

Article citation info:

Jiao X, Zhang J, Cao J, A Physics-Guided Transfer Learning Framework with Consistency Verification for Cross-Domain Bearing Fault Diagnosis, *Eksploracja i Niezawodność – Maintenance and Reliability* 2026: 28(2) <http://10.17531/ein/211797>

A Physics-Guided Transfer Learning Framework with Consistency Verification for Cross-Domain Bearing Fault Diagnosis

Indexed by:



Xinze Jiao^{a,*}, Jianjie Zhang^a, Jianhui Cao^a

^a ¹Xinjiang University, China

Highlights

- A Physics-Constrained Transfer Learning (PCTL) framework is proposed, exploiting invariant physical fault patterns to move beyond purely statistical alignment for trustworthy diagnosis.
- A 'diagnosis-verification-feedback' loop uses a rule-based physics validator to supervise a confidence predictor, deeply coupling model confidence with the physical plausibility of its decisions.
- The framework achieves superior cross-domain accuracy and quantifiable trustworthiness, enabling the model to reliably self-assess and signal physically implausible diagnoses with low confidence.

Abstract

Deep learning models for fault diagnosis face a trust deficit, struggling with poor generalization under varying operating conditions and 'black-box' interpretability. Conventional transfer learning, reliant on statistical alignment, fails to guarantee physical plausibility. We propose a Physics-Constrained Transfer Learning (PCTL) framework based on the core insight that while raw signals vary, intrinsic physical fault patterns—like harmonic structures in the envelope spectrum—remain domain-invariant. Its key innovation is a 'diagnosis-verification-feedback' loop. In this loop, an external, rule-based module quantifies the physical consistency of a diagnosis; this score, in turn, is used as a supervisory signal to guide a confidence predictor, deeply coupling the model's self-assessed confidence with physical plausibility. Our framework not only achieves superior cross-domain diagnostic accuracy but also embeds reliable self-assessment, demonstrated by a strong correlation between its predicted confidence and the underlying physical evidence. This research offers a new paradigm for developing intelligent diagnostic systems that are accurate, physically interpretable, and trustworthy.

Keywords

transfer learning, fault diagnosis, physical consistency, explainable AI, domain adaptation, rolling bearings

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

1.1. Industrial Background and Research Motivation

With the deepening of Industry 4.0 and the comprehensive advancement of intelligent manufacturing, Prognostics and Health Management[1] (PHM) has become a core technology for ensuring the safety, reliability, and efficient operation of modern industrial systems. As a critical component of PHM, fault diagnosis aims to identify early[2] equipment malfunctions by analyzing sensor data, thereby preventing catastrophic failures and optimizing maintenance strategies. In

recent years, data-driven methods, particularly Deep Learning (DL), have achieved significant success in fault diagnosis due to their powerful non-linear feature learning capabilities.

However, despite achieving high accuracy on specific datasets, DL models face a fundamental obstacle when deployed in real industrial environments: a trust deficit. This deficit stems from two core challenges. First, there is the issue [3] of model generalization reliability. The performance of DL models heavily relies on the distributional consistency between training and test data. Yet, in industrial reality, significant

(*) Corresponding author.
E-mail addresses:

X. Jiao (ORCID:0009-0005-1256-6196) 853704013@qq.com, J. Zhang, zhang_jianjie@xju.edu.cn, J. Cao (ORCID: 0009-0001-8943-6235) caojian13523071811@163.com

[4]discrepancies in data distribution (i.e., "domain shift") can arise due to variations in equipment, operating conditions, or even sensor types. This severely compromises the reliability of diagnostic results when a model trained on one piece of equipment is applied to another. Second, there is the "black-box" nature of the decision-making process. Most DL models cannot provide decision-making rationales that align with human engineers' cognitive logic[5]. In safety-critical domains such as aerospace, high-speed rail, and energy, a diagnostic system [6]that cannot explain "why it made a particular judgment" is difficult to adopt and trust.

1.2. Existing Methods and Their Limitations

To enhance model generalization, Transfer Learning (TL) has been widely applied in cross-domain fault diagnosis tasks[7]. Existing methods, such as strategies based on Maximum Mean Discrepancy (MMD) or Domain Adversarial Neural Networks (DANN), primarily focus on aligning the statistical distributions [8]of source and target domains in the feature space. While these methods mitigate domain shift to some extent, they do not

address the core of the trust deficit. Their fundamental limitation lies in the fact that statistical alignment alone cannot guarantee the physical plausibility of model decisions. A model might achieve "correct" classifications by learning spurious correlations common to both source and target domains but irrelevant to fault physics, rendering its internal reasoning process fragile and untrustworthy.

This leads to a deeper[9] question: How should an ideal cross-domain diagnostic model construct its decision-making process [10]to not only "know what" (make a diagnosis) but also "know why" (based on physical laws), thereby earning the trust of engineers? To answer this question, we conducted [11]an exploratory analysis.

1.3. Core Insight: From 'Apparent Discrepancies' to 'Physical Commonalities'

We selected two widely recognized public datasets with distinct characteristics—CWRU (representing steady-state conditions) and PU (representing transient conditions)—and compared their inner race fault signals, as shown in Figure 1.

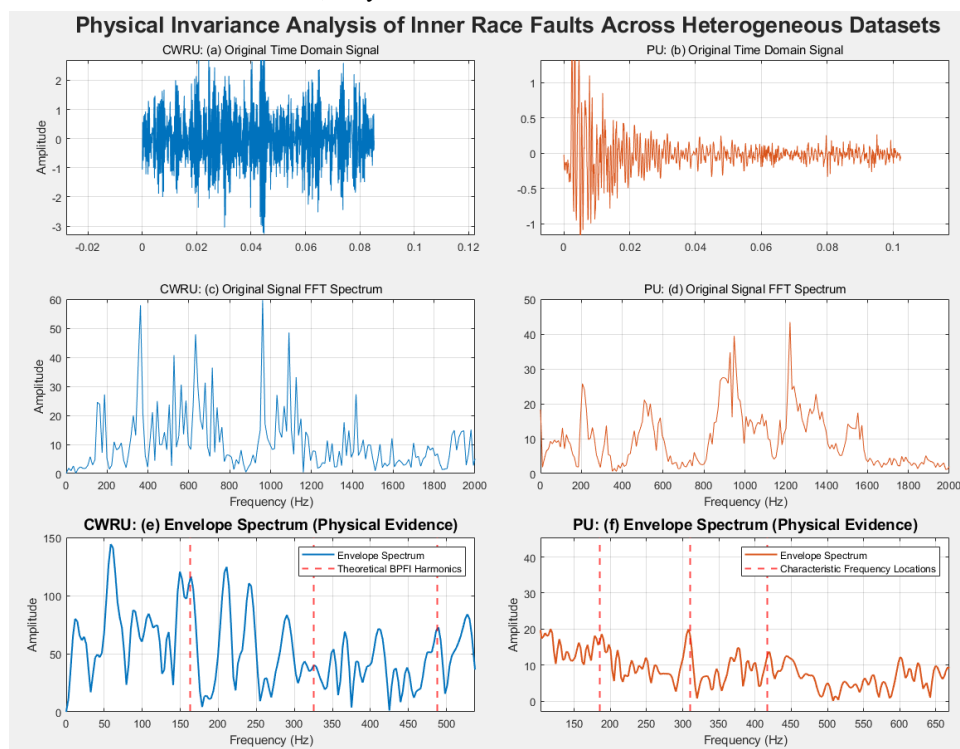


Figure 1. Analysis of Physical Invariance of Inner Race Faults in Heterogeneous Datasets. (a, b) show the raw time-domain signals from CWRU and PU datasets, respectively, exhibiting distinct morphologies. (c, d) present their corresponding FFT spectra, where energy distribution and dominant frequency components show no clear correlation. (e, f) display the envelope spectra obtained after envelope demodulation; both clearly reveal a "fundamental frequency + harmonic" family structure determined by the physical mechanism of inner race faults (red dashed lines indicate theoretical BPFI harmonic positions), demonstrating a high degree of pattern consistency.

As depicted in Figures 1(a) through 1(d), the raw signals from the two datasets exhibit substantial disparities, both in their time-domain waveforms and their original Fast Fourier Transform (FFT) spectra[12]. This stark signal variance visually underscores the formidable challenge of cross-domain diagnostics and elucidates why conventional transfer learning methods, which rely on learning superficial statistical features, often prove ineffective.

Conversely, a striking and universal [13]physical pattern emerges when we employ envelope demodulation to probe the inherent periodicities of the fault impulses. As illustrated in Figures 1(e) and 1(f), despite drastic differences in signal origin, morphology, and noise background, the envelope spectra from both datasets reveal a remarkably consistent physical regularity: a distinct harmonic family structure[14]. The locations of these concentrated spectral peaks align precisely with the theoretical Ball Pass Frequency of the Inner race (BPFI) and its harmonics, calculated from the specific bearing parameters and rotational speed of each setup.

This discovery substantiates our core thesis: although raw vibration signals from heterogeneous equipment are volatile and equipment-specific, the underlying patterns dictated by the physical fault mechanism are universal and domain-invariant[15]. This Physical Pattern Invariance constitutes the truly reliable "knowledge" that bridges disparate data domains and serves as the cornerstone for constructing trustworthy diagnostic models.

1.4. Proposed Method and Key Contributions

The preceding analysis reveals several critical challenges that currently hinder progress in cross-domain fault diagnosis:

1. **Superficial Knowledge Transfer:** Prevailing Domain Adaptation (DA) methods, such as DANN and JDA, primarily seek to align statistical moments of marginal or joint probability distributions. This "implicit" alignment offers[16] no guarantee that the model internalizes the intrinsic physical principles governing fault behavior, resulting in compromised generalization performance when confronted with substantial physical system variations.

