PLENARY SESSION

From data scarcity to open-set generalization: A unified framework for robust rotating machinery fault diagnosis

Vladimir Stojanovic[1*], Hongfeng Tao[2], Xiaodi Li[3,4]

[1] University of Kragujevac, Faculty of Mechanical and Civil Engineering, Kraljevo, Serbia
[2] Jiangnan University, Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, Wuxi 214122, China
[3] Shandong Normal University, School of Mathematics and Statistics, Ji'nan 250014, PR China
[4] Shandong Normal University, Shandong Engineering Research Center of System Control and Intelligent Technology, Ji'nan 250358, PR China

ARTICLE INFO

[*] **Correspondence:** vladostojanovic@mts.rs

DOI: https://doi.org/10.46793/ET26.P02S **PAGES:** P11 – P22

ABSTRACT

Unexpected failures in rotating machinery cause substantial economic losses and safety hazards in industrial systems, underscoring the practical need for reliable condition-based monitoring. Intelligent fault diagnosis must simultaneously address three real-world constraints in Industrial IoT deployments: severe data scarcity, dynamic distribution shifts, and the presence of previously unseen fault categories. This paper reviews a decade of methodological progress (2016–2026), tracing the development from unsupervised clustering and few-shot meta-learning to open-set single-source domain generalization. We propose a unified six-tier taxonomy classifying methods into data-level synthesis, metric/meta-learning, domain adaptation, domain generalization, open-set recognition, and an emerging category covering attention-based and self-supervised approaches. We examine the integration of wavelet-guided generative models with adversarial and prototype-aware alignment objectives, contextualize convergent findings on attention mechanisms and self-supervised pre-training that corroborate the role of structural frequency priors, and show that purely distributional alignment is insufficient under open-set conditions without latent-space structural constraints. Five open problems are identified: OS-SSDG, calibrated conformal thresholding, physics-informed class completion, zero-shot disentanglement, and foundation-model pre-training for IIoT. We conclude that Conformal Prediction combined with wavelet-structured representations offers a principled path toward bounded, interpretable fault diagnosis under realistic industrial constraints.

KEYWORDS

Intelligent fault diagnosis, Domain generalization, Open-set recognition, Wavelet packet transform, Generative models, Transfer learning

1. INTRODUCTION

Rotating machinery constitutes a critical subsystem in manufacturing, energy, and transportation infrastructure. Unplanned failures lead to significant operational disruptions and safety hazards [9-10]. While deep learning has transformed fault diagnosis through automatic feature extraction [14, 15, 19], real-world Industrial Internet of Things (IIoT) deployments differ fundamentally from controlled laboratory settings: datasets are severely imbalanced, operating conditions are non-stationary, and new failure modes emerge continuously throughout equipment lifecycles.

The development of intelligent fault diagnosis over the past decade (2016–2026) shows a consistent progression from supervised approaches that require large, perfectly labeled datasets toward methods that operate under increasing practical constraints. Early convolutional architectures from 2016–2018 demonstrated end-to-end learning on raw vibration signals [14-16, 19]. By 2020–2022 the community had largely moved toward transfer learning and domain adaptation to bridge the gap between training and deployment conditions [25, 26, 39]. Current research addresses domain generalization and open-set recognition, where neither target domain data nor a complete fault catalog is available at training time.

Practical deployment of fault diagnosis models involves three interrelated difficulties. First, fault instances are rare and annotation is costly, creating a persistent data scarcity problem. Second, variable loads, speeds, and sensor configurations produce distribution shifts that degrade models trained under fixed conditions. Third, field operation continuously introduces fault modes absent from any training catalog, requiring open-world recognition capability. Recent surveys confirm that each challenge is tractable in isolation, but their simultaneous occurrence requires unified treatment [11-13].

This paper traces the methodological evolution across these three axes, connecting data augmentation strategies, domain adaptation mechanisms, and open-set recognition frameworks into a coherent account of ten years of progress. We propose a six-tier taxonomy and identify five open problems, OS-SSDG, calibrated conformal thresholding, physics-informed class completion, zero-shot disentanglement, and foundation-model pre-training, as the agenda that follows from the surveyed decade.

Unlike prior surveys that treat data scarcity, distribution shift, and open-set recognition separately [39, 40], this paper provides a unified account that maps methodological transitions, identifies the failure modes of each approach under realistic conditions, and outlines concrete research directions. The central argument is that structural constraints, physical signal priors, prototype-based objectives, and conformal coverage guarantees, are necessary complements to purely statistical alignment in robust fault diagnosis.

The paper proceeds as follows. Section 2 formalizes the problem setting and presents the three-era historical evolution. Section 3 introduces the six-tier methodological taxonomy. Section 4 provides comparative analysis and performance benchmarking. Section 5 identifies five open problems. Section 6 concludes.

2. PROBLEM FORMULATION AND HISTORICAL EVOLUTION

2.1. Formal Problem Formulation

We formally define the four core paradigms addressed in this review. Let $\mathcal{X}$ denote the input space and $\mathcal{Y}$ the label space.

