (Balanced Training Samples)

Task	Source Domain			Target Domain		
	Motor Speed (r/min)	Fault Types	Number of Training Samples	Motor Speed (r/min)	Fault Types	Number of Training Samples
1	1797	10	2700	1772	10	300
2	1797	10	2700	1750	10	300
3	1797	10	2700	1730	10	300
4	1772	10	2700	1797	10	300
5	1772	10	2700	1750	10	300
6	1772	10	2700	1730	10	300
7	1750	10	2700	1797	10	300
8	1750	10	2700	1772	10	300
9	1750	10	2700	1730	10	300
10	1730	10	2700	1797	10	300
11	1730	10	2700	1772	10	300
12	1730	10	2700	1750	10	300

Table 3. Cross-Domain Fault Diagnosis Tasks (Unbalanced Training Samples)

Task	Source Domain			Target Domain		
	Motor Speed	Fault Types	Number of Training Samples	Motor Speed	Fault Types	Number of Training Samples
13	1797	10	2150	1772	10	300
14	1797	10	2150	1750	10	300
15	1797	10	2150	1730	10	300
16	1772	10	2150	1797	10	300
17	1772	10	2150	1750	10	300
18	1772	10	2150	1730	10	300
19	1750	10	2150	1797	10	300
20	1750	10	2150	1772	10	300
21	1750	10	2150	1730	10	300
22	1730	10	2150	1797	10	300
23	1730	10	2150	1772	10	300
24	1730	10	2150	1750	10	300

4.1.2. Experimental Results and Analysis

(1) Effectiveness of the Proposed Method

Table 4 displays the accuracy of the proposed CDF-WTI-RTL model in diagnosing faults for 24 cross-domain tasks. For the 12 tasks with balanced training samples, the average accuracy reached 99.63%, with tasks 2, 4, 5, 7, 9, 11, and 12 achieving 100% accuracy. For the 12 tasks with unbalanced training samples, the average accuracy was 98.38%, with tasks 19 and 21 achieving 100% accuracy. The accuracy results for all 24 tasks validate the effectiveness of the introduced CDF-WTI-RTL model. Additionally, the difference in accuracy between balanced and unbalanced training sample tasks highlights the challenge posed by insufficient fault data leading to unbalanced training samples, which negatively impacts the model's capacity to achieve ideal performance and results in a drop in accuracy.

Table 4. Fault Diagnosis Accuracy of CDF-WTI-RTL Model Across Tasks

Task	Accuracy (%)	Task	Accuracy (%)
1	99.50	13	98.50
2	100.00	14	99.50
3	98.50	15	98.50
4	100.00	16	99.50
5	100.00	17	97.50
6	98.50	18	94.50
7	100.00	19	100.00
8	99.50	20	98.50
9	100.00	21	100.00
10	99.50	22	97.50
11	100.00	23	98.50
12	100.00	24	98.00
Avg. Accuracy	99.63	Avg. Accuracy	98.38

(2) Superiority of the Proposed Method

To additionally verify the advantage of the introduced CDF-WTI-RTL approach in achieving ideal cross-domain fault diagnosis performance, comparative models were constructed based on classical methods for feature transfer learning, comprising TCA, JDA, BDA, GFK, as well as TJM. These models are described in Table 5. Each comparative model combines the source and target domain deep feature sets (DFS) extracted by the transfer ResNet network with a particular method for transfer learning. For instance, the DFS-TCA model inputs the labeled source domain DFS and unlabeled target domain DFS into TCA, outputs domain-adapted feature sets, and trains a random forest classifier which makes use of the labeled feature set belonging to the source domain. The trained classifier is then employed for classifying the feature set of the unlabeled target domain, producing fault diagnosis accuracy. Similar processes were followed for DFS-JDA, DFS-BDA, DFS-GFK, and DFS-TJM.

Table 5. Comparative Fault Diagnosis Models

Label	M1	M2	M3	M4	M5	M6
Fault Diagnosis Model	DFS-TCA	DFS-JDA	DFS-BDA	DFS-GFK	DFS-TJM	CDF-WTI-RTL

The comparative experimental results shown in Tables 6 and 7 indicate that:

(1) The proposed CDF-WTI-RTL model achieves significantly better fault diagnosis accuracy compared to other models. For tasks with balanced training samples (Tasks 1-12), all models achieved at least 90% accuracy. However, the accuracy of DFS-JDA and CDF-WTI-RTL models was notably higher than others. The CDF-WTI-RTL model achieved an average fault diagnosis accuracy of 99.63%, outperforming DFS-TCA, DFS-JDA, DFS-BDA, DFS-GFK, and DFS-TJM by 7.38%, 2.50%, 4.13%, 5.13%, and 4.34%, respectively.