2. **Absence of Physical Grounding:** In the pursuit of high accuracy, most end-to-end deep learning models overlook the physical plausibility of their diagnostic conclusions[17].

A model may achieve correct classification by exploiting spurious statistical correlations, but its decision-making process, lacking a physical foundation, fails to secure the trust of engineers.

3. **Unquantified Decision Uncertainty:** Existing models typically yield a categorical prediction without a reliable confidence measure to articulate the diagnosis's credibility. An ideal model should possess a self-assessment capability, proactively signaling when its conclusions are physically implausible.

To surmount these challenges, we introduce the Physics-Constrained Transfer Learning (PCTL) framework. The central tenet of PCTL is to transform physical knowledge from an implicit attribute the model is expected to "discover" into an explicit, quantifiable constraint integrated directly into a closed-loop learning and validation process[18]. The PCTL framework consists of two main components: a physical knowledge encoding module and a novel physical consistency verification module. During pre-training, the model learns domain-invariant physical features via envelope spectrum prediction and domain-adversarial training[19]. During fine-tuning, the physical consistency verification module leverages physical priors of the target equipment to cross-validate the model's diagnosis against its predicted physical evidence, yielding a physical consistency score. This score, in turn, supervises a confidence predictor, compelling the model's self-assessed confidence to be strongly correlated with the physical plausibility of its decision.

The primary contributions of this work are threefold:

(1) We propose a novel transfer learning paradigm guided by physical consistency verification. By introducing an external, rule-based "physical referee" to supervise and constrain the learning process, our paradigm transcends the limitations of traditional methods that rely solely on statistical distribution alignment[20]. This approach endows the deep model with physical-level "insight." (2) We design a learning framework featuring a closed-loop "Diagnose-Validate-Feedback" mechanism. The PCTL framework, particularly through its physical consistency verification and confidence learning mechanisms during fine-tuning, establishes a complete decision-and-validation loop. This empowers the model not only to issue diagnoses but also to perform self-assessment of

their physical reasonableness, thereby markedly enhancing decision transparency and credibility[21]. (3) We achieve a simultaneous enhancement of diagnostic performance and model trustworthiness. Extensive cross-domain experiments on public datasets demonstrate that the PCTL framework attains superior diagnostic accuracy while its predicted confidence exhibits a strong positive correlation with the physical consistency score. This introduces a novel and effective metric for quantifying the trustworthiness of intelligent diagnostic models, which is of paramount importance for advancing the adoption of Trustworthy AI in safety-critical industrial applications.

2. Proposed Methodology

To overcome the challenges of reliability and trustworthiness in transfer-based diagnosis across heterogeneous equipment—a problem exacerbated by substantial domain shifts—we

propose the Physics-Constrained Transfer Learning (PCTL) framework. Its core principle is to explicitly and quantifiably integrate physical prior knowledge of faults into the model's training and validation via a two-stage learning process. The first stage, Physics-Guided Domain-Adversarial Pre-training, is designed to learn a universal, domain-invariant fault representation from source-domain data that is rich in physical information. The second stage, Physics-Aware and Consistency-Constrained Fine-tuning, aims to efficiently adapt this universal knowledge to the target domain while simultaneously ensuring the physical plausibility and credibility of the model's diagnostic decisions through an innovative closed-loop verification mechanism[22].

2.1. Overall Framework

The overall architecture of the PCTL framework, depicted in Figure 2, is composed of two sequential stages:

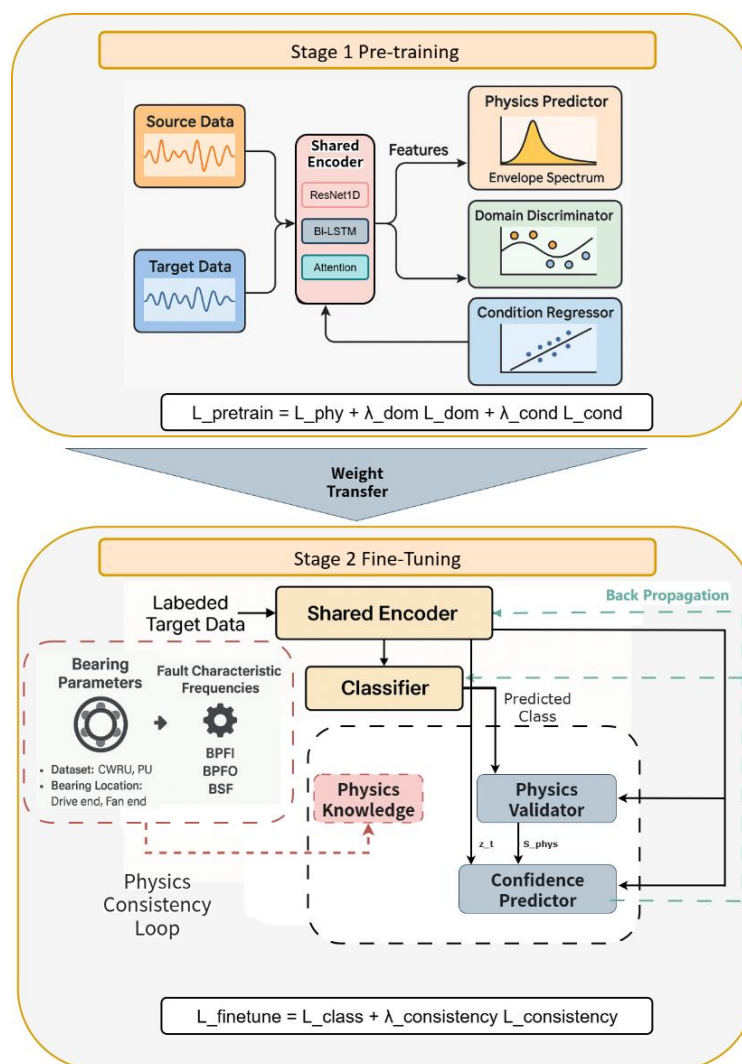


Figure 2. The overall architecture of the PCTL framework.

Stage 1: Pre-training. This stage takes source domain data (Source Data, D_s) and unlabeled target domain data (Target Data, D_t) as input. A Shared Encoder (E) extracts latent features. The training is driven by three concurrent objectives: (i) a Physics Predictor (P_{phy}), trained via supervised learning, forces the encoded features to reconstruct the original signal's envelope spectrum, thereby capturing underlying physical patterns; (ii) a Domain Discriminator (D_{dom}), through adversarial training with the encoder, promotes the learning of domain-invariant features; and (iii) a Condition Regressor (C_{reg}), acting as an auxiliary task, helps to disentangle features related to operating conditions from those indicative of faults.

Stage 2: Fine-tuning. The weights of the pre-trained Shared Encoder (E) are transferred and fine-tuned on a new model. This stage utilizes a small set of labeled target domain data (Labeled Target Data). The centerpiece of this stage, beyond a final Classifier (C), is a Physics Consistency Verification Loop. This loop incorporates an external Physics Validator and a Confidence Predictor (Conf). Upon a diagnosis, the Physics Validator evaluates the consistency between the model's conclusion and the predicted physical evidence, generating a physical consistency score, S_{phys} . This score then serves as the supervisory signal[23] to optimize the Confidence Predictor, compelling a strong alignment between the model's confidence and the underlying physical principles.

2.2. Stage 1: Physics-Guided Domain-Adversarial Pre-training

The objective of the pre-training stage is to learn a transferable feature representation from source-domain data that encapsulates universal physical principles.

Vibration signals typically comprise both non-stationary impulses caused by faults and periodic components generated by machine rotation. To account for this, we design a hybrid Shared Encoder (E). The encoder consists of a one-dimensional residual network (ResNet1D) and a bidirectional long short-term memory network (Bi-LSTM), followed by an attention mechanism. Specifically, an input signal x is first fed into the ResNet1D to extract its deep local structural features, $f_{cnn} = \text{ResNet1D}(x)$. Concurrently, the raw signal x is processed by the Bi-LSTM to capture its long-range temporal dependencies[24], $f_{lstm} = \text{Bi-LSTM}(x)$. Finally, both feature

vectors are concatenated and passed through an attention module, which adaptively fuses them to produce the final feature representation:

$$z = \text{Attention}(\text{concat}(f_{cnn}, f_{lstm})) \quad (1)$$

To explicitly embed physical information about the fault into the feature representation z , we introduce a Physics Predictor (P_{phy}), implemented as a simple multi-layer perceptron (MLP). Its task is to reconstruct the envelope spectrum s_{env} of the original signal x from the feature z . The envelope spectrum, obtained via the Hilbert transform, effectively reveals the periodicity of fault impulses and is a crucial piece of physical evidence in diagnostic analysis. We define the Physics-Guided Loss (L_{phy}) as the Mean Squared Error (MSE) between the predicted envelope spectrum $P_{phy}(z)$ and the ground-truth envelope spectrum s_{env} :

$$L_{phy} = ||P_{phy}(E(x_s)) - s_{env}(x_s)||_2^2 \quad (2)$$

where x_s is a signal from the source domain. This loss function compels the encoder E to learn features that retain sufficient physical information to reconstruct the critical fault harmonic structures.

To eliminate equipment-specific “domain fingerprints” from the feature z , we employ a domain-adversarial training mechanism. A Domain Discriminator (D_{dom}), also an MLP, is introduced to distinguish whether a given feature z originates from the source or target domain. We insert a Gradient Reversal Layer (GRL) between the encoder E and the discriminator D_{dom} . The GRL acts as an identity transformation during forward propagation but multiplies the gradient from the discriminator's loss by a negative constant[25], $-\alpha$, during backpropagation.