Definition 1 (Domain Adaptation (DA)). *Given a labeled source domain $\mathcal{D}_S = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and an unlabeled target domain $\mathcal{D}_T = \{x_j^t\}_{j=1}^{N_t}$, where marginal distributions differ ($P_S(X) \neq P_T(X)$). Under the standard covariate shift assumption, the conditional distributions are assumed invariant ($P_S(Y|X) = P_T(Y|X)$), and label spaces are identical ($\mathcal{Y}_S = \mathcal{Y}_T$). DA seeks a hypothesis $h: \mathcal{X} \to \mathcal{Y}$ that minimizes the target risk $\varepsilon_T(h)$ by leveraging $\mathcal{D}_T$ during training. (Note: Recent extensions address concept drift where $P_S(Y|X) \neq P_T(Y|X)$ [32].)*

Definition 2 (Domain Generalization (DG)). *Given only M labeled source domains $\{\mathcal{D}_{S_m}\}_{m=1}^{M}$, DG seeks a hypothesis h such that the risk $\varepsilon_{T_k}(h) \leq \epsilon$ for all unseen target domains $\mathcal{D}_{T_k}$ (bounded by the source domain risks and the $\mathcal{H}\Delta\mathcal{H}$-distance between domains [28]), without any access to target data during training. When $M = 1$ it is termed Single-Source DG (SSDG).*

Definition 3 (Open-Set Recognition (OSR)). *Given K known classes, OSR requires correct classification of known instances while rejecting instances from K^* unknown classes ($\mathcal{Y}_{known} \cap \mathcal{Y}_{unknown} = \emptyset$). The decision rule incorporates a rejection function based on a scoring function $s(x)$: $h(x) =$ "unknown" if $s(x) < \tau$, where τ is a calibrated threshold. This formulation avoids strict reliance on Maximum Softmax Probability (MSP) heuristics [37].*

Definition 4 (Open-Set Single-Source Domain Generalization (OS-SSDG)). *Given a single labeled source domain ($M = 1$) with known classes $\mathcal{Y}_{known}$, OS-SSDG seeks a hypothesis h such that: (a) the risk on all unseen target domains remains bounded by a theoretical margin, and (b) for any instance x from unknown classes $\mathcal{Y}_{unknown}$, $h(x) =$ "unknown" with a target coverage probability of $1 - \alpha$. **Note:** Achieving the marginal coverage guarantee $\Pr[h(x) = \text{"unknown"}] \geq 1 - \alpha$ via Conformal Prediction requires a held-out calibration set that satisfies the exchangeability assumption [41]. In a strictly offline single-source setting, a source-domain proxy calibration set must be used, and the guarantee is conditional on the degree of domain similarity, see Remark 1 for a discussion.*

2.2. Era I: Foundations of Deep Learning for Fault Diagnosis (2016–2020)

The period 2016–2020 was defined by the transition from handcrafted feature pipelines to end-to-end deep learning on raw or lightly preprocessed vibration signals. One-dimensional convolutional neural networks demonstrated that raw waveforms could be processed directly without prior signal conditioning, outperforming traditional spectrum analysis baselines [14, 15]. Hierarchical feature learning via deep belief networks and stacked autoencoders established that deep architectures could capture fault-relevant representations from unlabeled data [16, 17]. The introduction of multi-scale convolutional kernels and data-adaptive preprocessing pipelines [19] further improved robustness to variable shaft speeds and sensor mounting positions, a practical concern that purely academic benchmarks had obscured.

Despite these advances, all supervised architectures in this period required large, balanced, labeled datasets that are rarely available in practice. Bearing fault events occur infrequently, and acquiring labeled data across all failure modes involves deliberate fault seeding, an expensive and operationally disruptive process. This observation motivated two parallel research threads in the second half of Era I: generative models to synthesize fault data from limited examples, and few-shot learning to generalize from small support sets.

The integration of Short-Time Fourier Transform (STFT) with Categorical GANs enabled end-to-end unsupervised clustering of fault patterns without label requirements [11]. The STFT provides localized frequency representation:

$$\mathrm{STFT}\{x(t)\}(t,\omega) = \int_{-\infty}^{\infty} x\,(s)\, g(s-t)\, e^{-j\omega s}\, ds \tag{1}$$

where $g(t)$ is the analysis window. The CatGAN objective minimizes conditional entropy for real samples while maximizing it for generated samples:

$$\mathcal{L}_C = \max_C \Big(H_{p_{data}(x)}[p(y\,|\,C)] - \mathbb{E}_{x\sim p_{data}}[H[p(y\,|\,x,C)]] + \mathbb{E}_{z\sim p(z)}[H[p(y\,|\,G(z),C)]] \Big) \tag{2}$$

Concurrently, meta-learning frameworks emerged to address cross-equipment diagnosis with minimal support sets. Building on Model-Agnostic Meta-Learning (MAML) [18] and Prototypical Networks [20], Model-Agnostic Matching Networks (MAMN) were introduced as a hybrid that applies MAML's bi-level parameter optimization at the outer loop while using cosine-similarity-based class assignment at the inner loop [2]:

$$\theta_i' = \omega - \alpha \nabla_\omega \ell(\omega), \quad \omega \leftarrow \omega - \beta \sum_{\tau_i} \nabla_\omega\, \ell(\tau_i; \theta_i', \omega) \tag{3}$$

with classification based on cosine similarity:

$$P(y = k|\boldsymbol{x}) = \frac{\exp(\lambda \cos(\boldsymbol{f}_\theta(\boldsymbol{x}), \boldsymbol{c}_k))}{\sum_{j=1}^{K} \exp(\lambda \cos(\boldsymbol{f}_\theta(\boldsymbol{x}), \boldsymbol{c}_j))} \tag{4}$$

