Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD37615

Feature-Enhanced Fusion and Adversarial Domain Adaptation for Cross-Condition Few-Shot Fault Diagnosis of Bearings

Yuxin Li

Joint Southampton-Harbin University of Engineering Institute of Marine Engineering, Harbin Engineering University, Harbin, China
18814815205@163.com

Abstract. Rolling bearings are core components of rotating machinery, and their operational status is critical to the safe and reliable service of major equipment. However, the variable rotational speeds and loads in industrial environments cause significant data distribution drift; coupled with an extreme scarcity of labelled fault data under new operating conditions, this leads to a severe decline in the generalization performance of traditional deep learning models based on the assumption of independent and identically distributed (i.i.d.) data. To address the dual challenges, this paper proposes a fault diagnosis method based on Feature Enhancement Fusion and Adversarial Domain Adaptation (FEF-ADA). Firstly, a time-frequency spectrum is generated via the Continuous Wavelet Transform, and high-dimensional, discriminative fault features are deeply extracted whilst suppressing industrial background noise through a combination of hollow convolution, spectral convolution and a channel-spatial dual attention mechanism; secondly, a domain adversarial network with a gradient reversal layer is introduced to align the feature manifolds of the source and target domains via a maximum-minimum game mechanism, thereby extracting operating-condition-invariant representations; finally, prototype learning is introduced in the metric space to establish accurate classification decisions relying solely on a very small number of target-domain support samples. Experiments demonstrate that, under conditions of only 15 samples per class (15-shot), FEF-ADA achieved an average recognition rate of 89.70%, with a standard deviation of just 0.70% across five independent experiments. Its performance significantly outperforms traditional CNNs (73.41%), ProtoNet (78.66%) and DANN (82.47%), demonstrating excellent resistance to operating condition drift and significant engineering practical value.

Keywords: Bearing fault diagnosis, adversarial domain adaptation, few-shot learning, cross-operating conditions, feature fusion

1. Introduction

As Rolling bearings are critical components in modern industrial machinery, operating under severe alternating loads where minor damage can rapidly escalate into catastrophic failure [1]. While deep

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD37615

learning methods have significantly advanced intelligent fault diagnosis, their performance relies heavily on the independent and identically distributed (i.i.d.) assumption [2]. In practical environments, continuous variations in speed and load induce pronounced distribution shifts (covariate shift), severely degrading the cross-condition diagnostic accuracy of standard models [3]. Meanwhile, acquiring sufficient labeled fault data under newly encountered operating conditions is expensive and challenging, creating a typical few-shot learning bottleneck [1]. Conventional unsupervised domain adaptation often suffers from negative transfer under heavy noise and sample scarcity, whereas standard metric learning (e.g., ProtoNet) struggles with biased prototypes when domain discrepancies are unaligned [4].

To tackle these challenges, this paper proposes the Feature Enhancement Fusion combined with Adversarial Domain Adaptation (FEF-ADA) framework. The key contributions are summarized as follows: (1) Multi-Scale Feature Enhancement Backbone: Combines 2D continuous wavelet transform (CWT), hollow convolution, spectral convolution, and a channel-spatial dual-attention mechanism to suppress industrial noise while capturing discriminative time-frequency representations [5]. (2) Joint Adversarial Domain Adaptation and Prototypical Metric: Introduces a gradient reversal layer (GRL) to align domain manifolds, coupled with prototype metric learning to achieve accurate, condition-invariant classification using only a few target-domain samples [6, 7]. (3) Validation and Dynamic Attribution: Evaluated on the CWRU benchmark across 10 health states, demonstrating superior transfer accuracy and revealing the underlying confusion patterns among similar faults with different damage severities.

2. Theoretical foundations and related work

2.1. Current status of bearing fault diagnosis technology

Bearing fault diagnosis has evolved from traditional signal processing to data-driven intelligent approaches. Early methods relied heavily on time-frequency analysis (e.g., FFT, EMD) and dynamic characteristic frequencies, which demand substantial expert prior knowledge and lack robustness under severe noise and variable operating conditions [1]. While traditional machine learning introduced statistical classification, its performance ceiling remains constrained by manual feature engineering [2]. Recently, deep learning models (e.g., CNNs, ResNets) with attention-guided multi-scale feature fusion have achieved superior end-to-end fault representation [2]. However, most existing models still depend heavily on the independent and identically distributed (i.i.d.) assumption, limiting their practical transferability across complex operating conditions [8].