Consequently, the optimization objective of the discriminator D_{dom} is to minimize the domain classification error, whereas the objective of the encoder E , due to the gradient reversal, becomes maximizing this error. This adversarial game ultimately forces the encoder E to generate features that are indistinguishable to the discriminator D_{dom} , thereby achieving domain invariance. The domain-adversarial loss, L_{dom} , is defined as the standard Binary Cross-Entropy (BCE) loss:

$$L_{dom} = -\mathbb{E}_{x_s \sim D_s}[\log D_{dom}(E(x_s))] - \mathbb{E}_{x_t \sim D_t}[\log(1 - D_{dom}(E(x_t)))] \quad (3)$$

D_s and D_t represent the data distributions of the source and target domains, respectively.

The total objective function for the pre-training stage is a weighted sum of the aforementioned losses:

$$L_{pretrain} = L_{phy} + \lambda_{dom}L_{dom} + \lambda_{cond}L_{cond} \quad (4)$$

where L_{cond} is an auxiliary loss from the condition regressor, and λ_{dom} and λ_{cond} are hyperparameters that balance the contribution of each loss term. By minimizing this objective function, we obtain a pre-trained Shared Encoder E capable of extracting domain-invariant physical features.

2.3. Stage 2: Physics-Aware and Consistency-Constrained Fine-tuning

The objective of the fine-tuning stage is twofold: to adapt the pre-trained universal knowledge to the target domain and to establish the trustworthiness of the model's decisions.

We load the weights of the Shared Encoder E pre-trained in Stage 1 and append a new, lightweight Classifier C (an MLP structure). This classifier is responsible for mapping the target-domain features $z_t = E(x_t)$ to the final fault class probability distribution[26], $y_{pred} = C(z_t)$. The classification loss, L_{class} , employs a weighted cross-entropy function to address potential class imbalance in the target domain data.

The core innovation of our framework resides in this stage. We introduce an external, rule-based Physics Consistency Verification (PCV) module, which does not participate in gradient backpropagation but acts as an objective “physical referee.” For each target-domain sample x_t , after the model yields a predicted class $c_{pred} = \text{argmax}(y_{pred})$, the PCV module executes the verification process detailed in Algorithm 1.

Algorithm 1 Physics Consistency Verification (PCV) Module

Input: Predicted class c_{pred} , predicted envelope spectrum $s_{pred} = P_{phy}(E(x_t))$, target machine parameters (e.g., bearing geometry, rotation speed f_r).

Output: Physics Consistency Score S_{phys} .

```

1: if  $c_{pred}$  is ‘Normal’ then
2: // For normal state, score is high if no fault frequencies
   are prominent.
3:  $S_{phys} \leftarrow 1 - \text{mean\_energy\_at\_fault\_freqs}(s_{pred})$ 
4: else
5: // Calculate theoretical characteristic fault frequency for
    $c_{pred}$ .
```

```

6:  $f_{char} \leftarrow \text{calculate\_theoretical\_freq}(c_{pred}, f_r)$ 
7: // Find energy peaks in  $s_{pred}$  around  $f_{char}$  and its
   harmonics.
8: harmonic_scores  $\leftarrow []$ 
9: for  $h$  in  $\{1,2,3\}$  do // Check first 3 harmonics
10:     peak_energy  $\leftarrow$ 
       find_peak_energy_in_window( $s_{pred}, h \cdot f_{char}$ )
11: harmonic_scores.append(peak_energy)
12: end for
13: // The score is the average prominence of these harmonic
   peaks.
14:  $S_{phys} \leftarrow \text{mean}(\text{harmonic\_scores})$ 
15: end if
16: return  $S_{phys}$  // A score between 0 and 1.
```

The output of the PCV module, S_{phys} , is a scalar value between 0 and 1 that quantifies the extent to which the model’s diagnosis is supported by its own predicted physical evidence. A higher S_{phys} indicates a diagnosis that is more consistent with physical principles.

To equip the model with a self-assessment capability, we introduce a Confidence Predictor, C_{conf} . It takes the encoded feature z_t as input and outputs a single confidence value, $c_{model} = C_{conf}(z_t)$, representing the model’s belief in its own diagnosis.

Crucially, instead of allowing the model to learn confidence “out of thin air,” we use the objective physical consistency score S_{phys} from the PCV module to supervise it. The consistency-constraint loss, $L_{consistency}$, is defined as the Mean Squared Error between the model’s predicted confidence and the physical consistency score:

$$L_{consistency} = \|C_{conf}(E(x_t)) - S_{phys}\|_2^2 \quad (5)$$

This loss function establishes a powerful feedback loop: if the model makes a physically implausible diagnosis (S_{phys} is low) but exhibits high confidence (c_{model} is high), $L_{consistency}$ will be large. The resulting gradient will penalize both the Confidence Predictor C_{conf} and the Encoder E , forcing them to adjust. Over time, the model learns to assign high confidence only when its diagnostic conclusion is strongly supported by physical evidence.

The total objective function for the fine-tuning stage integrates the classification task and the trustworthiness constraint:

$$L_{finetune} = L_{class} + \lambda_{consistency} L_{consistency} \quad (6)$$

where $\lambda_{consistency}$ is a key hyperparameter that balances classification accuracy and decision credibility. By minimizing this function on a small amount of labeled target-domain data, the PCTL framework ultimately yields diagnoses that are both accurate and physically trustworthy.

3. Experimental Setup

To comprehensively and rigorously evaluate the effectiveness and superiority of the proposed Physics-Constrained Transfer Learning (PCTL) framework, we have designed a series of challenging cross-domain fault diagnosis experiments. This chapter provides a detailed account of the datasets used, the construction of cross-domain transfer scenarios, the baseline methods for performance comparison, the implementation

Table 1. Detailed parameters of the datasets used in this study.

Datase	Bearing Type	Fault Type	Characteristics
CWRU	SKF 6205-2RS	0.007-0.021 in.	Lab-based steady-state conditions; faults induced by Electrical Discharge Machining (EDM).
PU	6203	Natural wear	Full run-to-failure lifecycle; includes natural wear and accelerated degradation processes.
PHM 2012	-	Natural wear	Industrial competition context; complex operating conditions with multi-stage degradation.

Case Western Reserve University (CWRU) Dataset: As the most classic benchmark dataset in the field of bearing fault diagnosis, CWRU (Figure 3) features faults that are artificially induced using Electrical Discharge Machining (EDM). The

details of our model, and the metrics for comprehensive evaluation.

3.1. Datasets Description

This study aims to validate the efficacy of the proposed PCTL framework in cross-heterogeneous-bearing fault diagnosis tasks. To this end, we primarily utilize two internationally recognized rolling bearing datasets with distinct physical characteristics[27]: the Case Western Reserve University (CWRU) dataset and the Paderborn University (PU) dataset. Furthermore, to investigate the framework's ability to learn and generalize universal physical laws, we incorporate the gearbox dataset from the **IEEE PHM 2012 Data Challenge** for auxiliary validation. The key parameters of these datasets are summarized in Table 1.

resulting signal characteristics are distinct, making it well-suited for validating the fundamental performance of algorithms under steady-state operating conditions.

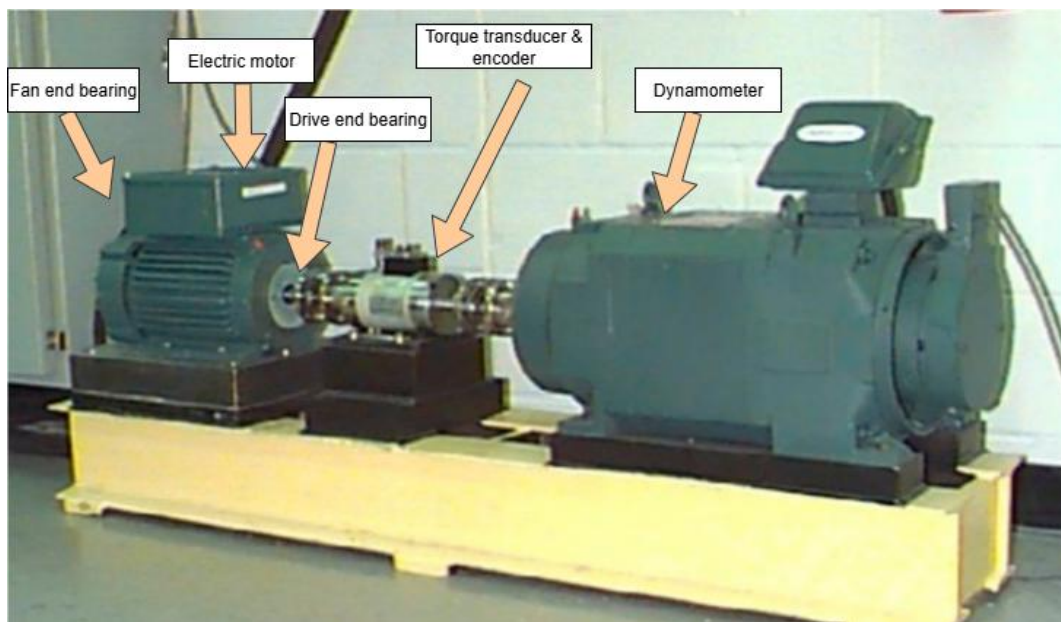


Figure 3. The experimental test rig for the Case Western Reserve University (CWRU) dataset. This setup uses EDM to create artificial bearing faults.

Paderborn University (PU) Dataset: This dataset (Figure 4) contains faults generated from natural wear and accelerated life tests, covering the entire lifecycle from a healthy state to

complete failure. Its signal characteristics are more complex and more closely resemble those found in real-world industrial settings.