(2) For tasks with unbalanced training samples (Tasks 13-24), the proposed model was the only one to achieve satisfactory cross-domain fault diagnosis accuracy, while all other models showed significant degradation. The average accuracy of the CDF-WTI-RTL model was 98.38%, surpassing DFS-TCA, DFS-JDA, DFS-BDA, DFS-GFK, and DFS-TJM by 10.05%, 7.50%, 8.34%, 10.30%, and 6.67%, respectively. Figures 2 and 3 illustrate the average fault diagnosis accuracy of comparative models across Tasks 1-12 and Tasks 13-24, respectively.

(3) Comparing the experimental results of DFS-TJM and CDF-WTI-RTL demonstrates that the improvements proposed in this study, based on TJM, enhance domain adaptation capability and improve the discriminative

power of feature data, thereby significantly increasing fault diagnosis accuracy.

Table 6. Fault Diagnosis Accuracy (Tasks 1-12)

Task	Cross-Domain Fault Diagnosis Accuracy (%)					
	M1	M2	M3	M4	M5	M6
1	95.00	96.50	95.50	93.50	97.50	99.50
2	92.50	97.00	93.00	92.00	95.50	100.00
3	83.00	84.50	80.00	85.50	86.50	98.50
4	97.50	98.50	98.50	96.50	99.50	100.00
5	99.50	100.00	99.50	98.50	99.50	100.00
6	87.00	99.50	100.00	99.50	98.50	98.50
7	96.50	97.50	99.50	96.50	95.50	100.00
8	92.00	98.00	95.00	92.50	93.00	99.50
9	94.50	100.00	98.50	97.50	95.50	100.00
10	83.00	96.50	92.00	91.00	94.50	99.50
11	88.00	97.50	95.50	86.50	92.50	100.00
12	98.50	100.00	99.00	97.50	95.50	100.00
Avg. Accuracy	92.25	97.13	95.50	94.50	95.29	99.63

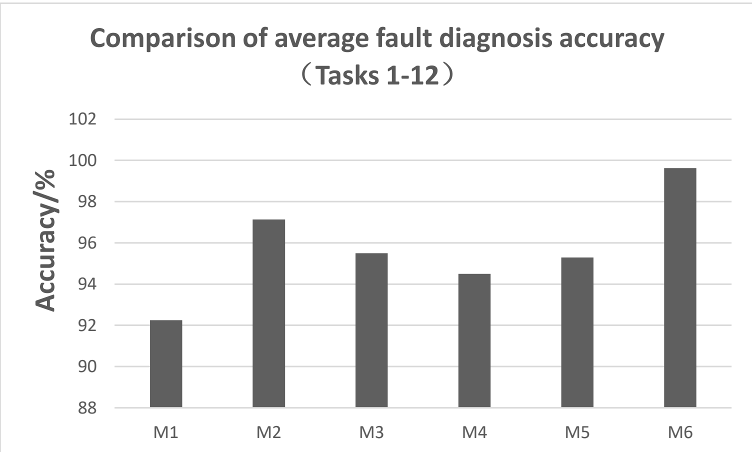

Figure 2. Comparison of Average Accuracy for Cross-Domain Fault Diagnosis (Tasks 1-12)

Table 7. Fault Diagnosis Accuracy (Tasks 13-24)

Task	Cross-Domain Fault Diagnosis Accuracy (%)					
	M1	M2	M3	M4	M5	M6
13	96.50	95.00	92.50	91.00	96.50	98.50
14	92.00	83.00	85.00	82.50	91.50	99.50
15	83.50	81.50	78.50	74.50	85.00	98.50
16	96.50	97.50	96.50	93.50	97.50	99.50

17	95.50	99.00	98.50	97.50	96.50	97.50
18	92.00	100.00	99.50	95.00	94.50	94.50
19	93.50	92.50	91.00	88.50	93.00	100.00
20	88.50	93.50	92.50	91.00	96.50	98.50
21	87.50	97.50	98.00	95.50	96.50	100.00
22	74.50	81.50	83.00	84.50	83.50	97.50
23	75.50	83.00	80.50	78.50	85.00	98.50
24	84.50	86.50	85.00	85.00	84.50	98.00
Avg. Accuracy	88.33	90.88	90.04	88.08	91.71	98.38