2.2. The challenge of data distribution drift in cross-operating-condition diagnostics

In actual industrial operating conditions, changes in operating conditions (such as variations in rotational speed, torque, radial load and lubricant temperature) inevitably lead to non-linear modulation of the bearing's natural vibration frequency response, impact pulse period and signal-to-noise ratio. Mathematically, let the source domain sample space and marginal distribution be $D_s=\{(x_i^s,y_i^s)\}_{i=1}^{N_s},x^s\sim P_s(X)$, and the target domain be $D_t=\left\{x_j^t\right\}_{j=1}^{N_t},x^t\sim P_t(X)$. Due to changes in operating parameters: $P_s(X)\neq P_t(X)$ or $P_s(Y|X)\neq P_t(Y|X)$. That is, there is significant covariate shift. If a model trained in the source domain is directly deployed in the target domain, the features it extracts are often dominated by condition-specific noise, resulting in a severe mismatch of the decision boundary and a significant decline in recognition accuracy [9].

DOI: 10.54254/2755-2721/2026.LD37615

2.3. Theoretical foundations of adversarial domain adaptation and few-sample learning

To eliminate distribution biases between domains, domain adaptation techniques based on transfer learning have been introduced into fault diagnosis. Among these, adversarial domain adaptation draws upon the game-theoretic principles of Generative Adversarial Networks (GANs). Its core lies in constructing a feature generator (extractor) G_f and a domain discriminator G_d :

• The domain discriminator G_d strives to distinguish whether a feature vector originates from the source domain or the target domain;

• The feature extractor G_d, on the other hand, uses a gradient reversal layer (GRL) to invert the sign of the domain discriminator's gradient during backpropagation ($-\lambda$), thereby encouraging the features it generates to be as ambiguous as possible G_d. When the game ultimately converges to a Nash equilibrium, the extracted feature space constitutes a domain-invariant manifold [3, 7].

Few-shot learning aims to achieve accurate induction using only a very small number of labelled samples (K-shot, e.g. $K \in$ [5, 15]) [10]. Prototypical networks, based on metric learning, map samples to an embedding space and define the mean of the support set for each class as the prototype vector for that class [4]:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\theta(x_i) \tag{1}$$

Classification is achieved by constructing a Softmax distribution based on the distance (e.g., Euclidean distance) between the query sample under test and the prototypes of each class. The organic integration of the macro-level distribution alignment capability of Adversarial Domain Adaptation with the micro-level metric-based classification capability of Prototypical Networks constitutes an optimal technical approach for overcoming the "cross-scenario + sample scarcity" bottleneck [6].

3. Construction of the Feature Enhancement Fusion with Adversarial Domain Adaptation (FEF-ADA) model

Addressing the engineering challenges in small-sample fault diagnosis of rolling bearings across operating conditions—namely, the inability to characterize weak fault features under strong noise, data distribution shifts caused by operating condition drift, and reduced model generalization performance due to a scarcity of labelled samples in the target domain—this chapter proposes the Feature-Enhanced Fusion and Adversarial Domain Adaptation (FEF-ADA) diagnostic model (Figure 1). Through a progressive architecture comprising multi-scale dual-domain feature extraction, attention-guided feature enhancement and fusion, adversarial domain adaptation for distribution alignment, and prototype-based small-sample classification, the model achieves high-precision bearing fault diagnosis in scenarios characterized by sample scarcity and operating condition drift.