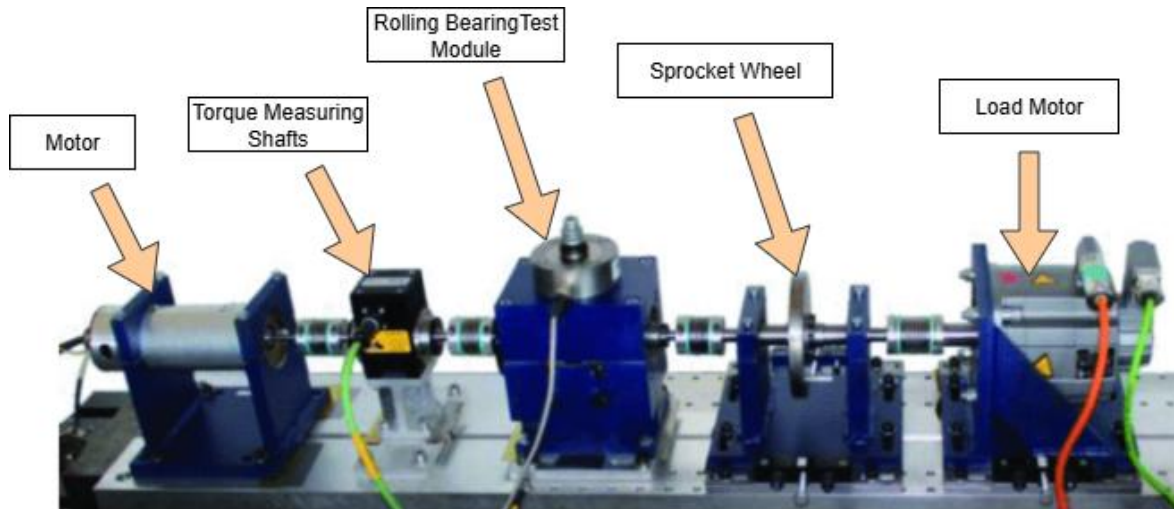


Figure 4. The full run-to-failure bearing test rig for the Paderborn University (PU) dataset, used to collect data from natural degradation processes.

IEEE PHM 2012 Data Challenge Dataset: This dataset (Figure 5) originates from a real-world industrial challenge. It is particularly noteworthy that the primary fault objects in this dataset are gearboxes. We include it in our study to construct a highly challenging transfer scenario. This allows us to test whether our PCTL framework can successfully disentangle and

learn fundamental vibration degradation laws—independent of specific components—from a complex system containing multiple vibration sources (e.g., gear meshing, shaft rotation, bearings) and effectively transfer this knowledge for accurate bearing diagnosis.

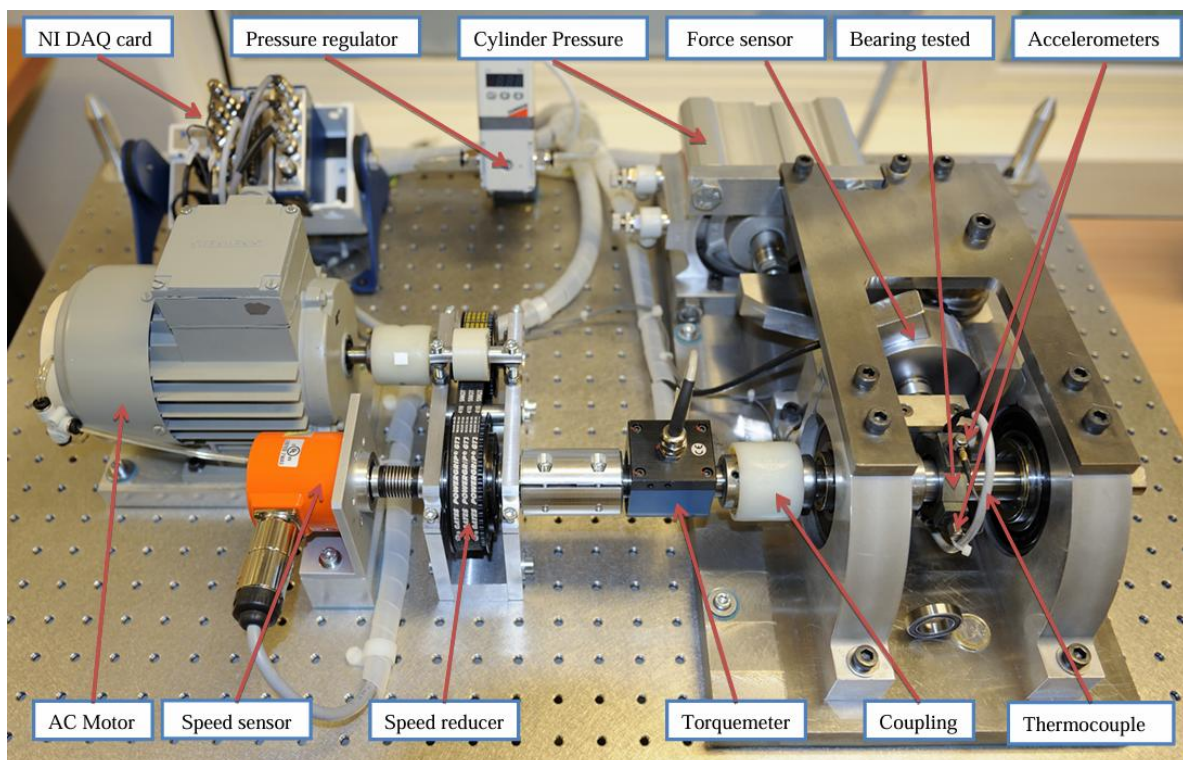


Figure 5. The gearbox fault prognostics test rig for the IEEE PHM 2012 Data Challenge.

3.2. Cross-Domain Scenarios

To comprehensively evaluate the performance of the PCTL framework, we have designed a series of cross-domain transfer tasks of increasing difficulty. In each task, we adhere to the standard definition of transfer learning: the dataset used for pre-training is defined as the source domain D_s , with its labeled data

denoted as $\{(x_s, y_s)\}$; the dataset used for fine-tuning and testing is defined as the target domain D_t , with its data denoted as $\{x_t\}$. experiments primarily focus on bidirectional transfer between the two core bearing datasets (CWRU and PU) and also construct cross-component transfer tasks from the gearbox dataset to the bearing datasets. The specific task configurations are detailed in Table 2.

Table 2. Configuration of the cross-domain transfer tasks.

Task	Source Domain (Pre-training)	Target Domain (Fine-tuning)	Core Challenge
T1	CWRU	PU	Steady-state $\rightarrow$ Variable conditions; Artificial fault $\rightarrow$ Natural wear
T2	PU	CWRU	Variable conditions $\rightarrow$ Steady-state; Natural wear $\rightarrow$ Artificial fault
T3	PHM 2012	CWRU	Cross-component transfer; Industrial $\rightarrow$ Lab (steady-state)
T4	PHM 2012	PU	Cross-component transfer; Industrial $\rightarrow$ Lab (variable conditions)

In all tasks, we follow the Unsupervised Domain Adaptation (UDA) setting for pre-training, where source-domain data is labeled and target-domain data is unlabeled. During the fine-tuning stage, we use a small, labeled subset of the target domain to update the model parameters.

3.3. Baselines for Comparison

To validate the superiority of our proposed PCTL framework, we select several representative categories of methods for comparison:

1. No Transfer:

Source Only: A model is trained exclusively on source-domain data and then directly tested on the target domain. This baseline measures the magnitude of the domain shift.

Target Only: A model is trained from scratch using only the small, labeled subset of the target-domain data.

2. Conventional Transfer Learning:

Fine-tuning: A model is pre-trained on the source-domain data, after which the entire pre-trained encoder is fine-tuned using the small, labeled subset of the target domain.

DANN (Domain-Adversarial Neural Network): A classic domain-adversarial network that learns domain-invariant features through adversarial training with a gradient reversal layer.

JDA (Joint Distribution Adaptation): A classic method that aligns both marginal and conditional distributions simultaneously.

3. State-of-the-Art Contrastive Learning:

TF-C (Time-Frequency Consistency): A state-of-the-art self-supervised contrastive learning framework for time series. TF-C pre-trains an encoder by treating the representations of the time-domain and frequency-domain views of the same signal sample as a positive pair, and those from different samples as negative pairs. It uses a contrastive loss (e.g., NT-Xent) to maximize the agreement between positive pairs. For our comparison, this pre-trained encoder is then fine-tuned on the labeled target data, representing a powerful baseline for representation learning.

4. Ablation Study: To verify the necessity of each innovative component in the PCTL framework, we design the following variants:

PCTL w/o Phy: The PCTL framework without the physics-guided loss (L_{phy}) during the pre-training stage.

PCTL w/o Dom: The PCTL framework without the domain-adversarial loss (L_{dom}) during the pre-training stage.

PCTL w/o Cons: The PCTL framework without the core innovation of the fine-tuning stage—the physical consistency constraint loss ($L_{consistency}$).

3.4. Implementation Details

All experiments were conducted in a unified hardware and software environment to ensure fair comparison.

Data Preprocessing: Raw vibration signals were segmented into samples of 1024 data points. We employed overlapping sampling to augment the number of samples. All samples were standardized using z-score normalization.

Network Architecture: The ResNet1D in the Shared Encoder contains 4 residual blocks and outputs a 128-dimensional feature vector. The hidden layer dimension of the Bi-LSTM is 128. All MLPs (used for the predictor, discriminator, classifier, etc.) consist of 2 fully connected layers with ReLU activation functions.