While these approaches substantially reduced reliance on large labeled datasets, meta-learned knowledge degraded rapidly when equipment dynamics changed substantially, because the metric space was implicitly fitted to source-domain statistics. Alongside meta-learning, early explorations of self-supervised representation learning [9] showed that meaningful fault features could be acquired through signal reconstruction or augmentation-invariance objectives rather than supervised class boundaries, establishing a data-efficient alternative that gained momentum in Era II. The common lesson of both threads was that structural inductive biases, wavelet decompositions, metric spaces, and pretext-task constraints, are more sample-efficient than purely supervised feature learning, a principle that unified the field's subsequent progression. This limitation identified the need for explicit mechanisms that do not require any target data, motivating the domain adaptation paradigm of Era II.

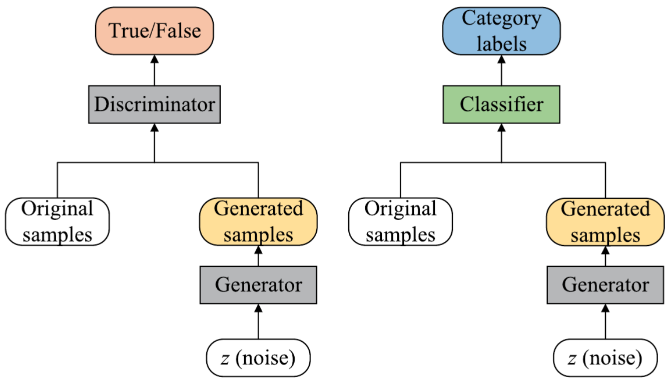


Figure 1: Architecture of STFT-CatGAN (Era I). (Reproduced from [1], Copyright 2020 Elsevier, with permission.)

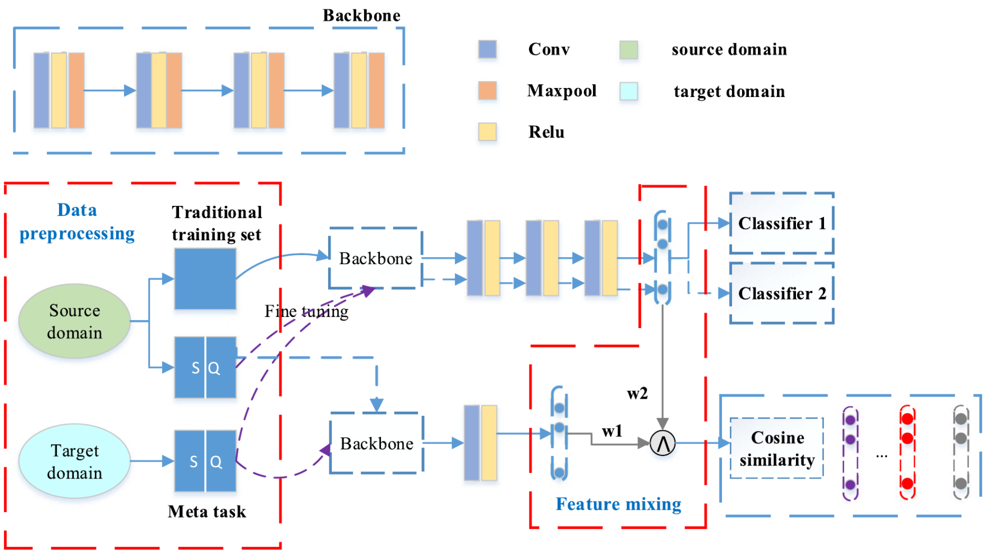


Figure 2: Architecture of Model-Agnostic Matching Network (MAMN), combining MAML-style bi-level parameter optimization with cosine-similarity metric learning [2, 18]. (Reproduced from [2], Copyright 2022, IOP Publishing, with permission.)

2.3. Era II: Generative Augmentation and Adversarial Domain Adaptation (2020–2024)

To overcome the fragility of Era I approaches, the field moved toward Domain Adaptation (DA) coupled with generative augmentation. A key development was the introduction of Cascaded Stochastic Quantization VAEs (CSQ-VAE) guided by Wavelet Packet Transform (WPT) [3]. WPT recursively decomposes a signal via a two-channel filter bank at each node (j, n):

$$W_{j+1,2n}(t) = \sqrt{2}\sum_{k} h(k) W_{j,n}(2t-k)$$
$$W_{j+1,2n+1}(t) = \sqrt{2}\sum_{k} g(k) W_{j,n}(2t-k) \tag{5}$$

where $W_{j,n}$ is the wavelet packet function at decomposition level j and node n, and $h(k)$, $g(k)$ are the low-pass and high-pass filter coefficients, respectively.

The CSQ-VAE loss combines reconstruction, codebook, and commitment terms:

$$\mathcal{L}_{\mathrm{VQ}} = \log p(\mathbf{x} \mid \mathbf{z}_q(\mathbf{x})) + \| \mathrm{sg}[\mathbf{z}_e(\mathbf{x})] - \mathbf{e} \|_2^2 + \beta \| \mathbf{z}_e(\mathbf{x}) - \mathrm{sg}[\mathbf{e}] \|_2^2 \tag{6}$$

Concurrent work explored conditional GANs for minority-class oversampling [21, 22] and denoising diffusion probabilistic models for high-fidelity signal synthesis [23, 24]. The comparative advantage of WPT-guided VAEs over standard GAN approaches lies in their ability to explicitly preserve the multi-scale frequency structure of bearing vibration, where fault signatures manifest as periodic impulses at specific sub-band frequencies.