DOI: 10.54254/2755-2721/2026.LD37615

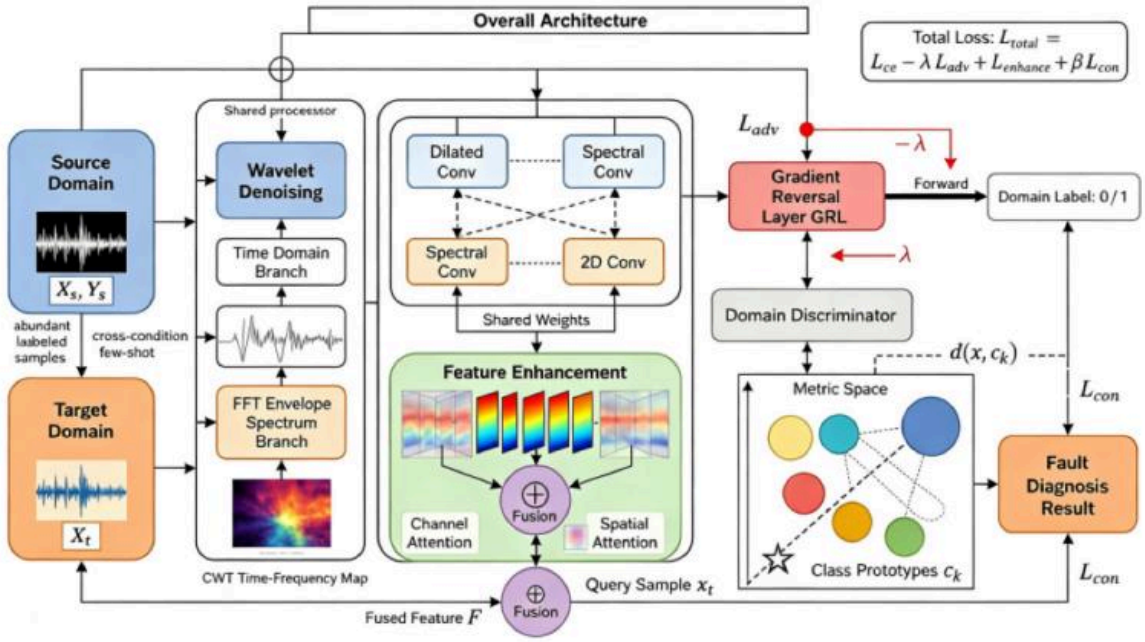


Figure 1. Overall architecture of the Feature-Enhanced Fusion and Adversarial Domain Adaptation (FEF-ADA) model

3.1. Overall architectural design of the FEF-ADA model

The overall architecture of FEF-ADA comprises four functional modules, as illustrated in Figure 1: (1) Hared Feature Preprocessing: Extracts raw multi-domain vibration features via parallel time-domain and continuous wavelet transform (CWT) time-frequency branches. (2) Feature Enhancement and Multi-Scale Fusion: Employs dilated and spectral convolutions with shared weights, coupled with channel-spatial dual attention (CBAM) to suppress noise and fuse discriminative representations [5]. (3) Adversarial Domain Adaptation: Aligns feature distributions across domains through a minimax game between the domain discriminator and the feature extractor via a gradient reversal layer (GRL) [7]. (4) Prototypical Metric Classification: Computes Euclidean distances to class prototypes in the metric space, enabling accurate few-shot decision-making on target-domain queries [4].

3.2. Loss function and collaborative optimization objectives

To achieve the co-optimization of feature discriminability, domain invariance and prototype compactness, the total training loss function of FEF-ADA is defined as follows:

$$L_{total}=L_{ce}-\lambda L_{adv}+L_{enhance}+\beta L_{con} \tag{2}$$

where L_{ce} represents the standard classification cross-entropy loss for labelled samples in the source domain, ensuring that the feature extractor possesses sufficient ability to distinguish between fault categories:

$$L_{ce} = -\frac{1}{N_s}\sum_{i=1}^{N_s}\sum_{k=1}^{C} I(y_i^s = k)\log P(y = k|x_i^s) \tag{3}$$

DOI: 10.54254/2755-2721/2026.LD37615

L_{adv}(DomainAdversarialLoss) represents the binary domain adversarial loss, whilst λ is the balancing hyperparameter:

$$L_{adv}=-\frac{1}{N}\sum_i [d_i \log G_d(G_f(x_i))+(1-d_i)\log(1-G_d(G_f(x_i)))] \quad (4)$$

During forward propagation, features are output directly to the discriminator; during backpropagation, GRL transforms the gradients entering the feature extractor into $-\lambda\frac{\partial L_{adv}}{\partial G_f}$, forcing the feature extractor to learn domain-invariant features [7].