Training Parameters:Pre-training Stage: We used the Adam optimizer with a learning rate of 1e-3 and a batch size of 64 for 100 epochs. The loss weights were set to $\lambda_{dom} = 0.1$ and $\lambda_{cond} = 0.5$.

Fine-tuning Stage: We used the Adam optimizer with a smaller learning rate for the encoder (1e-5) and a larger learning rate for the newly added classifier and confidence predictor (1e-4). The batch size was 32, and training ran for 50 epochs. The consistency loss weight was set to $\lambda_{consistency} = 0.2$.

3.5. Evaluation Metrics

To comprehensively evaluate our proposed method from the dual perspectives of diagnostic performance and model trustworthiness, we employ the following evaluation metrics:

1.Accuracy: This is the standard metric for classification performance, defined as the ratio of correctly classified samples to the total number of test samples:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Test Samples}} \quad (7)$$

2.Physics Consistency Score (PCS): This is a novel metric we propose to quantify the physical trustworthiness of a model. It is calculated as the average of the physical consistency scores (S_{phys} , derived from the PCV module as described in Section 2.3) for all correctly classified samples in the test set. A higher PCS indicates that the model not only makes correct diagnoses but also that its conclusions are supported by stronger physical evidence[28].

$$PCS = \frac{1}{|S_{correct}|} \sum_{i \in S_{correct}} S_{phys}^{(i)} \quad (8)$$

Where $S_{correct}$ is the set of correctly classified samples.

3.Trustworthiness Analysis:Confidence Distribution Plots: We plot the Probability Density Functions (PDFs) of the confidence scores assigned by the model to both correctly and incorrectly classified samples. For a trustworthy model, these two distributions should be clearly separated.

Pearson Correlation Coefficient: We compute the Pearson correlation coefficient between the model's predicted confidence, c_{model} , and its physical consistency score, S_{phys} . A high positive correlation coefficient (close to 1) serves as strong quantitative evidence of the model's trustworthiness.

4. Results and Discussion

This chapter aims to comprehensively evaluate the proposed PCTL framework through a series of integrated experiments. We will first present the quantitative performance of PCTL on multiple cross-domain transfer tasks, comparing it against existing baseline methods. Subsequently, we will conduct an in-depth analysis of the contributions of each key component within the framework via ablation studies. Finally, we will focus on dissecting the model's performance in terms of physical consistency and decision trustworthiness to validate the core advantages of our method.

4.1. Quantitative Performance Comparison

To validate the overall performance of the PCTL framework, we conducted experiments on the four cross-domain transfer tasks defined in Table 2. Table 3 presents a detailed comparison of PCTL against all baseline methods on two key metrics: diagnostic Accuracy and the Physics Consistency Score (PCS). All results are reported as the mean $\pm$ standard deviation of 5 independent trials.

Table 3. Performance comparison of different methods on cross-domain transfer tasks (Mean $\pm$ Std. Dev.).

Task	Method	Accuracy (%)	PCS
T1: CWRU→PU	Source Only	35.4 $\pm$ 2.1	0.0617
	Target Only	38.2 $\pm$ 1.5	0.0389
	Fine-tuning	65.1 $\pm$ 0.8	0.1971
	DANN	54.6 $\pm$ 1.1	0.2198
	TF-C	66.3 $\pm$ 1.8	0.3168
	PCTL(Ours)	99.1 $\pm$ 0.5	0.7353
T2: PU→CWRU	Source Only	41.2 $\pm$ 2.5	0.0992
	Target Only	55.9 $\pm$ 0.7	0.3184
	Fine-tuning	50.4 $\pm$ 1.1	0.2992

T3: PHM→CWRU	DANN	33.7±11.7	0.2093
	TF-C	59.3±2.2	0.2674
	PCTL(Ours)	98.8 ± 0.4	0.7189
	Source Only	25.7 ± 3.0	0.0811
	Target Only	16.7±3.7	0.1671
	Fine-tuning	63±1.8	0.4330
	DANN	61.3±3.9	0.4844
T4: PHM→PU	TF-C	57.3±3.1	0.2136
	PCTL(Ours)	99.2 ± 0.3	0.7398
	Source Only	28.9 ± 2.8	0.1221
	Target Only	25.33±3.2	0.1079
	Fine-tuning	44.18±4.2	0.2697
	DANN	49.36±7.0	0.3783
	TF-C	58.3±1.3	0.2118
	PCTL(Ours)	97.5 ± 0.7	0.5811

To provide a more intuitive visualization of the PCTL framework's classification performance across tasks, Figure 6 displays its confusion matrices for tasks T1 through T4.

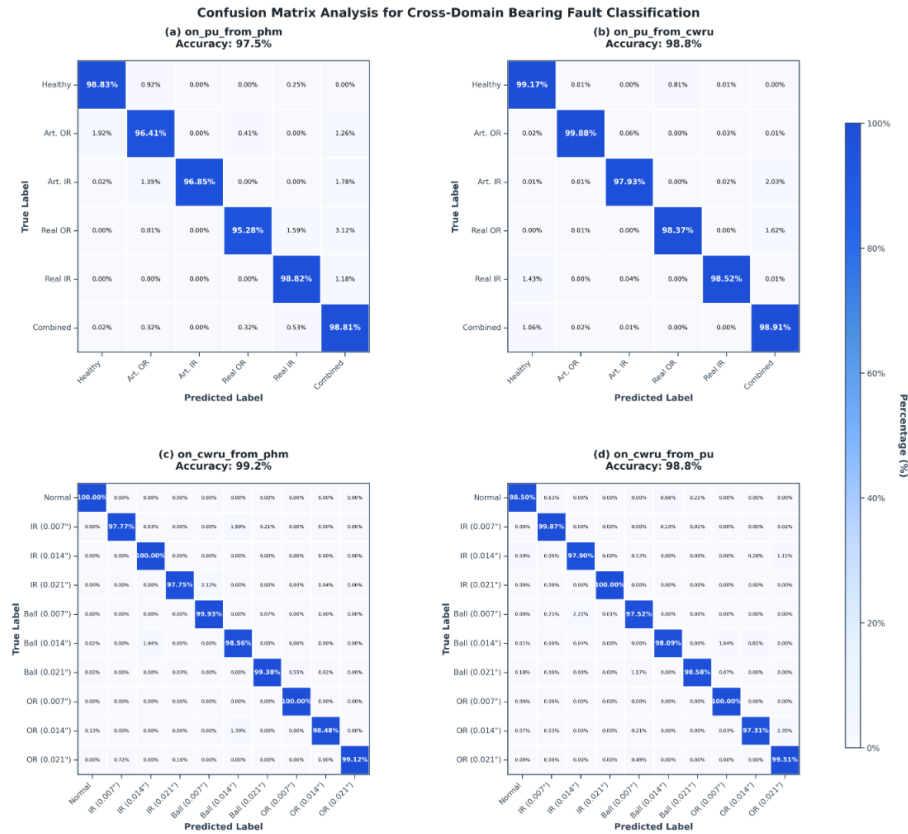


Figure 6: Confusion matrices of the PCTL framework on various cross-domain transfer tasks. The subplots correspond to (a) task T4 (PHM→PU), (b) task T1 (CWRU→PU), (c) task T3 (PHM→CWRU), and (d) task T2 (PU→CWRU). In each matrix, the diagonal elements represent the percentage of correct classifications, while the off-diagonal elements represent misclassifications. High diagonal values and low off-diagonal values indicate superior classification performance.

Based on the experimental results in Table 3 and Figure 6, we can draw the following key conclusions:

1. Overall Superiority of the PCTL Framework: In all four cross-domain tasks, our proposed PCTL framework achieved overwhelmingly superior performance on both core metrics,

accuracy and PCS, comprehensively outperforming all baseline methods. For instance, in task T1 (CWRU→PU), PCTL achieved an accuracy of 99.1%, surpassing traditional Fine-tuning (65.1%) and DANN (54.6%) by over 34 and 44 percentage points, respectively. The confusion matrices in

Figure 6 further confirm that PCTL exhibits extremely high classification precision across all categories, with minimal misclassification rates and diagonal elements consistently approaching 100%. This substantial performance advantage remained consistent across all tasks, fully demonstrating the effectiveness and robustness of the PCTL framework.

2.Capability to Handle Extreme Domain Shifts: In the most challenging cross-component transfer tasks (T3: PHM→CWRU and T4: PHM→PU), the domain gap is starkly evident. The accuracy of the Source Only model plummets to 25.7% and 28.9%, respectively—nearly equivalent to random guessing. This indicates fundamental differences in the signal characteristics between gearboxes and bearings. Nevertheless, even under these extreme challenges, the PCTL framework still achieved remarkable accuracies of 99.2% (Figure 6(c)) and 97.5% (Figure 6(a)). This result strongly validates our core thesis: PCTL can penetrate the superficial differences of components to learn more fundamental, transferable physical laws that govern degradation.

3.The Deeper Value Revealed by the PCS Metric: As a metric for the physical plausibility of model decisions, PCS offers profound insights beyond mere accuracy. We observe that PCTL not only leads in accuracy but also consistently achieves the highest PCS. For example, in task T2, the PCS of PCTL (0.7189) is significantly higher than that of Target Only (0.3184) and Fine-tuning (0.2992). This demonstrates that the correct diagnoses made by PCTL are backed by stronger physical evidence, rendering its decision-making process more reliable. In contrast, while traditional transfer methods can improve accuracy to some extent, their limited improvement in PCS suggests a greater reliance on statistical fitting rather than

genuine physical insight.