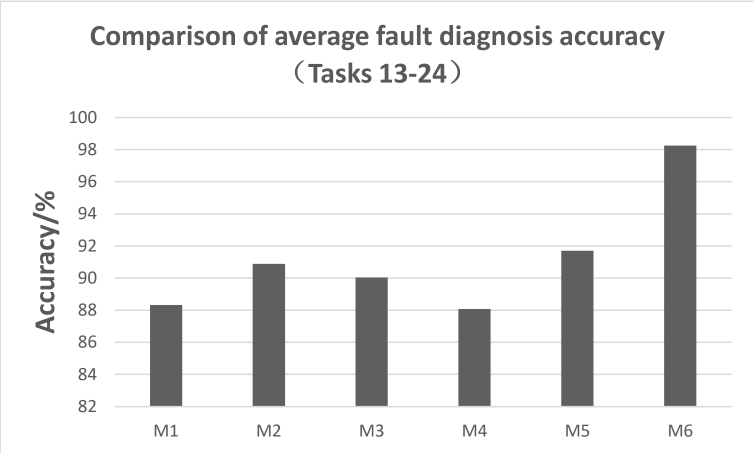

Figure 3. Comparison of Average Accuracy for Cross-Domain Fault Diagnosis (Tasks 13-24)

4.2. Experimental Analysis Based on the SQ-MFS Mechanical Fault Test Bench Bearing Fault Dataset

4.2.1. Experimental Data and Task Setup

For further verification of the proposed method's adaptability, the SQ-MFS mechanical fault test bench [3] bearing fault dataset was utilized for experimental analysis. For each category of bearing vibration data, 360 samples were randomly selected, with each sample consisting of 5000 consecutive data points. According to table 8, 300 samples were used for training purposes, and the leftover 60 samples were allocated for testing. The 300 training samples were used to fine-tune the transfer ResNet network. Based on this dataset and its partitioning, two types of cross-domain fault diagnosis tasks were designed: tasks with balanced training samples (SQ-task1-2) and tasks with unbalanced training samples (SQ-task3-4), as shown in table 9. For the unbalanced training sample tasks, the numbers of training samples for fault types (categories 1-10) were 300, 280, 260, 240, 220, 200, 180, 160, 140, and 120, respectively, with a total of 2100 training samples.

Table 8. Description of the SQ-MFS Mechanical Fault Test Bench Bearing Fault Dataset

Bearing Condition	Defect Size (mm)	Training/Test Samples at different motor speed		Class Label
		1200r/min	1500 r/min	
Normal	0	300/60	300/60	1
Inner Race Fault	0.05	300/60	300/60	2
	0.1	300/60	300/60	3
	0.2	300/60	300/60	4
Outer Race Fault	0.05	300/60	300/60	5
	0.1	300/60	300/60	6
	0.2	300/60	300/60	7
Ball Fault	0.05	300/60	300/60	8
	0.1	300/60	300/60	9
	0.2	300/60	300/60	10

Table 9. Cross-Domain Fault Diagnosis Tasks Based on the SQ-MFS Bearing Fault Dataset

Task	Source Domain			Target Domain		
	Speed (r/min)	State Types	Sam-ples	Speed	State Types (r/min)	Samp-les
SQ-task1	1200	10	3000	1500	10	600
SQ-task2	1500	10	3000	1200	10	600
SQ-task3	1200	10	2100	1500	10	600
SQ-task4	1500	10	2100	1200	10	600

4.2.2. Experimental Results and Analysis

The experimental outcomes are displayed in Table 10 , demonstrate the superior fault diagnosis accuracy of the proposed CDF-WTI-RTL model compared to other models across four cross-domain tasks (SQ-task1-4). The average accuracy of the proposed model reached 91.36%, outperforming DFS-TCA (82.38%), DFS-JDA (85.25%), DFS-BDA (84.13%), DFS-GFK (81.38%), and DFS-TJM (87.00%). For SQ-task1, the CDF-WTI-RTL model achieved 92.50%, surpassing the comparative models. Similarly, for SQ-task4, it achieved the highest accuracy of 92.50%, demonstrating consistent superior performance. These results further confirm the effectiveness of the proposed improvements in enhancing domain adaptation capability and improving the discriminative power of feature data, even under challenging conditions such as unbalanced training samples.

Table 10. Comparison of Fault Diagnosis Accuracy for SQ-Tasks1-4

Task	Cross-Domain Fault Diagnosis Accuracy (%)					
	M1	M2	M3	M4	M5	M6
SQ-task1	85.50	88.50	86.50	83.50	87.00	92.50
SQ-task2	82.00	84.50	83.00	82.00	89.50	91.00
Avg. Accuracy (balanced training samples)	83.75	86.5	84.75	82.75	88.25	91.75
SQ-task3	79.50	83.00	81.50	81.00	85.50	90.50
SQ-task4	82.50	85.00	85.50	79.00	86.00	92.50
Avg. Accuracy (unbalanced training samples)	82.38	85.25	84.13	81.38	87.00	91.50

The experimental results also highlight the practical viability of the proposed method for IIoT-enabled predictive maintenance. With an average accuracy of 91.5% under unbalanced training conditions(SQ-task3-4), the CDF-WTI-RTL model demonstrates strong potential for deployment in real-world industrial settings where labeled fault data is scarce and operating conditions vary frequently. Compared to state-of-the-art transfer learning methods such as TCA and JDA, the proposed method achieves superior cross-domain generalization, which is essential for large-scale IIoT systems where edge devices operate under diverse and dynamic environments.

