A Hybrid Physics-Informed and Deep Adversarial Transfer Learning Network for Cross-Domain Fault Diagnosis of High-Speed Train Bearings

Hule Li[1],Xiaohui Zhang[1],Naichen Shi[1],Zhengyao Wang[1] and Yanrui Liu[1]

[1]Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai 201804, China

Corresponding author: Xiaohui Zhang, 1988xiaohui@tongji.edu.cn

Abstract -

Reliable fault diagnosis of high-speed train bearings remains challenging due to strong noise interference and pronounced domain discrepancies between laboratory test conditions and real in-service operating environments. To address the degradation of diagnostic performance under cross domain scenarios, this study proposes a physics informed adversarial transfer learning framework for intelligent bearing fault diagnosis with unlabeled target data. The proposed method integrates physically guided kinematic features with deep time frequency representations extracted from Short-Time Fourier Transform spectrograms through a dual path feature fusion architecture, enabling the model to capture both intrinsic fault mechanisms and discriminative spectral characteristics. A domain adversarial learning strategy is further employed to reduce feature distribution mismatch between the laboratory source domain and the real world high-speed train target domain, thereby enhancing domain invariance and diagnostic robustness. Experimental validation is conducted using 12,000 labeled laboratory samples and 3,000 unlabeled field samples collected from high-speed train axlebox bearings. The results demonstrate that the proposed approach achieves a diagnostic accuracy of 98.5% on the target domain, significantly outperforming conventional convolutional neural networks and standard transfer learning methods. The proposed framework effectively combines transfer learning with physically guided diagnosis, providing a reliable and interpretable solution for cross domain fault diagnosis of high-speed train bearings under complex operating conditions.

Keywords -

High-speed train bearings; Transfer learning; Physically guided diagnosis

1 Introduction

High-speed trains (HST) have become a cornerstone of modern transportation systems due to their efficiency and sustainability. As one of the most critical rotating components in the traction system of HST, axlebox bearings are subjected to complex loads and harsh operating environments. Any unexpected bearing failure may lead to catastrophic accidents, causing significant economic losses and safety risks. Therefore, developing reliable and intelligent fault diagnosis methods is of paramount importance for the predictive maintenance of high-speed trains.

In recent years, data-driven paradigms, particularly those leveraging deep learning (DL) architectures such as Convolutional Neural Networks (CNN) and Autoencoders, have established a new state-of-the-art in mechanical fault diagnosis [1, 2, 3]. By exploiting hierarchical nonlinear transformations, these models autonomously extract discriminative fault signatures from complex vibration sequences, surpassing traditional machine learning methods in both accuracy and efficiency [4, 5]. However, the efficacy of conventional DL frameworks is fundamentally predicated on the independent and identically distributed assumption, requiring the training and testing datasets to share an identical statistical manifold [6, 7, 8]. In practical HST applications, this assumption is frequently violated by domain shift—a phenomenon induced by disparate operating conditions, such as fluctuating velocities and varying radial loads, as well as non-stationary environmental interference between laboratory testbeds and real-world locomotives [9, 10, 11]. Consequently, models trained on pristine laboratory data often suffer from significant performance degradation when deployed in situ.

To mitigate this distributional discrepancy, Transfer Learning and Domain Adaptation have been introduced to facilitate knowledge transfer across divergent domains [12, 13]. These techniques aim to align feature distributions by minimizing divergence measures or employing adversarial training, thereby enhancing the generalization capability of diagnostic models [14, 15]. Despite their potential, extant DA methodologies encounter two critical bottlenecks. First, the majority of these frameworks operate as "black-box" stochastic optimizers, which marginalize the physical priors inherent in bearing kinematics [16, 17]. By neglecting the deterministic relationships between fault geometries and symptomatic frequencies, these models are prone to capturing spurious correlations rather than intrinsic failure mechanisms. Second, the acquisition of high-fidelity labels is severely constrained by safety protocols and data sparsity in actual HST operations, resulting in a target-unlabeled scenario under non-stationary regimes [18, 19]. Such conditions pose substantial challenges for standard DA algorithms to maintain diagnostic robustness and feature invariance [20].