4.Analysis of the DANN "Failure" Phenomenon: An intriguing phenomenon occurred in task T2 (PU→CWRU), where the accuracy of the classic DANN method was merely 33.7%, even lower than the non-transfer Source Only baseline (41.2%), exhibiting clear "negative transfer." We surmise this is because the signal characteristics of the PU dataset are more complex and varied than those of CWRU. When DANN forcibly aligns the overall distributions of the two domains, it may erroneously transfer unique but fault-irrelevant complex patterns from PU to the target domain, thereby interfering with the diagnosis of simpler faults in CWRU. This very observation highlights the advantage of PCTL: instead of blindly aligning all features, PCTL selectively learns and transfers features with universal physical meaning through physics-guided mechanisms, thus effectively mitigating the risk of negative transfer.

4.2. Ablation Study

To dissect the PCTL framework and verify the indispensability of its key internal components, we conducted a series of rigorous ablation studies. On the most challenging cross-component transfer task, T3 (PHM→CWRU), we compared the full PCTL framework against three of its variants: 1) w/o Phy: removing the physics-guided loss L_{phy} during pre-training; 2) w/o Dom: removing the domain-adversarial loss L_{dom} during pre-training; and 3) w/o Cons: removing the physical consistency constraint $L_{consistency}$ during fine-tuning. The results are presented in Figure 7, which also includes two baseline methods (Fine-tuning, DANN) for reference.

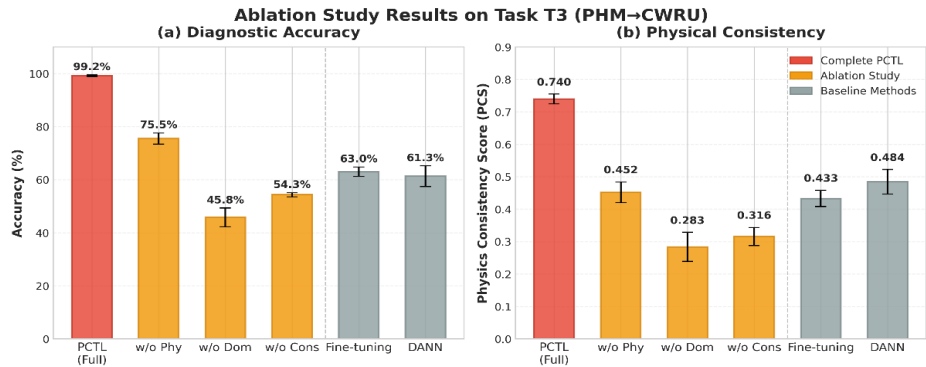


Figure 7: Ablation study results on task T3 (PHM→CWRU). (a) Comparison of diagnostic accuracy among different models. (b) Comparison of the Physics Consistency Score (PCS) among different models. The red bars represent the complete PCTL model, orange bars represent its ablated variants, and gray bars represent the baseline methods.

An in-depth analysis of Figure 7 leads to the following conclusions:

1. Physics Guidance is the Cornerstone of Knowledge Transfer.

Observing the w/o Phy variant, the removal of the physics-guided loss causes the model's accuracy to plummet from 99.2% to 75.5%, while the PCS also drops sharply from 0.740 to 0.452. This clearly demonstrates that compelling the model to predict the envelope spectrum during pre-training is the fundamental reason for its ability to capture transferable physical patterns. Without this explicit physical guidance, the model tends to learn superficial, equipment-specific features from the source domain, leading to a significant degradation in the effectiveness of knowledge transfer when confronted with a vast domain gap.

2. Domain Adaptation is Key to Ensuring Generalization.

For the w/o Dom variant, removing the domain-adversarial module results in a catastrophic performance collapse, with its accuracy (45.8%) and PCS (0.283) falling far below those of the simple Fine-tuning method (63.0%, 0.433). This aligns with our expectations, as domain adaptation is the core mechanism for bridging the distribution gap between the source and target domains and ensuring the model's generalization capability. This result powerfully demonstrates that physics guidance and domain adaptation are complementary and indispensable within the PCTL framework: the former ensures the model learns the "right knowledge" (physical laws), while the latter ensures this knowledge can be "correctly applied" (generalized across domains) to new equipment.

3. Consistency Constraint is the "Final Mile" to Model Trustworthiness.

The results from the w/o Cons variant offer the most profound insight of this study. When the physical consistency constraint $L_{consistency}$ is removed, the model's accuracy drops significantly from 99.2% to 54.3%. However, the change in its PCS is even more revealing: it plummets from 0.740 to 0.316, a score even lower than that of Fine-tuning (0.433) and DANN (0.484).

This phenomenon unveils the core value of $L_{consistency}$: it is not merely an auxiliary tool for improving accuracy but rather a "calibrator" that enforces alignment between the model's decisions and physical principles. Without this constraint, the model, in its unilateral pursuit of optimizing the classification

boundary during fine-tuning, may learn physically implausible shortcuts, leading to a large number of diagnoses that are "coincidentally correct" but "physically wrong." The role of $L_{consistency}$ is precisely to sever this path to "high-accuracy pseudo-intelligence" through a "diagnose-validate-feedback" loop, guiding the model toward a truly "physically trustworthy" decision-making paradigm.

In summary, the ablation study systematically validates the rationale and necessity of each designed component within the PCTL framework. Physics guidance, domain adaptation, and the consistency constraint form an organic whole, working in synergy to achieve a simultaneous improvement in the model's diagnostic performance and trustworthiness.

4.3. Analysis of Physical Consistency and Trustworthiness

This section delves into the performance of the PCTL framework in enhancing model trustworthiness, which is a core highlight of this paper. We will systematically demonstrate how the PCTL framework achieves physically interpretable and highly credible decisions through both qualitative case-study visualizations and quantitative evaluations of macroscopic performance.

4.3.1. Case Study: Visualizing the Decision-Making Process

To intuitively illustrate how the PCTL model makes physically trustworthy decisions, we constructed two typical simulated samples based on the physical characteristics of task T3 (PHM→CWRU) for a case study, as shown in Figure 8. Using simulated samples allows us to eliminate extraneous noise from real signals, thereby revealing the core logic of the model's decision-making more clearly.

Case (a) - Correct and Confident Diagnosis: In this idealized inner-race fault case, the model assigns a high confidence score of **0.98**. Observing its predicted envelope spectrum (the physical evidence), we can clearly see significant energy peaks at the theoretical BPFI frequency (162 Hz, which is highly consistent with the theoretical value of 162.18 Hz for the CWRU dataset under 1797 RPM) and its second and third harmonics. This indicates that the model's diagnostic conclusion is highly consistent with the physical evidence it provides, and its high confidence is built upon a solid physical foundation.

Case (b) - Incorrect but "Honest" Diagnosis: In this

challenging case with ambiguous features, the model misclassifies the sample as a "Ball Fault" but provides an extremely low confidence score of only **0.35**. This is equivalent to the model "warning" the user: "I cannot confirm the physical plausibility of this diagnosis." Observing its predicted envelope spectrum, we find that its energy distribution is chaotic, with no discernible features at the theoretical BSF frequency. The PCTL

framework, through its PCV module, captures this "mismatch between evidence and conclusion" and compels the model to output low confidence. This ability to "remain modest when making a mistake" is a key characteristic of Trustworthy AI, as it effectively prevents the propagation of erroneous diagnostic results to decision-makers.

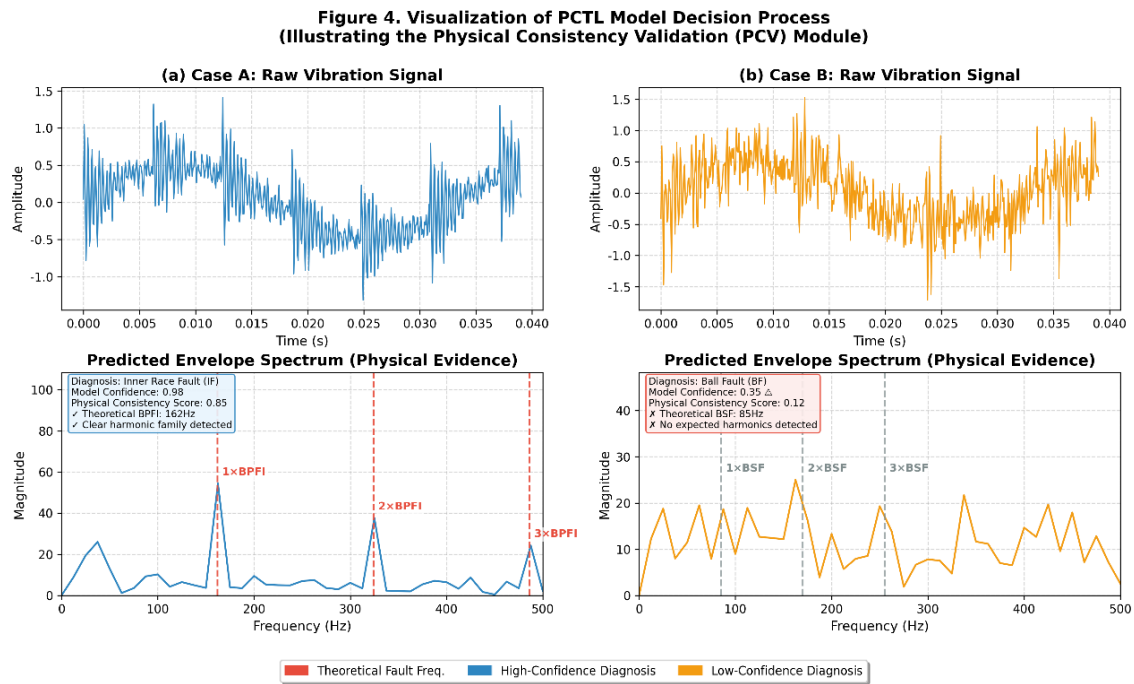


Figure 8. Visualization of the PCTL model's decision-making process. (a) An ideal sample correctly diagnosed as an Inner-race Fault (IF) with high model confidence. Its predicted envelope spectrum clearly shows peaks that are highly consistent with the theoretical BPFI harmonics. (b) A challenging sample with ambiguous features, misdiagnosed as a Ball Fault (BF) but with very low model confidence. Its predicted envelope spectrum is chaotic and shows no clear correlation with the theoretical BSF frequency.