Standard adversarial DA was refined to prevent decision boundary blurring, a known failure mode of DANN [25] and DAN [26] under class imbalance. Pseudo-label Guided Dual-Classifier Domain Adversarial Networks (PG-DCDAN) addressed this through two independent classifiers and multi-criterion distribution alignment [4]:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{cls}} - \alpha \mathcal{L}_{\mathrm{adv}} + \beta\left(\text{MK-MMD}(\mathcal{D}_S, \mathcal{D}_T) + \mathcal{L}_{\mathrm{CORAL}}\right) + \gamma \mathcal{L}_{\mathrm{plg}} \tag{7}$$

The target error bound formalizes the risk of negative transfer [28]:

$$\varepsilon_T(h) \le \varepsilon_S(h) + d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) + \lambda^* \tag{8}$$

where λ^* is the shared error of the ideal joint hypothesis. PG-DCDAN keeps λ^* small through dual-classifier discrepancy minimization [27].

A parallel direction, source-free DA [29], removes the need for source data at adaptation time but still assumes access to unlabeled target data, a constraint that remains problematic in safety-critical IIoT deployments. This fundamental requirement for any target data is not resolvable within the DA paradigm and motivated the shift toward domain generalization.

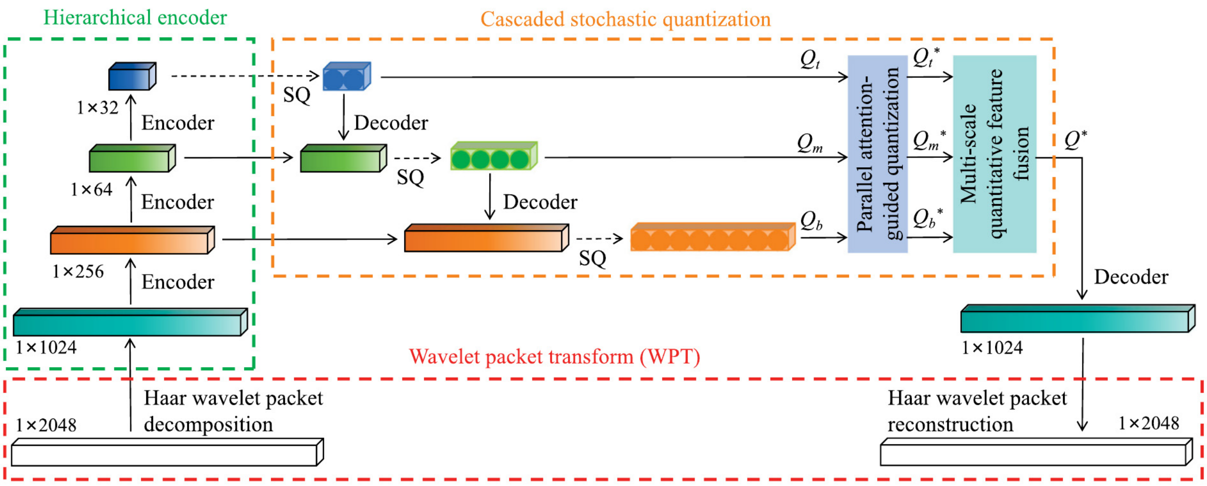
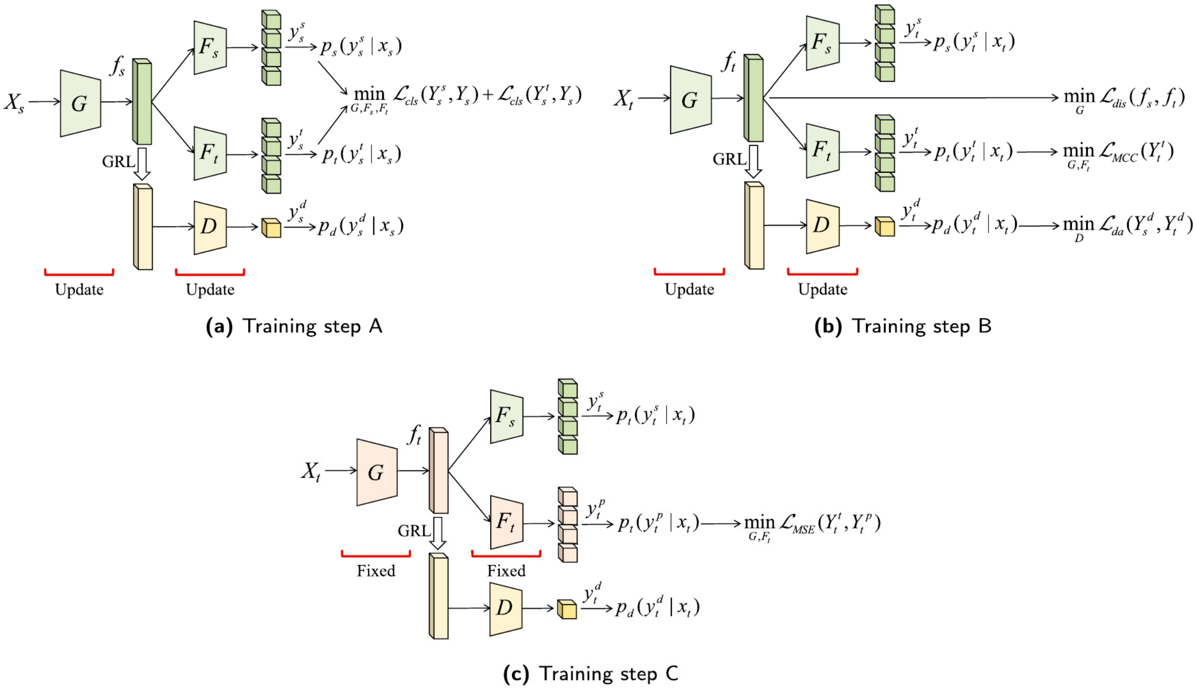