$L_{enhance}$ represents the feature enhancement loss, L_{con} represents the prototype consistency loss; λ and β are the balancing weights for the domain adversarial term and the prototype constraint term, respectively. Through the collaborative optimization of multiple losses, the model simultaneously possesses the ability to distinguish fault features, adapt to distributions across operating conditions, and generalize with few samples.

3.3. Feature enhancement and multi-scale fusion module

To address the issues of incomplete feature representation in a single signal domain and the tendency for weak fault features to be drowned out by background noise under high-noise conditions, this module constructs a multi-scale feature extraction architecture based on time-frequency spectra obtained via the continuous wavelet transform. It enhances fault features and performs deep feature fusion through multi-layer convolutions, thereby improving the model's ability to distinguish fault features [2].

The training process employs a phased alternating optimization strategy: first, the feature extractor and classifier are pre-trained on the source-domain dataset; subsequently, target-domain samples are introduced to conduct adversarial domain adaptive training, thereby adapting to the data distribution of the target domain whilst retaining fault classification capability; ultimately, the FEF-ADA model, capable of small-sample diagnosis across operating conditions, is obtained [8].

4. Experimental validation and results analysis

4.1. Introduction to the experimental dataset and data pre-processing

The benchmark Case Western Reserve University (CWRU) bearing dataset (SKF6205 deep-groove ball bearing) is utilized to validate the proposed model. The dataset encompasses four distinct motor load and rotational speed conditions (0–3 hp at 1730–1797 rpm), comprising one healthy condition and nine distinct fault states across three locations (inner ring, outer ring, and rolling elements) and three severity levels (7, 14, and 21 mils), thereby formulating a 10-class diagnostic task. Raw single-channel vibration signals are segmented using a sliding window and converted into two-dimensional time-frequency energy spectra via the Continuous Wavelet Transform (CWT) with complex Morlet wavelets to serve as standardized inputs.

4.2. Task definition and evaluation metrics for cross-operating-condition small-sample tasks

Task Setup: The source domain is configured with a sufficient number of labelled samples (known operating conditions), whilst the target domain simulates new operating conditions encountered in industrial settings. Within the target domain, a strict " K-shot$(K=15)$ " is implemented: for each class, only 15 labelled samples are randomly selected to form the support set for parameter

DOI: 10.54254/2755-2721/2026.LD37615

optimization, whilst all remaining target domain samples are allocated to the query set (or test set) for performance evaluation [11].

Evaluation criterion: The average classification accuracy (Acc) is adopted as a comprehensive performance metric:

$$\mathrm{Acc}=\frac{\mathrm{N}_{\mathrm{correct}}}{\mathrm{N}_{\mathrm{total}}}\times 100\% \tag{5}$$

Eliminating Randomness: To thoroughly eliminate random errors arising from the random initialization of network weights and the random sampling of small support sets, all independent experiments were repeated five times, with the algorithm's accuracy and stability assessed using "mean ± standard deviation" (Mean±Std) [12].

4.3. Comparative experiments and model stability

To evaluate cross-condition few-shot diagnostic performance, FEF-ADA was benchmarked against standard CNN, ProtoNet [4], and DANN [7] under $K=15$ on the CWRU dataset. All models shared identical CWT inputs and hyperparameter search ranges. To eliminate randomness, five independent runs with reinitialized parameters and resampled target support sets were conducted. As shown in Table 3, the standard CNN achieves only 73.41% accuracy due to severe condition-induced covariate shift [8]. ProtoNet improves accuracy to 78.66% via metric learning but struggles with unaligned domain distributions [4]. DANN achieves 82.47% by introducing adversarial alignment [7].