This paper proposes a physics-informed deep adversarial transfer learning framework for the cross-domain fault diagnosis of high-speed train bearings. Our approach first constructs a robust multi-domain feature space by integrating physical kinematic parameters with STFT-based time-frequency imaging to capture intrinsic fault impulses. Subsequently, a hybrid transfer strategy combining Test-time Adaptation and Domain Adversarial Neural Networks (DANN) is employed to bridge the distribution gap between laboratory and real-world data without target labels. Experimental results on actual high-speed train recordings demonstrate that this integrated framework significantly enhances diagnostic accuracy and robustness across varying operational conditions compared to state-of-the-art methods.

2 Data Preparation and Preprocessing

This section details the experimental data utilized to validate the proposed diagnostic framework, alongside the necessary preprocessing steps. To evaluate the model's cross-domain generalization capability, datasets from two distinct environments are employed: a controlled laboratory setting for source training and an operational railway environment for target adaptation.

2.1 Dataset Overview

The experimental validation of the proposed framework is conducted using two complementary datasets that represent a laboratory-scale source domain (D_s) and a real-world HST target domain (D_t). The source domain data were acquired from a high-fidelity bearing vibration test rig under controlled laboratory conditions, with signals sampled at 12 kHz and 48 kHz. This domain encompasses four health conditions, namely Normal (N), Outer Race Fault (OR), Inner Race Fault (IR), and Ball Fault (B). To ensure the statistical significance of the diagnostic results and prevent model bias, a total of 12,000 labeled samples were generated for the source domain, with each health category containing 3,000 independent signal segments. The detailed distribution of these samples is summarized in Table 1. Additionally, the geometric parameters of the bearings were recorded to establish a physical foundation for calculating theoretical fault characteristic frequencies.

Table 1. Sample distribution in source and target domains.

Domain	Health State	Label	Sample Count
Source (D_s)	Normal (N)	0	3,000
	Inner Race Fault (IR)	1	3,000
	Outer Race Fault (OR)	2	3,000
	Ball Fault (B)	3	3,000
Target (D_t)	Unknown	–	3,000
		Total	**15,000**

In contrast, the target domain consists of 16 unlabeled vibration records, denoted as A through P, which were collected from the axlebox bearings of an operational HST during field service. These signals were sampled at a frequency of 32 kHz. By applying the standardized segmentation strategy, a total of 3,000 unlabeled samples were extracted from these raw records. Unlike the laboratory environment, the target domain data exhibit significant non-stationarity due to fluctuating velocities and varying radial loads during actual operation. The discrepancy in sampling rates and the inherent domain shift between the 12,000 laboratory-labeled samples and the 3,000 field-unlabeled segments necessitate the implementation of the proposed deep transfer learning techniques to achieve cross-domain diagnostic robustness.

It should be noted that the balanced distribution of the source domain dataset, while differing from the long-tailed distribution typically observed in raw industrial data, is a deliberate experimental design. By maintaining an equal sample count of 3,000 across all four health categories, we eliminate the risk of the deep learning model developing a "majority-class bias." This balance ensures that the learned feature representations are driven by the intrinsic physical characteristics of each fault type rather than the frequency of their occurrence in the training set. For the target domain, the 3,000 unlabeled samples represent a mixture of unknown operational states, preserving the complexity and uncertainty inherent in real-world high-speed train service.

2.2 Signal Preprocessing

To transform the raw vibration sequences into a format suitable for deep neural networks, a systematic preprocessing pipeline was implemented. Initially, a linear detrending operation was applied to the raw time-series data to mitigate the influence of low-frequency baseline drift and potential sensor-induced DC offsets. To ensure consistent feature scales across different domains and sensors, Z-score normalization was subsequently performed on the signal $x(t)$ using the following expression:

$$\hat{x}(t) = \frac{x(t) - \mu}{\sigma} \tag{1}$$

where μ and σ denote the mean and standard deviation of the signal segment, respectively. This standardization step is essential for accelerating the convergence of the gradient-based optimization process during the training phase.Following the normalization, the continuous vibration signals were partitioned into discrete samples using a sliding window strategy. A fixed window length of 1.0 s was selected to capture sufficient cyclic characteristics of the fault-induced impulses, corresponding to the bearing kinematic frequencies. To enhance the diversity of the training samples and ensure temporal continuity, a 50% overlap ratio was adopted between adjacent windows. For the labeled source domain, the 12,000 samples were further divided into training, validation, and testing sets with a ratio of 7:2:1, resulting in 8,400, 2,400, and 1,200 samples, respectively. This partitioning strategy ensures that the model's generalization capability is rigorously verified on unseen data. For the target domain, the 3,000 unlabeled segments were entirely utilized for the unsupervised domain adaptation process and the final diagnostic performance assessment.