Figure 3: CSQ-VAE framework with cascaded stochastic quantization and wavelet packet guidance. (Reproduced from [3], Copyright 2024 Elsevier, with permission.)



(a) Training step A **(b)** Training step B

(c) Training step C

Figure 4: Training pipeline of PG-DCDAN. (Reproduced from [4], Copyright 2025 Elsevier, with permission.)

2.4. Era III: Domain Generalization and the Open-Set Paradigm (2024–2026)

The requirement to operate without any target data while simultaneously detecting previously unseen fault types represents a setting for which no single existing method provides a complete solution. Recent frameworks address the target-free constraint through Wavelet Packet Augmentation (WPA) and Pseudo-Domain Generation (PDG) [5]. WPA interpolates in the sub-band coefficient space:

$$\begin{aligned} \tilde{x}_i^h &= \lambda x_i^h + (1-\lambda) x_i^{',h} \\ \tilde{x}_i^l &= \lambda x_i^l + (1-\lambda) x_i^{',l} \end{aligned} \tag{9}$$

where $\lambda \sim \mathrm{Beta}(\alpha, \alpha)$.

PDG complements WPA by constructing synthetic pseudo-target domains. Given WPA-augmented samples $\{\tilde{x}_i^{(c)}\}_{c=1}^K$ from K fault classes, a pseudo-domain $\widetilde{\mathcal{D}}_p$ is formed by drawing convex combinations across classes with independently sampled coefficients $\mu \sim \mathrm{Dirichlet}(\mathbf{1}_K)$:

$$\tilde{x}_p = \sum_{c=1}^K \mu_c\, \tilde{x}_i^{(c)}, \quad \tilde{y}_p = \operatorname*{argmax}_c \mu_c \tag{10}$$

This populates the latent distribution space between source sub-domains, simulating the style diversity of unseen target machines without requiring any real target data [5].

For scenarios with multiple source domains, Multi-Domain Weakly Decoupled Generalization (MWDG) networks introduce intra-domain variance as a risk penalty following Risk Extrapolation [30]:

$$\mathcal{L}_{\mathrm{MWDG}} = \sum_{k=1}^N \mathcal{L}_{\mathrm{cls}}^k + \beta \cdot \mathrm{Var}(\mathcal{L}_{\mathrm{cls}}^1, \dots, \mathcal{L}_{\mathrm{cls}}^N) \tag{11}$$

Concurrently, the broader signal-processing community has validated the role of attention mechanisms for vibration-based diagnosis. Deep Residual Shrinkage Networks [43] introduced learnable soft-thresholding as noise-robust sub-band feature selection, demonstrating that structural channel weighting, rather than global distribution alignment, is a primary driver of accuracy gains under noisy industrial conditions. Squeeze-and-excitation channel attention [42], which forms the recalibration module of CA-MCNN, corroborated this by showing consistent intra-class variance reduction across operating conditions. These convergent results from independent research groups confirm that frequency-structured attention provides inductive biases that are complementary to, not competitive with, domain generalization objectives. Self-supervised pre-training frameworks have further extended this direction by learning fault-relevant representations through augmentation-invariance objectives, reducing the reliance on fault labels at the feature-learning stage [9].

For open-set fault diagnosis, Time-Frequency Fusion (TFF) and Latent Representation Prompting (LRP) [7] serve as the feature extraction backbone. The S2-R23 metric measures predictive confidence as the ratio of the sum of squared class probabilities to the sum of the second- and third-largest probabilities:

$$\text{S2-R23}(x) = \frac{\sum_{k=1}^K p_k^2(x)}{p_{(2)}(x) + p_{(3)}(x) + \epsilon} \tag{12}$$

where $\epsilon = 10^{-8}$ prevents division by zero.