Benefiting from the synergistic combination of multi-scale feature enhancement, domain alignment, and prototypical constraints, FEF-ADA achieves the highest average accuracy of 89.70%, outperforming DANN by 7.23%. Furthermore, FEF-ADA exhibits exceptional stability with a minimal standard deviation of 0.0070 across the five runs (ranging between 88.65% and 90.59%), verifying its robustness against the "prototype collapse" often observed in few-shot metric methods [6].

Table 3. Comparative experimental results of different methods

Method	Mean Accuracy	Standard Deviation
CNN	0.7341	0.0215
ProtoNet	0.7866	0.0182
DANN	0.8247	0.0146
Method described in this paper	0.8970	0.0070

4.4. Key parameter ablation and sensitivity analysis

Ablation studies were conducted under a lightweight benchmark ($N_{max}=100$) to evaluate the impact of core hyperparameters: support set size ($K-shot$), adversarial loss weight (λ), and sliding window step size ($STEP$). As summarized in Table 4, several key insights emerge: (1) K-shot Sensitivity: Increasing$K$ from 5 to 15 monotonically improves accuracy from 79.60% to 85.72%, underscoring the critical role of prototype anchor samples in constraining manifold drift under extreme target data scarcity. (2) Adversarial Weight Trade-off: Optimal performance is maintained across $\lambda \in [0.06,0.12]$ (around 86%). However, increasing λ to 0.18 causes accuracy to

DOI: 10.54254/2755-2721/2026.LD37615

drop significantly to 83.14%. This reflects a trade-off: insufficient adversarial weight fails to bridge domain gaps, whereas excessive alignment forces the extractor to smooth out essential class-discriminative vibration features in pursuit of domain confusion [3]. (3) Window Step Size: Setting *STEP=128* achieves the optimal accuracy (85.72%). Smaller steps (*STEP=64*) introduce sample redundancy, while larger steps(*STEP=192*) degrade performance (85.45%) due to reduced training diversity.

Table 4. Results of key hyperparameter ablation and sensitivity analysis (with restricted sample size, Nmax=100)

Experimental Group	K-SHOT	ADV-WEIGHT	STEP	Mean ± Standard Deviation
K-shot=5	5	0.12	128	0.7960 ± 0.0070
K-shot=10	10	0.12	128	0.8323 ± 0.0173
K-shot=15	15	0.12	128	0.8572 ± 0.0130
Adv = 0.06	15	0.06	128	0.8615 ± 0.0121
Adv=0.12	15	0.12	128	0.8572 ± 0.0130
Adv=0.18	15	0.18	128	0.831 ± 0.0109
Step=64	15	0.12	64	0.8462 ± 0.0100
Step=128	15	0.12	128	0.8572 ± 0.0130
Step=192	15	0.12	192	0.8545 ± 0.0102

4.5. Fine-grained confusion matrix analysis

To evaluate the discriminative capability across sub-categories, Figure 3 presents the normalized confusion matrix on the target query set. The model achieves near-perfect classification for distinctive health states, such as normal conditions (Class 0: 1.00) and specific fault modes (Classes 4 and 7 $\geq$ 0.98).