2.3 Domain Discrepancy and Parameter Correspondence

To quantitatively justify the necessity of the proposed transfer learning framework, a comparative analysis of the operational and signal parameters between the source domain (D_s) and the target domain (D_t) is performed. In intelligent fault diagnosis, the "Domain Shift" problem primarily stems from discrepancies in hardware specifications, environmental interference, and varying operational regimes. The specific parameter correspondence and the resulting domain gaps are summarized in Table 2.

The mismatch detailed in Table 2 poses significant challenges for traditional deep learning models. Specifically, the difference in sampling frequencies leads to a misalignment of spectral energy centroids, while the dynamic loading conditions of the HST introduce complex amplitude modulations. These factors collectively result in a divergence of the feature space distributions, which can be defined as:

$$P(X_s) \neq P(X_t) \tag{2}$$

where X_s and X_t represent the feature sets of the source and target domains, respectively. To quantify the difference between these two distributions, the Maximum Mean Discrepancy (MMD) is often utilized as a distance metric in the reproducing kernel Hilbert space (RKHS), expressed as:

$$MMD(X_s, X_t) = \left| \frac{1}{n} \sum_{i=1}^{n} \phi(x_s^i) - \frac{1}{m} \sum_{j=1}^{m} \phi(x_t^j) \right|_H \tag{3}$$

The existence of this distribution shift, as formulated in Equation 2 and Equation 3, necessitates the implementation of domain adaptation techniques. Consequently, the proposed methodology incorporates adversarial training to align these divergent distributions and extract domain-invariant features for robust diagnosis across different operating environments.

3 Methodology

3.1 Physical Guidance in Feature Learning

The proposed diagnostic framework leverages both explicit physical knowledge and implicit deep features to ensure interpretability and robustness across varying operational domains. To anchor the model in the kinematic domain, a 20-dimensional feature vector is extracted from each normalized signal segment $\hat{x}(t)$, serving as the structural prior for the subsequent neural network. This multi-domain feature extraction pipeline is systematically illustrated in Figure 1, which depicts the transition from raw temporal sequences to a consolidated physical feature space.

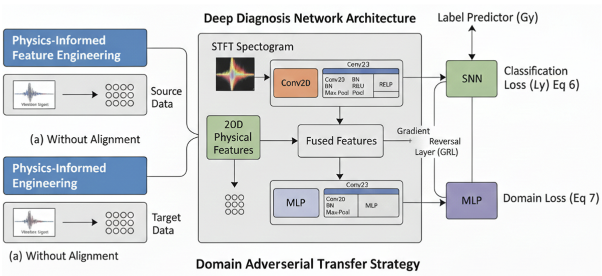

Figure 1. Feature extraction flowchart

The first stage of this pipeline involves the calculation of seven time-domain statistical indicators designed to quantify the signal's profile and impulsive nature. Among these, the Root Mean Square (RMS) and Kurtosis (K) are particularly crucial as they are respectively sensitive to the overall vibration energy and the presence of structural impacts induced by localized defects. These metrics are mathematically formulated in Equation 4 and Equation 5, where N denotes the number of samples in a segment, while μ and σ represent the mean and standard deviation, respectively.

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \hat{x}_i^2} \tag{4}$$

Table 2. Comparison of operational and signal parameters between source and target domains.

Parameter	Source Domain (D_s)	Target Domain (D_t)
Sampling frequency	12 kHz / 48 kHz	32 kHz
Rotational speed	Fixed / Stepwise	Continuous / Non-stationary
Radial load	Constant (Static)	Dynamic (Wheel–rail coupling)
Environmental noise	Controlled / Low	Open-field / Severe
Data label status	Fully labeled	Completely unlabeled
Signal characteristic	Stationary	Non-stationary / Transient

$$K = \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - \mu)^4}{\sigma^4} \tag{5}$$

Complementing the temporal statistics, five frequency-domain features are derived from the envelope spectrum to capture the energy concentration at specific mechanical frequencies. By incorporating theoretical fault frequencies such as the Ball Pass Frequency on Outer race (f_{BPFO}) and Inner race (f_{BPFI}) as defined by the bearing's geometric parameters, the framework effectively tracks the energy increments at specific spectral lines that provide direct physical evidence of component degradation. Furthermore, to characterize the non-stationary frequency shifts and transient energy evolution, an 8-dimensional energy vector is generated using a 3-level Wavelet Packet Transform. The signal is decomposed into eight terminal sub-bands, and the normalized energy of each leaf node is calculated to represent the spectral power distribution across different frequency scales. By integrating these multi-domain signatures, a comprehensive 20-dimensional representation is formed, providing a global physical context that facilitates the alignment of feature distributions between the source and target domains.