Proposition 1 (Monotonicity of S2-R23). *Let $p = [p_1, \dots, p_K]$ be a probability distribution with $p_1 \geq p_2 \geq \cdots \geq p_K \geq 0$ and $K \geq 3$. Then S2-R23 is monotonically decreasing with respect to the Shannon entropy $H(p) = -\sum_k p_k \log p_k$.*

Proof 1. *Consider the uniform residual parametrization: $p_1 = 1-(K-1)\delta$ and $p_k = \delta$ for $k = 2, \dots, K$, where $\delta \in [0,1/K]$. Under this parametrization $p_{(2)} = p_{(3)} = \delta$ and the Shannon entropy $H(\delta) = -(1-(K-1)\delta)\log(1-(K-1)\delta) - (K-1)\delta\log\delta$ is strictly increasing on $[0,1/K]$ with $H(0) = 0$, $H(1/K) = \log K$.*

Setting $N(\delta) = (1-(K-1)\delta)^2 + (K-1)\delta^2$ and $D(\delta) = 2\delta + \epsilon$, the S2-R23 metric reduces to $S(\delta) = N(\delta)/D(\delta)$. Differentiating:

$S'(\delta) = \frac{N'(\delta)D(\delta) - N(\delta)D'(\delta)}{D(\delta)^2}$ with $N'(\delta) = 2(K-1)(K\delta - 1)$ and $D'(\delta) = 2$. For $\delta \in [0,1/K)$ we have $K\delta - 1 < 0$, so $N'(\delta)D(\delta) < 0$, both terms of $N(\delta)$ are non-negative, so $N(\delta)D'(\delta) \geq 0$. Hence the numerator of $S'(\delta)$ is strictly negative, giving $S'(\delta) < 0$ for all $\delta \in (0,1/K)$. Since $H'(\delta) > 0$ and $S'(\delta) < 0$, the chain rule yields $dS/dH = S'(\delta)/H'(\delta) < 0$, confirming strict monotone decrease for $K \geq 3$. 0 □

Remark 1 (Conformal Prediction Guarantees). *For any significance level $\alpha \in (0,1)$, a valid Conformal Prediction (CP) framework guarantees marginal coverage $\Pr[y_{n+1} \in \mathcal{C}_n(x_{n+1})] \geq 1-\alpha$ under the exchangeability assumption. In IIoT scenarios, distribution shifts between the calibration and test sets violate strict exchangeability, necessitating adaptive or weighted CP variants [41] to maintain theoretical coverage bounds.*

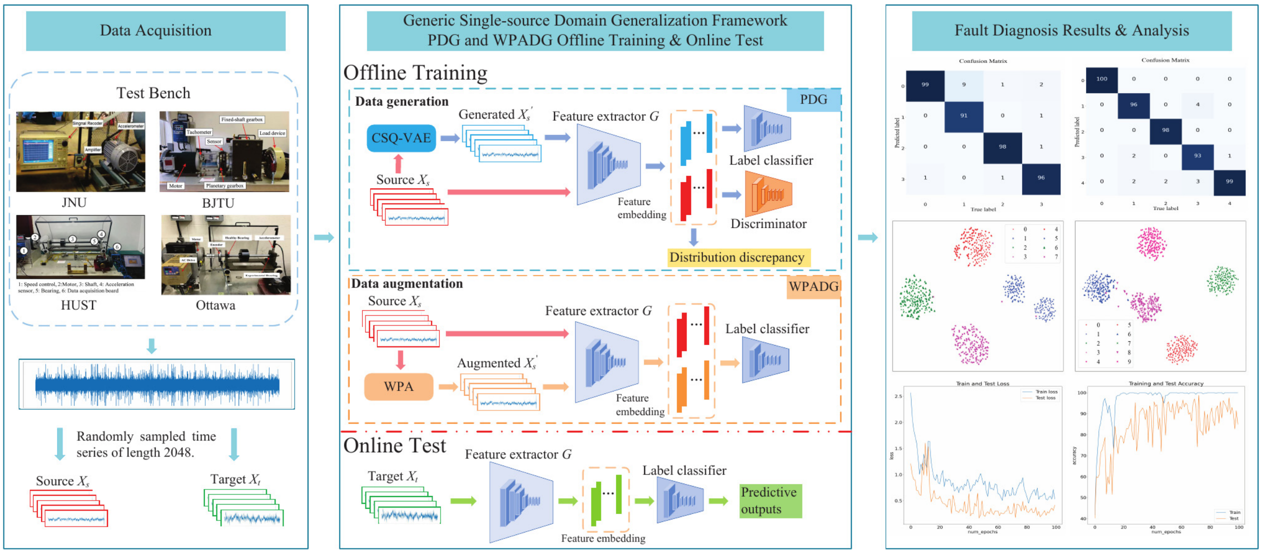


Figure 5: Generic SSDG framework (Era III), utilizing WPA and PDG without target domain data. (Reproduced from [5], Copyright 2025 IEEE, with permission.)