5. Conclusion

This paper proposed a bearing fault diagnosis approach across domains utilizing wavelet time-frequency maps and deep feature transfer learning, named CDF-WTI-RTL. To fully validate the effectiveness, adaptability, and superiority of the proposed cross-domain fault diagnosis method, experiments were conducted using two bearing fault datasets under conditions of both balanced and unbalanced training samples. Comparative analyses were performed against classical feature transfer learning methods. The proposed CDF-WTI-RTL model has demonstrated, through experimental outcomes, its capability to achieve high fault diagnosis accuracy in conditions where training samples are either balanced or unbalanced. The average fault diagnosis accuracy across the two datasets(unbalanced training samples) reached 98.38% and 91.5%, respectively, significantly outperforming the average fault diagnosis accuracy of the comparative models. Furthermore, integration with IIoT architectures offers a promising pathway toward resilient, human-centric predictive maintenance systems aligned with Industry 5.0 principles.

Acknowledgements

This paper has received funding from the "Green and Blue Project" of Jiangsu universities in 2025.

References

[1] *Chaleshtori, Amir Eshaghi, and Abdollah Aghaie*. A novel bearing fault diagnosis approach using the Gaussian mixture model and the weighted principal component analysis, Reliability Engineering & System Safety 242 (2024): 109720.

[2] *Chauhan Sumika, Govind Vashishtha, Rajesh Kumar et al.* An adaptive feature mode decomposition based on a novel health indicator for bearing fault diagnosis, Measurement 226 (2024): 114191.

[3] *Alonso-González Miguel, Vicente García Díaz, Benjamìn López Pérez, et al.* Bearing fault diagnosis with envelope analysis and machine learning approaches using CWRU dataset, IEEE Access 11 (2023): 57796-57805.

[4] *Maurya M, Panigrahi I, Dash D, et al.* Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review. Soft Computing, 2024, 28(1): 477-494.

[5] *Xu Xun, Lu Yuqian, Vogel-Heuser Birgit, et al.* Industry 4.0 and Industry 5.0—Inception, conception and perception. Journal of Manufacturing Systems, 2021, 61: 530-535.

[6] *H. Kitey, P. Chandankhde and K. Jajulwar.* An IIoT Base Platform for Industrial Fault Detection: Implementation and Performance Analysis -A Comprehensive Review.2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 1836-1841.

[7] *Bitam, T.; Yahiaoui, A.; Boubiche, D.E.; Martínez-Peláez, R.et al.* P. Artificial Intelligence of Things for Next-Generation Predictive Maintenance. Sensors 2025, 25, 7636.

[8] *Haiyue Fu, Di Yu, Changshu Zhan, et al.* Unsupervised rolling bearing fault diagnosis method across working conditions based on multiscale convolutional neural network, Measurement Science and Technology, 2024, 35 (3).

[9] *Hou Peng, Zhang Jianjie, Jiang Zhangzheng, et al*. A Bearing Fault Diagnosis Method Based on Dilated Convolution and Multi-Head Self-Attention Mechanism, Applied Sciences, 2023, 13 (23).

[10] *Zhao Yi, Zhang Xinchang, Feng Weiming, et al.* Deep learning classification by ResNet-18 based on the real spectral dataset from multispectral remote sensing images, Remote Sensing, 2022, 14(19): 4883.

[11] *He K, Zhang X, Ren S, et al*. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.

[12] *Wei Fulin, Xu Xueyuan, Jia Tianyuan, et al.* A multi-source transfer joint matching method for inter-subject motor imagery decoding, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 1258-1267.

[13] *Pan Sinno Jialin, Tsang Ivor W, Kwok James T, Yang Qiang,* Domain Adaptation via Transfer Component Analysis, IEEE Transactions on Neural Networks, 2011, 22(2):199-210. TCA

[14] *Wu Jianping, Jiang Bin, Chen Hongtian, et al.* Sensors information fusion system with fault detection based on multi-manifold regularization neighborhood preserving embedding, Sensors, 2019, 19(6): 1440.

[15] *Sugiyama M.* Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, Journal of machine learning research, 2007, 8(5).