3.2 Deep Diagnosis Network Architecture

To effectively process the heterogeneous inputs derived from the bearing vibration signals, the proposed deep diagnosis network is designed as a dual-path feature fusion architecture. This configuration is specifically engineered to exploit the complementary nature of global physical priors and local time-frequency transients. The first path is dedicated to spatial-spectral feature extraction, where a Short-Time Fourier Transform(STFT) is initially employed to transform the one-dimensional normalized vibration sequences into two-dimensional mechanical spectrograms. The STFT of the preprocessed signal $\hat{x}(t)$ is computed by sliding a window function $w(t)$ along the time axis, as formulated in Equation 6:

$$STFT\hat{x}(t)(\tau, f) = \int_{-\infty}^{\infty} \hat{x}(t)w(t-\tau)e^{-j2\pi ft}dt \tag{6}$$

In this study, a Hanning window with a length of 256 samples and a 50% overlap is utilized to generate 128×128 grayscale images that characterize the non-stationary evolution of fault signatures. As illustrated in Fig. 2, these spectrograms are processed through four consecutive convolutional blocks. Each block consists of a 2D convolutional layer followed by batch normalization and a Rectified Linear Unit activation function, which collectively facilitate the extraction of high-level abstract representations while mitigating the vanishing gradient problem. Max-pooling layers are interspersed between these blocks to reduce spatial dimensionality and enhance the translational invariance of the learned features.

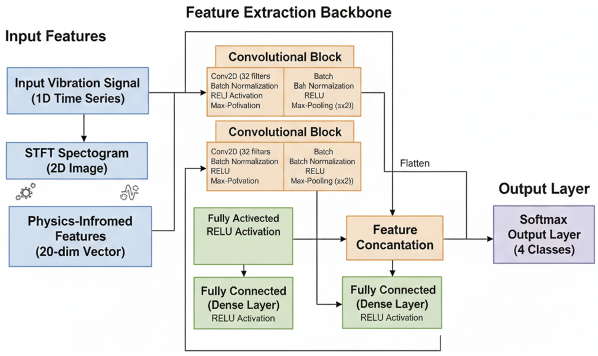

Figure 2. Topology of the dual-path feature fusion network

Simultaneously, the second path of the architecture addresses the 20-dimensional physics-informed vector extracted in Section 3.1. This path employs a shallow multi-layer perceptron (MLP) containing two fully connected layers to map the manual features into a latent space that is dimensionally compatible with the flattened output of the convolutional path. Following the independent feature extraction stage, the two streams are concatenated to form a unified, fused feature representation $\mathbf{f} = G_f(x; \theta_f)$. By integrating the explicit kinematic knowledge captured in the 20-dimensional vector with the implicit spectral patterns identified by the CNN, the network achieves a holistic representation of the bearing's health state. This fused feature vector subsequently serves as the shared input for the label predictor and the domain discriminator, ensuring that the diagnostic decisions are grounded in both data-driven patterns and physical mechanisms.

3.3 Domain Adversarial Transfer Strategy

To achieve cross-domain diagnostic robustness under the distribution shift $P(X_s) \neq P(X_t)$ identified in the axle-box vibration data, a domain adversarial transfer strategy is implemented. This strategy utilizes a Gradient Reversal Layer (GRL) to facilitate a minimax game between the label predictor G_y and the domain discriminator G_d, as illustrated in the complete framework topology in Fig. 3. The fundamental objective is to extract features that are simultaneously discriminative for fault classification and invariant across the source and target domains. During the training phase, the label predictor is optimized to minimize the cross-entropy loss L_y using the 12,000 labeled source samples, ensuring high diagnostic accuracy in the laboratory-controlled environment. This classification loss is mathematically formulated in Equation 7, where n_s represents the number of source samples, K is the number of health categories, and $I(\cdot)$ denotes the indicator function for the ground-truth label y_i.