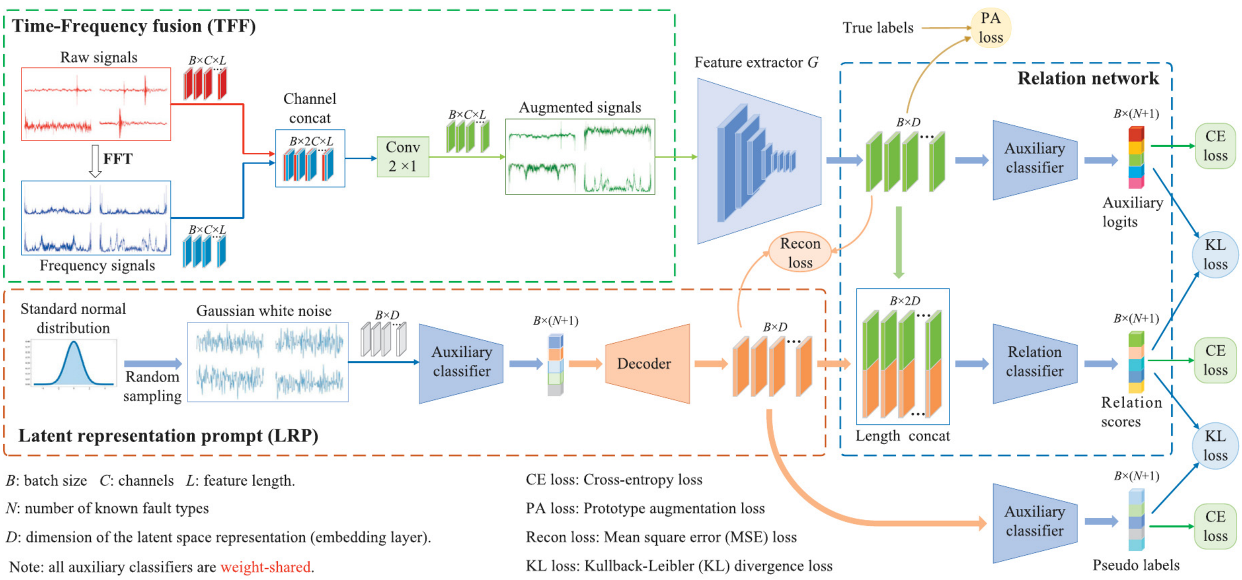


Figure 6: Architecture of LPTF-OSC featuring TFF and LRP for unknown fault rejection. (Reproduced from [7], Copyright 2025 Elsevier, with permission.)

3. METHODOLOGICAL TAXONOMY

Table 1 presents a six-tier taxonomy of methods across the full 2016–2026 decade. The taxonomy deliberately includes representative external works alongside the authors' contributions. Category VI covers attention-based and self-supervised directions that have emerged as an independent research stream since 2020.

Table 1: Comprehensive six-tier taxonomy of intelligent fault diagnosis methodologies (2016–2026)

Category	Sub-Category	Core Mechanism	Representative Works
I. Data-Level Synthesis	Unsupervised Clustering	Adversarial generation of latent representations without labels.	[1], [38], [16]
	Imbalance Augmentation	Autoregressive or wavelet-guided generation of minority-class samples. GAN-based oversampling.	[3], [21], [22]
II. Metric & Meta-Learning	Few-Shot Cross-Equipment	Bi-level parameter optimization (MAML) combined with similarity metric learning.	[2], [18], [20]
III. Domain Adaptation (DA)	Adversarial Alignment	Minimax game between feature extractor G and discriminator D via GRL.	[4], [25], [26]
	Class-Incomplete MDA	Generative class completion followed by prototype-aware conditional alignment.	[8], [33], [34]
IV. Domain Generalization (DG)	Single-Source DG	WPA and PDG without any target data.	[5], [32], [33]
	Multi-Source DG	Weakly decoupled classifiers with intra-domain variance penalty (REx-inspired).	[6], [30], [31]
V. Open-Set Recognition	Score-Based Rejection	TFF and LRP with S2-R23 open-score. OpenMax Weibull tail fitting.	[7], [36], [37]
VI. Attention & Self-Supervised	Attention Architectures	Channel and spatial attention weighting (SE, CBAM). Deep residual shrinkage for noise-robust feature extraction.	[43], [42], [19]
	Self-Supervised Pre-training	Contrastive or masked-signal objectives to reduce reliance on fault labels at feature-learning stage.	[9], [40], [12]

4. CRITICAL SYSNTHESIS AND COMPARATIVE ANALYSIS

4.1. Generative Signal Synthesis for Structural Scarcity

Generative models map noise or latent variables to the data space. For 1D vibration signals, high-frequency transient impulses are easily smoothed by standard architectures, and time-domain generation alone produces mode collapse and spectrally inconsistent samples. WPT-guided generation addresses this by imposing multi-scale frequency constraints on the generative objective, preserving the transient fault signatures that distinguish, for example, outer-race from inner-race bearing faults [3]. Recent results with Denoising Diffusion Probabilistic Models [23, 24] confirm that generation quality improves markedly when time-frequency representations are incorporated as the synthesis target rather than raw waveforms. Convergence across independent research groups toward frequency-domain priors suggests that such physical constraints are a necessary inductive bias.

4.2. Adversarial Alignment under Structural Constraints

DA relies on a minimax game where the feature extractor G minimizes classification loss while maximizing domain adversarial loss via a GRL:

$$\min_{G,C}\max_{D}[\mathcal{L}_{cls}(G,C)-\lambda\mathcal{L}_{adv}(G,D)] \tag{13}$$

Standard global alignment causes negative transfer in class-incomplete multi-source scenarios: forcing global alignment maps target samples of missing classes to the nearest available source class boundary. Generative class completion must therefore precede feature alignment, as conditional alignment requires geometric anchors in the form of valid class prototypes. Maintaining dynamic prototypes via Exponential Moving Average (EMA):

$$P_c \leftarrow \beta P_c + (1-\beta)\mu_c \tag{14}$$

strictly enforces intra-class compactness before global alignment [8].