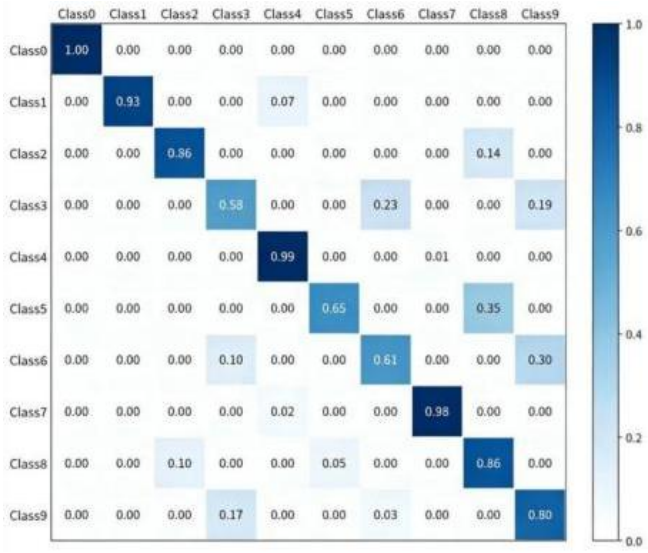


Figure 3. Confusion matrix

DOI: 10.54254/2755-2721/2026.LD37615

However, moderate misclassification is observed among certain subclasses, particularly between Class 3 (0.58) and Classes 6/9, as well as between Class 5 (0.65) and Class 8. These errors primarily arise from the subtle distinction between "same fault type with varying damage severities" (e.g., 7 mils vs. 14/21 mils). Under severe operating condition drift and limited target support samples ($K=15$), the non-linear modulation of impact frequencies and load variations makes fine-grained feature boundaries inherently ambiguous, leading to local prototype overlaps.

5. Conclusion

This paper addresses the engineering bottlenecks caused by data distribution drift resulting from variable load and speed in the industrial operation of mechanical equipment, as well as the extreme scarcity of fault-labelled data under new operating conditions, by proposing the Feature Enhancement Fusion Adversarial Domain Adaptation (FEF-ADA) bearing fault diagnosis architecture. This study introduces innovations in both theory and architecture, integrating the high-resolution representational capabilities of CWT continuous time-frequency analysis whilst effectively suppressing environmental interference through multi-scale dual-domain convolution and cross-dimensional attention mechanisms [5]; utilizing adversarial domain adaptation and gradient inversion layers to eliminate distribution shifts across operating conditions; and combining prototype metric learning to achieve high-performance closed-loop diagnosis requiring only 15 target samples.

This is strongly supported by experimental metrics. Cross-operating-condition 10-classification experiments on the CWRU benchmark dataset confirmed that FEF-ADA achieved an outstanding average diagnostic accuracy of 89.70%, with a standard deviation as low as 0.0070. Compared with traditional CNNs, ProtoNet and DANN, the recognition accuracy achieved significant increases of 16.29%, 11.04% and 7.23%, respectively. Ablation experiments and fine-grained confusion matrix analysis clarified the hyperparameter constraint boundaries and the confounding features of dynamic characteristics under "same fault type but different damage sizes". This framework effectively overcomes the reliance on vast amounts of labelled data for target operating conditions, demonstrating excellent industrial transferability and robustness. Future work will focus on introducing contrastive disentanglement to further address the challenges of distinguishing ultra-subtle ball defects and fine-grained differentiation between different damage sizes.

References

[1] Lei, Y., et al. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, 106587.
[2] Zhu, Y. (2026). A study on rolling bearing fault diagnosis based on feature fusion and deep learning [D]. Wuhan: Hubei University of Technology.
[3] Liu, M., et al. (2020). Triplet loss-guided adversarial domain adaptation for bearing fault diagnosis. *Sensors, 20*(1), 320.
[4] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems (NeurIPS), 30*.
[5] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 3-19.
[6] Yang, F., et al. (2020). Few-shot rolling bearing fault diagnosis with metric-based meta-learning. *Sensors, 20*(22), 6437.
[7] Ganin, Y., et al. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research, 17*(1), 2096–2030.
[8] Huang, X. (2025). A study on intelligent fault diagnosis methods for rolling bearings driven by multi-operating-condition few-shot data [D]. Chongqing:Chongqing Jiaotong University.

DOI: 10.54254/2755-2721/2026.LD37615

[9] Tang, X., et al. (2024). Domain adaptation for bearing fault diagnosis based on SimAM and an adaptive weighting strategy. *Sensors, 24*(13), 4251.
[10] Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *International Conference on Machine Learning (ICML)*, PMLR, 1126–1135.
[11] Wu, W., et al. (2025). Source-free domain adaptation based on label reliability for cross-domain bearing fault diagnosis. arXiv preprint arXiv:2503.08749.
[12] Zhang, A., et al. (2021). Few-shot bearing fault diagnosis based on model-agnostic meta-learning. *IEEE Transactions on Industrial Electronics, 68*(7), 6227–6236.