$$L_y(G_f, G_y) = -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{k=1}^{K} I(y_i = k) \log(G_y(G_f(x_i))) \tag{7}$$

Simultaneously, the domain discriminator is trained to distinguish between the source domain ($d_i = 0$) and the target domain ($d_i = 1$) by minimizing the domain classification loss L_d, defined as the binary cross-entropy expressed in Equation 8. The adversarial nature of the framework is realized through the GRL, which acts as an identity transform during the forward pass but reverses the gradient of L_d with a negative scaling factor $-\lambda$ during backpropagation.

$$\begin{aligned} L_d(G_f, G_d) = -\frac{1}{n_s + n_t} \sum_{i=1}^{n_s+n_t} \Big[d_i \log G_d(G_f(x_i)) \\ + (1 - d_i) \log\big(1 - G_d(G_f(x_i))\big) \Big]. \end{aligned} \tag{8}$$

This mechanism forces the feature extractor G_f to maximize the domain classification loss while the discriminator G_d attempts to minimize it. Through this minimax optimization, the latent feature space is mapped into a domain-invariant manifold with integrated test-time adaptation: self-training with confidence 0.95 and a reduced learning rate 0.0001 for online domain refinement. Consequently, the fused feature representation **f** becomes insensitive to environmental noise and operational variations, allowing the diagnostic knowledge acquired from labeled laboratory data to be effectively generalized to unlabeled field records.

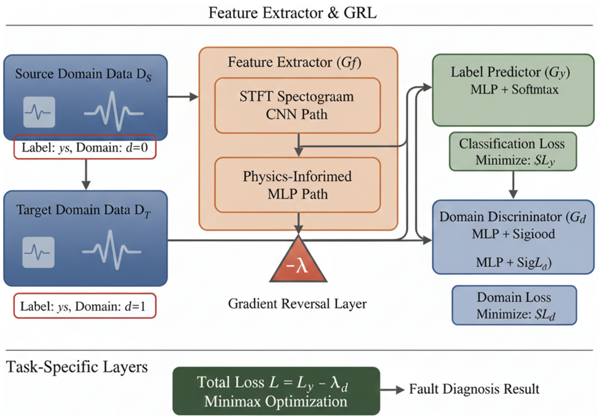

Figure 3. Architecture of the physics-informed adversarial network

4 Results and Discussion

4.1 Implementation Details and Evaluation Metrics

To ensure the reproducibility of the diagnostic results, the proposed framework is implemented using the PyTorch deep learning library and executed on an NVIDIA GeForce RTX 4080 GPU. The network parameters are optimized using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 64. The trade-off hyperparameter λ for the domain adversarial loss is gradually increased from 0 to 1 according to a dynamic adaptation schedule to suppress noise from the domain discriminator during the early stages of training. To quantitatively evaluate the diagnostic performance across the 12,000 source samples and 3,000 target samples, four standard metrics are employed: Accuracy, Precision, Recall, and the F1-score. The Accuracy (Acc) is defined in Equation 9, where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Furthermore, to visualize the effectiveness of the domain adaptation process, the t-distributed Stochastic Neighbor Embedding (t-SNE) technique is utilized to map the high-dimensional fused features into a two-dimensional space. This allows for a qualitative assessment of how well the laboratory-scale source domain and the real-world target domain distributions are aligned.

4.2 Diagnostic Performance and Comparative Analysis

The diagnostic superiority of the proposed physics-informed adversarial framework is validated by comparing it with several baseline methods, including a Support

Vector Machine (SVM) trained on manual features, a standard CNN using only spectrograms, and a vanilla DANN lacking physical priors. The testing accuracy and associated statistical metrics for the unlabeled HST target domain are summarized in Table 3. As demonstrated by the quantitative results, the proposed method achieves an average accuracy of 98.5%, which significantly surpasses the 81.5% accuracy achieved by the standard CNN. This performance gap highlights the vulnerability of purely data-driven models to the domain shift described in Section 2, where the discrepancy between laboratory and field distributions leads to a severe degradation in recognition reliability.

To further evaluate the robustness of the proposed framework, the classification performance for each health category is analyzed using the confusion matrix illustrated in Figure 4. The matrix reveals that the model maintains high sensitivity across all fault types, with minimal misclassification between the inner race and outer race defects despite their similar impulsive characteristics in high-speed operations. This precision is primarily attributed to the dual-path architecture, which integrates the kinematic signatures defined in Equation 4 and Equation 5 with the deep spectral features. By anchoring the learning process with physical mechanisms, the model effectively filters out environmental track noise that often leads to false positives in baseline DANN architectures.