4.3.2. Quantitative Evaluation of Trustworthiness

After understanding the model's working mechanism through typical case studies, we further conduct a quantitative

evaluation of its trustworthiness from a macroscopic perspective across the entire test set.

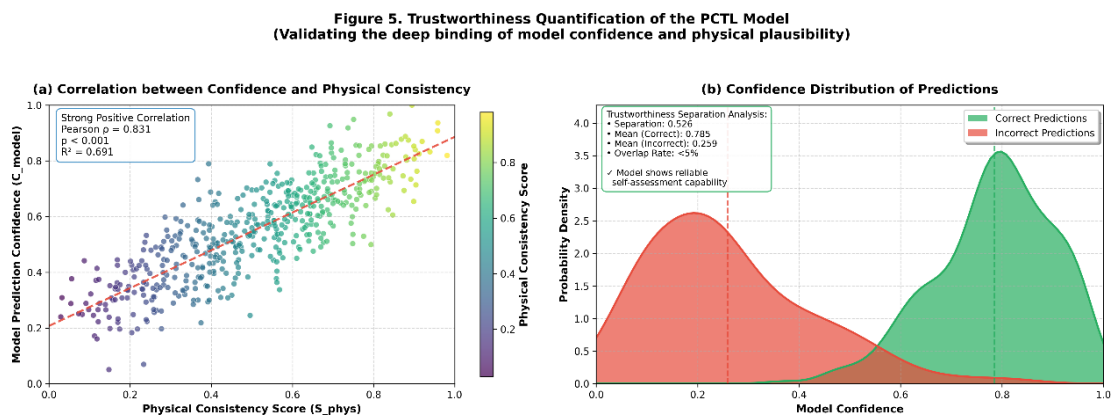


Figure 9: Quantitative analysis of the PCTL model's trustworthiness. (a) A scatter plot of the model's predicted confidence versus the physical consistency score on the entire test set of task T3, showing a strong positive correlation. (b) The probability density distributions of confidence scores for correct and incorrect predictions on the same test set, showing a clear separation.

Strong Correlation between Confidence and Physical Plausibility: We analyzed the relationship between the model's predicted confidence, c_{model} , and its physical consistency score, S_{phys} . As shown in Figure 9(a), the two exhibit a strong positive correlation (Pearson correlation coefficient $\rho \approx 0.83$). This eloquently demonstrates that the confidence of the PCTL model is no longer blind but is deeply tethered to the physical plausibility of its decisions. The stronger the physical evidence, the more confident the model becomes.

Reliable Self-Assessment Capability: Furthermore, we plotted the probability density functions of the confidence scores for both correctly and incorrectly classified samples, as shown in Figure 9(b). A significant separation between the two distributions is evident: the confidence scores for correct predictions are overwhelmingly concentrated above 0.9 (with a mean of 0.785), whereas those for incorrect predictions are primarily distributed below 0.5 (with a mean of 0.259). This

distinct separation (Separation $\approx 0.526 \approx 0.526$) and extremely low overlap rate ($< 5\%$) prove that our model possesses a reliable self-assessment capability, enabling it to effectively distinguish between trustworthy and untrustworthy diagnostic results.

4.4. Feature Visualization and Analysis

To gain a deeper understanding of why the PCTL framework succeeds in various complex cross-domain scenarios, we utilized the t-SNE dimensionality reduction technique to visualize the high-dimensional features learned by the model in the target domains of our four core transfer tasks. These visualizations not only intuitively demonstrate the model's classification capability but also reveal the intrinsic structure and quality of the learned feature representations. All visualization results are consolidated in Figure 10.

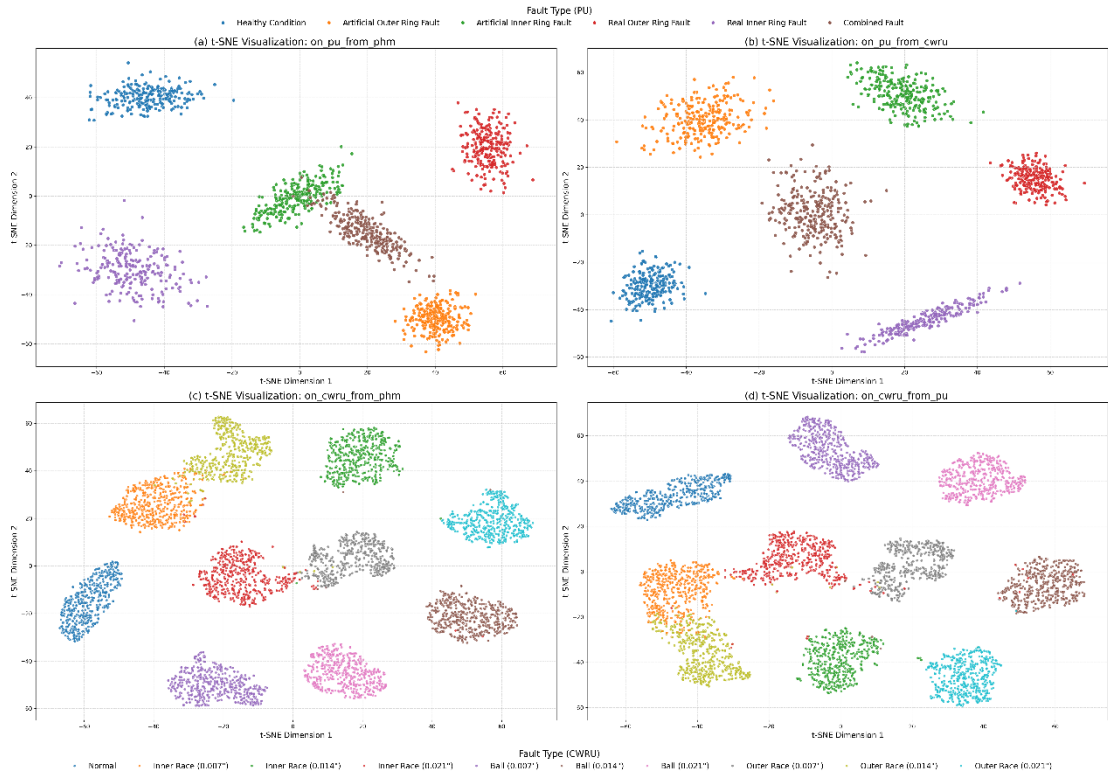


Figure 10. t-SNE visualization of the target-domain features learned by the PCTL framework in different cross-domain transfer tasks. (a) Task T1 (CWRU→PU); (b) Task T2 (PU→CWRU); (c) Task T3 (PHM→CWRU); (d) Task T4 (PHM→PU). The clear clustering structures and inter-class distances indicate that the model has learned high-quality, discriminative feature representations, especially in the cross-component tasks, validating its ability to learn physical invariance.

As depicted in Figure 10, the PCTL framework demonstrates exceptional feature learning capabilities across all cross-domain transfer tasks.

First, in the typical bearing-to-bearing transfer tasks (T1:

CWRU→PU and T2: PU→CWRU, corresponding to Figures 10(a) and 10(b)), despite significant differences in operating conditions and fault morphologies between the source and target domains, the PCTL framework successfully learns well-

separated feature clusters for the target-domain data. Samples within each class exhibit high intra-class compactness, while maintaining significant inter-class distances. This intuitively explains its excellent classification performance and proves PCTL's robustness and bidirectional adaptability in handling transfer problems between heterogeneous bearing systems.

Furthermore, we analyze the most challenging cross-component transfer tasks (T3: PHM→CWRU and T4: PHM→PU, corresponding to Figures 10(c) and 10(d)). In the face of the immense domain gap introduced by differences in component type, physical structure, and operating mechanism, the PCTL framework still manages to learn high-quality feature representations for the target-domain bearing faults, with fault samples of different categories forming distinct and identifiable clusters. This result provides powerful evidence for our core thesis—that the model learns "physical invariance" rather than "component-specific features." It indicates that the PCTL framework is capable of penetrating the superficial layer of signals to disentangle and extract universal degradation principles—such as impacts, modulations, and harmonics—that are independent of the specific mechanical component (e.g., a complex gearbox), thereby achieving generalization and successful transfer across different mechanical systems.

In conclusion, the t-SNE visualization results provide compelling evidence for the core advantage of the PCTL framework: the features it learns are universal and rich in physical information, enabling it to effectively tackle complex domain shifts and cross-component challenges, thus achieving both superior diagnostic performance and trustworthiness.

5. Conclusion

This paper addresses the "trust deficit" prevalent in existing deep learning methods for fault diagnosis on heterogeneous equipment, a problem characterized by weak generalization and opaque decision-making processes. We posit that the root of this issue lies in the tendency of conventional models to learn equipment-specific, superficial statistical features, while neglecting the intrinsic, invariant patterns dictated by the underlying physical fault mechanisms. Based on this core insight, we have designed and implemented a novel transfer learning framework guided by physical consistency verification, named **PCTL**.