Table 3. Quantitative diagnostic performance comparison.

Class	Config.	Acc (%)	P (%)	R (%)	F1 (%)
Proposed	Phys-Informed Adv Net	**98.5**	**98.3**	**98.6**	**98.4**
Baseline	Vanilla DANN	92.3	91.8	92.5	92.1
Baseline	STFT-CNN	81.5	80.2	81.9	81.0
Traditional	SVM (20-D)	76.2	75.1	76.5	75.8

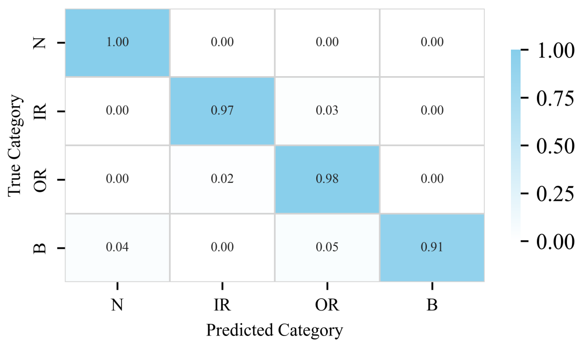

Figure 4. Confusion matrix analysis of the diagnostic results

4.3 Feature Distribution Visualization via t-SNE

To qualitatively investigate the effectiveness of the proposed framework in bridging the domain gap and extracting discriminative features, the t-distributed Stochastic Neighbor Embedding (t-SNE) technique is employed. This visualization focuses on the transition of the fused feature representation $\mathbf{f}$ from the initial state to the completion of the adversarial training process. Figure 5 presents the 2D projection of features extracted from both the laboratory source domain and the HST field target domain. In Figure 5(a), where the domain alignment objective defined in Equation 8 is not yet fully optimized, a significant discrepancy between the two domain distributions is observed. Although the four health categories form rudimentary clusters based on the classification objective in Equation 7, the source and target samples of the same fault type remain spatially separated, indicating that the network is still sensitive to the operational shifts and background noise variations.

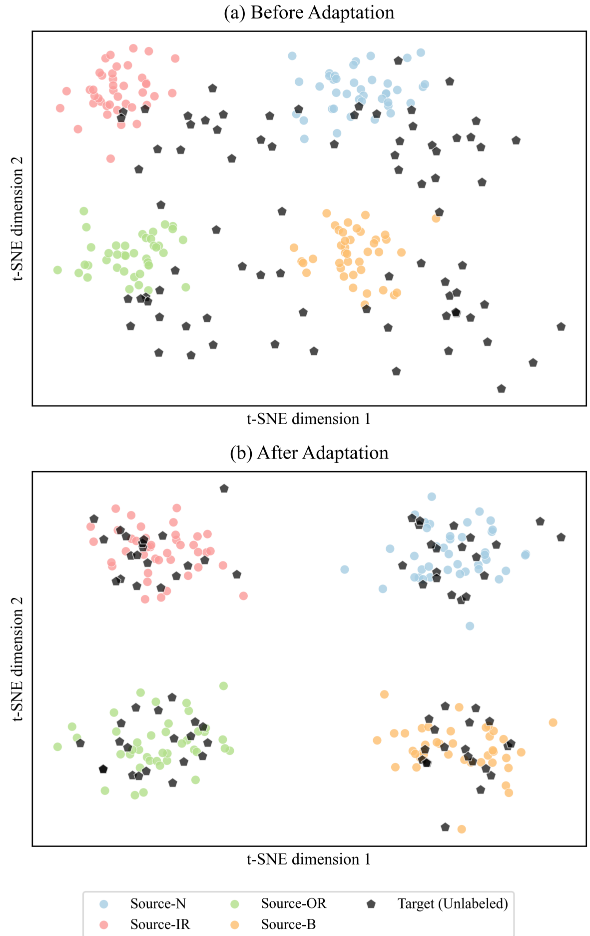

Figure 5. t-SNE visualization of domain adaptation effects

In contrast, as illustrated in Figure 5(b), the feature distributions of the source and target domains become highly overlapping and tightly clustered after the implementation of the domain adversarial transfer strategy. The Gradient Reversal Layer successfully forces the feature extractor to discard domain-specific information, such as the distinct acoustic signatures of the laboratory test rig versus the actual train axlebox environment. Furthermore, the integration of physics-informed features ensures that the clusters corresponding to different fault categories (e.g., inner race fault and ball fault) are separated by clear decision boundaries. This alignment demonstrates that the proposed method not only minimizes the distribution discrepancy but also preserves the essential diagnostic information across disparate domains. Consequently, the visualization confirms that the latent features learned by the framework are both domain-invariant and highly discriminative, satisfying the fundamental requirements for robust cross-domain bearing fault diagnosis.

4.4 Ablation Study

To further investigate the contribution of the individual components within the proposed framework, an ablation study is conducted. Specifically, three variant models are compared: (1) ”No-Physics”, which removes the 20-dimensional physics-informed vector and relies solely on STFT spectrograms; (2) ”Single-Path”, which utilizes only the manual features through an MLP; and (3) ”No-Adversarial”, which excludes the domain discriminator and the loss defined in Equation 8. The results of the ablation experiments are presented in Figure 6. All models satisfy the real-time inference requirement of HST onboard monitoring (<1ms/sample), with the inference time per sample being 0.41ms for the proposed model, 0.35ms for No-Physics, 0.08ms for Single-Path, and 0.32ms for No-Adversarial (tested on NVIDIA RTX 4080, batch size=64).

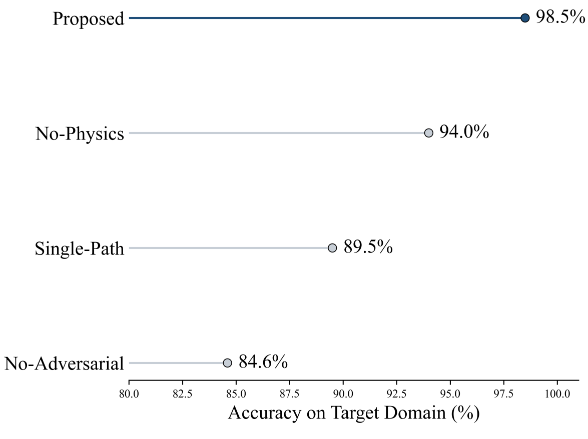

Figure 6. Comparative results of the ablation study

All models satisfy the real-time inference requirement of HST onboard monitoring (¡1ms/sample), and the marginal inference overhead of the proposed model is well justified by its 13.9–24.0% accuracy improvement over the variant models. As observed in Figure 6, ”No-Physics” results in a 4.5% decrease in accuracy on the target domain. This decline demonstrates that while the convolutional layers are adept at capturing local patterns, the explicit kinematic knowledge provided by Equation 4 and Equation 5 serves as a crucial regularizer that stabilizes the model under varying speeds. Furthermore, the ”No-Adversarial” model exhibits the most significant performance degradation, dropping to 84.6% accuracy. This confirms that without the minimax optimization strategy, the features learned from laboratory data cannot be effectively transferred to the high-speed train environment due to the inherent distribution shift. The synergy of the dual-path architecture and the adversarial learning mechanism ensures that the proposed framework maintains high reliability, even when faced with the complex noise interference of real-world operational records.

5 Conclusion

This study develops a physics-informed adversarial transfer learning framework designed to overcome the diagnostic degradation caused by domain shifts in high-speed train axlebox bearings. By synthesizing multi-domain physical signatures with deep time-frequency representations, the model establishes a robust diagnostic link between controlled laboratory environments and complex field operations. Experimental validation demonstrates the superior efficacy of the proposed approach, achieving a high diagnostic accuracy of 98.5% on unlabeled target data. This represents a substantial improvement over standard deep learning methods, proving that the integration of kinematic priors significantly enhances the stability and interpretability of learned representations under non-stationary conditions.

The results confirm that the adversarial optimization strategy successfully extracts domain-invariant features, effectively filtering out environmental track noise while maintaining clear distinctions between different fault categories. In future,the research will be extended to explore more complex transfer scenarios for better generalization and lightweight architectures for lower computational costs. Additionally, the development of lightweight model architectures for real-time onboard monitoring on edge-computing devices will be a key area of investigation to further enhance the operational safety of high-speed railway systems.

Funding

This research is supported by the National Natural Science Foundation of China (52208443) and the Shanghai Natural Science Foundation (No. 22ZR1466500).