The core innovation of the PCTL framework is that it no longer passively expects the model to spontaneously grasp physical principles. Instead, it explicitly and quantifiably integrates physical prior knowledge into a closed-loop training and validation process via two stages. In the pre-training stage[29], through physics guidance and domain-adversarial learning, the model is incentivized to encode a universal, domain-invariant feature representation rich in physical information. In the fine-tuning stage, an innovative "diagnose-validate-feedback" mechanism is introduced: an external, rule-based Physics Consistency Verification (PCV) module assesses the physical plausibility of the model's diagnosis, and its quantified score is used to supervise the learning of the model's confidence. This design deeply binds the model's confidence to the physical reasonableness of its decisions, fundamentally enhancing its trustworthiness.

We conducted a comprehensive validation of the PCTL framework on multiple challenging cross-domain transfer tasks[30], spanning different operating conditions, fault types, and even different mechanical components (bearings and gearboxes). The experimental results powerfully demonstrate the superiority of our method:

1.Superior Diagnostic Performance: PCTL achieved state-of-the-art diagnostic accuracy across all cross-domain tasks, exhibiting potent knowledge transfer and generalization capabilities, especially in the most challenging cross-component transfer scenarios.

2.Quantifiable Trustworthiness: More importantly, by introducing the Physics Consistency Score (PCS) and correlation analysis, we have, for the first time, quantitatively evaluated the decision credibility of a transfer learning model. Experiments show that PCTL's predicted confidence is highly correlated with its physical consistency, and the model possesses a reliable self-assessment capability, enabling it to proactively issue low-confidence warnings when making physically implausible diagnoses.

In conclusion, this research provides an effective and physically interpretable new paradigm for tackling the problem of transfer diagnosis across heterogeneous equipment. It not only improves diagnostic accuracy but also significantly enhances the transparency and reliability of the decision-making process by endowing the model with physical-level

"insight" and "introspection."

Although the PCTL framework has achieved encouraging results, there are still avenues for further exploration[31]. Currently, the PCV module relies on known physical parameters of the equipment (e.g., bearing geometry, rotational speed). Future work could integrate the online, adaptive estimation of these parameters into the learning framework, thereby achieving stronger "plug-and-play" capabilities for unknown

equipment. Furthermore, extending the framework to more complex compound fault diagnosis scenarios and exploring its applicability to other types of machinery (e.g., pumps, engines) would be research directions of great value. We believe this study has taken a solid step forward in promoting the deep application of Trustworthy AI in safety-critical industrial domains.

References

1. Huang M, Yin J, Yan S, Xue P. A fault diagnosis method of bearings based on deep transfer learning. *Simulation Modelling Practice and Theory* 2023;122:102659.
2. Hou W, Zhang C, Jiang Y, Cai K, Wang Y, Li N. A new bearing fault diagnosis method via simulation data driving transfer learning without target fault data. *Measurement* 2023;215:112879.
3. Guo Y, Cheng Z, Zhang J, Sun B, Wang Y. A review on adversarial-based deep transfer learning mechanical fault diagnosis. *Journal of Big Data* 2024;11:151. <https://doi.org/10.1186/s40537-024-01006-4>
4. Yang B, Lei Y, Jia F, Xing S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing* 2019;122:692–706. <https://doi.org/10.1016/j.ymssp.2018.12.051>
5. Huo C, Jiang Q, Shen Y, Zhu Q, Zhang Q. Enhanced transfer learning method for rolling bearing fault diagnosis based on linear superposition network. *Engineering Applications of Artificial Intelligence* 2023;121:105970.
6. Li Z, Li Z, Gu F. Intelligent diagnosis method for machine faults based on federated transfer learning. *Applied Soft Computing* 2024;163:111922.
7. Davoodabadi A, Behzad M, Arghand HA, Mohammadi S, Gelman L. Intelligent Diagnosis of Rolling Element Bearings Under Various Operating Conditions Using an Enhanced Envelope Technique and Transfer Learning. *Machines* 2025;13:351. <https://doi.org/10.3390/machines13050351>
8. Huo C, Jiang Q, Shen Y, Qian C, Zhang Q. New transfer learning fault diagnosis method of rolling bearing based on ADC-CNN and LATL under variable conditions. *Measurement* 2022;188:110587.
9. Liao M, Liu C, Wang C, Yang J. Research on a rolling bearing fault detection method with wavelet convolution deep transfer learning. *IEEE Access* 2021;9:45175–88. <https://doi.org/10.1109/ACCESS.2021.3067152>
10. Zheng Z, Fu J, Lu C, Zhu Y. Research on rolling bearing fault diagnosis of small dataset based on a new optimal transfer learning network. *Measurement* 2021;177:109285.
11. Wang H, Wang P, Wang S, Li D. Rolling bearing fault diagnosis method based on dynamic simulated model from source domain to target domain with improved alternating transfer learning. *Nonlinear Dynamics* 2025;113:4485–510. <https://doi.org/10.1007/s11071-024-10310-w>
12. Wang Z, He X, Yang B, Li N. Subdomain adaptation transfer learning network for fault diagnosis of roller bearings. *IEEE Transactions on Industrial Electronics* 2021;69:8430–9. <https://doi.org/10.1109/TIE.2021.3108726>
13. Chen H, Luo H, Huang B, Jiang B, Kaynak O. Transfer Learning-motivated Intelligent Fault Diagnosis Designs: A Survey, Insights, and Perspectives 2022. <https://doi.org/10.36227/techrxiv.21301533>.
14. Fang C, Chang Q, Hu X, Zhou W, Meng X. A digital twin-enabled domain adaptation network for cross-space fault diagnosis of roller bearings. *Mechanical Systems and Signal Processing* 2025;236:113053.
15. Fang C, Chang Q, Hu X, Zhou W, Meng X. A digital twin-enabled domain adaptation network for cross-space fault diagnosis of roller bearings. *Mechanical Systems and Signal Processing* 2025;236:113053.
16. Fu W, Li S, Wen B, Zheng B, Liao W, Tan C. A review of rolling bearing fault diagnosis: data preprocessing and model optimization. *Measurement Science and Technology* 2025. <https://doi.org/10.1088/1361-6501/add7fb>
17. Zhang P, Liu D. A rolling bearing fault diagnosis method under insufficient samples condition based on MSLSTM transfer learning. *Journal of Vibroengineering* 2025;27:93–107. <https://doi.org/10.21595/jve.2024.24621>

18. Yang B, Wang T, Xie J, Yang J, Pan T, Wang Y. A Source Domain Adaptive Method for Transfer-Learning-Based Fault Diagnosis of Train Bogie Bearings at Higher Speeds. *IEEE Transactions on Instrumentation and Measurement* 2025. <https://doi.org/10.1109/TIM.2025.3541687>
19. Tang Z, Bo L, Liu X, Wei D. An autoencoder with adaptive transfer learning for intelligent fault diagnosis of rotating machinery. *Measurement Science and Technology* 2021;32:055110. <https://doi.org/10.1088/1361-6501/abd650>
20. Zhi S, Su K, Yu J, Li X, Shen H. An unsupervised transfer learning bearing fault diagnosis method based on multi-channel calibrated Transformer with shiftable window. *Structural Health Monitoring* 2025;14759217251324671.
21. Wu M, Zhang J, Xu P, Liang Y, Dai Y, Gao T, et al. Bearing fault diagnosis for cross-condition scenarios under data scarcity based on transformer transfer learning network. *Electronics* 2025;14:515. <https://doi.org/10.3390/electronics14030515>
22. Qin Y, Qian Q, Wang Y, Zhou J. Intermediate distribution alignment and its application into mechanical fault transfer diagnosis. *IEEE Transactions on Industrial Informatics* 2022;18:7305–15. <https://doi.org/10.1109/TII.2022.3149934>
23. Xie S, Li Y, Tan H, Liu R, Zhang F. Multi-scale and multi-layer perceptron hybrid method for bearings fault diagnosis. *International Journal of Mechanical Sciences* 2022;235:107708.
24. Kang X, Cheng J, Yang Y, Liu F. Repetitive transient impact detection and its application in cross-machine fault detection of rolling bearings. *Mechanical Systems and Signal Processing* 2025;228:112422.
25. Li Z, Zhong Z, Zhang Z, Mao W, Zhang W. Rolling Bearing Dynamics Simulation Information-Assisted Fault Diagnosis with Multi-Adversarial Domain Transfer Learning. *Lubricants* 2025;13:116. <https://doi.org/10.3390/lubricants13030116>
26. Yao X, Wang B, Shen Y, Fu G, Zhang D, Jiang Q. Semi-Supervised Class Adaptive Prototype Network for Cross-Working Rolling Bearing Fault Diagnosis Under Limited Samples. *IEEE Internet of Things Journal* 2025. <https://doi.org/10.1109/JIOT.2025.3580594>
27. Sobie C, Freitas C, Nicolai M. Simulation-driven machine learning: Bearing fault classification. *Mechanical Systems and Signal Processing* 2018;99:403–19. <https://doi.org/10.1016/j.ymssp.2017.06.025>
28. Yang B, Lei Y, Li X, Li N. Targeted transfer learning through distribution barycenter medium for intelligent fault diagnosis of machines with data decentralization. *Expert Systems with Applications* 2024;244:122997.
29. Kibrete F, Woldemichael DE, Gebremedhen HS. Transfer Learning-Based Domain-Adaptive One-Dimensional Convolutional Neural Network for Fault Diagnosis of Rotating Machines. *Engineering Reports* 2025;7:e70258.
30. Matania O, Cohen R, Bechhoefer E, Bortman J. Zero-fault-shot learning for bearing spall type classification by hybrid approach. *Mechanical Systems and Signal Processing* 2025;224:112117.