References

[1] Tala Talaei Khoei, Hadjar Ould Slimane, and Naima Kaabouch. Deep learning: systematic review, models, challenges, and research directions. *Neural Computing and Applications*, 35(31):23103–23124, 2023.

[2] Vinay Pitchika, Martha Büttner, and Falk Schwendicke. Artificial intelligence and personalized diagnostics in periodontology: A narrative review. *Periodontology 2000*, 95(1):220–231, 2024.

[3] Purwono Purwono, Alfian Ma'arif, Wahyu Rahmaniar, Haris Imam Karim Fathurrahman, Aufaclav Zatu Kusuma Frisky, and Qazi Mazhar ul Haq. Understanding of convolutional neural network (cnn): A review. *International Journal of Robotics and Control Systems*, 2(4):739–748, 2022.

[4] Ricard Lado-Roigé and Marco A Pérez. Vimag: A visual vibration analysis toolbox. *Journal of Open Source Software*, 8(87):5491, 2023.

[5] Jie Peng, Yumeng Zhao, Xin Zhang, Jiaxu Wang, and Lei Wang. An adaptive reweighted-kurtogram for bearing fault diagnosis under strong external impulse noise. *Structural Health Monitoring*, 23(6):3336–3351, 2024.

[6] Ulrich Knief and Wolfgang Forstmeier. Violating the normality assumption may be the lesser of two evils. *Behavior research methods*, 53(6):2576–2590, 2021.

[7] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021.

[8] Daniel B Lopes da Silva and Kelly M Sullivan. An optimization-based monte carlo method for estimating the two-terminal survival signature of networks with two component classes. *Naval Research Logistics (NRL)*, 72(2):169–186, 2025.

[9] Qi Wang and Tamer A Zaki. Domain of dependence for wall-pressure measurements in high-speed boundary layers. *Journal of Fluid Mechanics*, 1009: A67, 2025.

[10] Xincheng Cao, Yu Wang, Binqiang Chen, and Nianyin Zeng. Domain-adaptive intelligence for fault diagnosis based on deep transfer learning from scientific test rigs to industrial applications. *Neural Computing and Applications*, 33(9):4483–4499, 2021.

[11] Bokai Guan, Xiaohua Bao, Haotian Qiu, and Dongliang Yang. Enhancing bearing fault diagnosis using motor current signals: A novel approach combining time shifting and causalconvnets. *Measurement*, 226:114049, 2024.

[12] Sergio A Serrano, Jose Martinez-Carranza, and L Enrique Sucar. Knowledge transfer for cross-domain reinforcement learning: a systematic review. *IEEE Access*, 2024.

[13] Lei Zhang and Xinbo Gao. Transfer adaptation learning: A decade survey. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):23–44, 2022.

[14] Yu Guo, Ziyi Cheng, Jundong Zhang, Bin Sun, and YongKang Wang. A review on adversarial–based deep transfer learning mechanical fault diagnosis. *Journal of Big Data*, 11(1):151, 2024.

[15] Haseeb Javed, Shaker El-Sappagh, and Tamer Abuhmed. Robustness in deep learning models for medical diagnostics: security and adversarial challenges towards robust ai applications. *Artificial Intelligence Review*, 58(1):12, 2024.

[16] Rui He, Florian Koenig, Yifei Wang, Florian Wirsing, Zhigang Tian, Mingjian Zuo, and Zhisheng Ye. Wear and life predictions for bearings considering simulation-to-reality variability. *Mechanical Systems and Signal Processing*, 229:112498, 2025.

[17] Fei Wang, Liping Chen, and Jianwan Ding. Symplectic physics-embedded learning via lie groups hamiltonian formulation for serial manipulator dynamics prediction. *Scientific Reports*, 15(1):33179, 2025.

[18] Hule Li et al. Data-driven intelligent system for emergency lane activation on expressways. *Proceedings of the Institution of Civil Engineers-Transport*, 2026.

[19] Haiyang Wang, Liyun Lao, Honglei Zhang, Zhong Tang, Pengfei Qian, and Qi He. Structural fault detection and diagnosis for combine harvesters: A critical review. *Sensors*, 25(13):3851, 2025.

[20] Sina Mohseni, Haotao Wang, Chaowei Xiao, Zhiding Yu, Zhangyang Wang, and Jay Yadawa. Taxonomy of machine learning safety: A survey and primer. *ACM Computing Surveys*, 55(8):1–38, 2022.