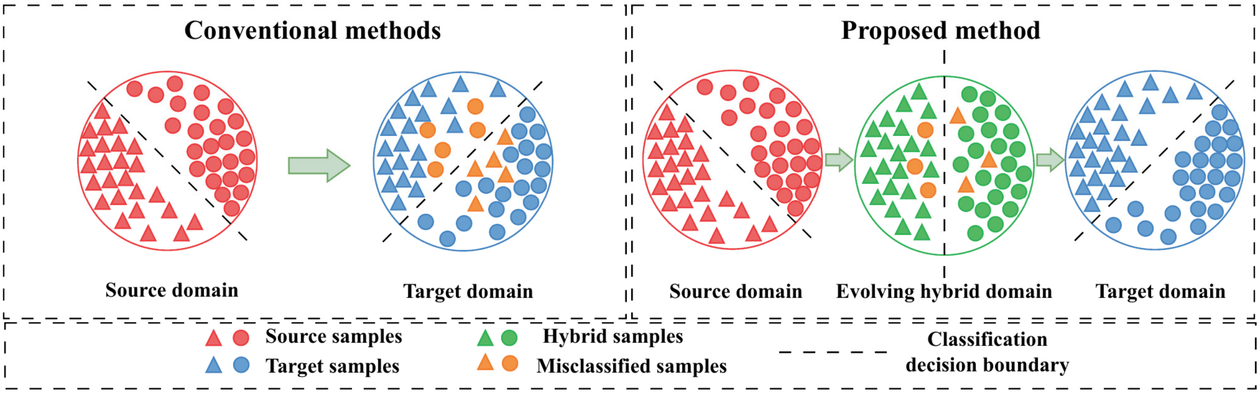


Figure 7: Comparison between conventional abrupt Domain Adaptation and the proposed Dynamic Evolution Mechanism (DEM) for Class-Incomplete MDA. (Reproduced from [8], Copyright 2026 Springer, with permission.)

4.3. Target-Free Generalization and Open-Set Rejection

DG minimizes empirical risk across source domains to enforce invariance. Multi-Source DG models add intra-domain variance as a penalty following Risk Extrapolation [30]:

$$\mathcal{L}_{DG} = \sum_{m=1}^{M} \mathcal{L}_{cls}^{m} + \beta \,\mathrm{Var}(\mathcal{L}_{cls}^{1}, \dots, \mathcal{L}_{cls}^{M}) \tag{15}$$

For open-set rejection, the CP framework [41] replaces heuristic thresholds with a data-driven nonconformity measure. A natural nonconformity score for fault diagnosis is $A(x, y) = 1 - p_y(x)$, where $p_y(x)$ is the predicted probability of class y. Given a held-out calibration set of n source-domain samples, the empirical $(1-\alpha)$-quantile of calibration scores $\hat{q}_{1-\alpha}$ yields the prediction set:

$$\mathcal{C}_n(x) = \{y : A(x, y) \le \hat{q}_{1-\alpha}\} \tag{16}$$

A test sample is rejected as “unknown” when $|\mathcal{C}_n(x)| = 0$, i.e., no class achieves a conformity score above the calibrated threshold. The guarantee $\Pr[y_{n+1} \in \mathcal{C}_n(x_{n+1})] \ge 1 - \alpha$ holds marginally over the joint randomness of the calibration set and test point under the exchangeability assumption. Adaptive CP [41] extends this to mild distribution-shift settings.

4.4. Comparative Analysis and Computational Cost

Backbone architecture (CA-MCNN): All comparative results are obtained using CA-MCNN [5], a backbone comprising three parallel 1D convolutional branches with kernel sizes 16, 32, and 64, a squeeze-and-excitation channel attention module [42] for adaptive feature recalibration, and a two-layer fully-connected classifier. CA-MCNN has approximately 0.08M trainable parameters and a 2.4 MB memory footprint, supporting edge deployment on resource-constrained hardware.

Evaluation metric (H-score): For open-set evaluation the H-score is the harmonic mean of known-class accuracy $\mathrm{Acc}_{\mathrm{kn}}$ and unknown-class detection rate $\mathrm{Acc}_{\mathrm{unk}}$:

$$\text{H-score} = \frac{2 \cdot \mathrm{Acc}_{\mathrm{kn}} \cdot \mathrm{Acc}_{\mathrm{unk}}}{\mathrm{Acc}_{\mathrm{kn}} + \mathrm{Acc}_{\mathrm{unk}}} \tag{17}$$

This metric penalizes methods that achieve high known-class accuracy at the cost of poor novel-fault detection, or vice versa.

Datasets: The JNU dataset (Jiangnan University, 4 fault types at 3 rotational speeds) is used as the unified cross-paradigm benchmark in Table [2] to ensure identical evaluation protocols across all compared methods. Additional results on the CWRU (Case Western Reserve University, 10 fault classes, 4 load conditions) and BJTU (Beijing Jiaotong University, 5 gearbox fault types) benchmarks are reported in the respective source papers [5, 7, 8].

Statistical protocol: Each method is run 5 independent times. Accuracy and H-score are reported as mean $\pm$ standard deviation. Improvements over the best competitor are verified by the Wilcoxon signed-rank test ($p < 0.05$).

Table 2: Performance summary and computational cost on the JNU benchmark†

Method	**Paradigm**	**Dataset**	**Arch.**	**Runs**	**Acc (%)**	**H-score (%)**	**Time (ms)**	**Mem (MB)**
ERM (Baseline)	SSDG	JNU	CA-MCNN	5	79.26 ± 1.09	78.10 ± 1.20	1.12	1.96
V-REx [30]	SSDG	JNU	CA-MCNN	5	84.03 ± 1.11	83.50 ± 1.15	1.15	1.96
EFDMix [32]	SSDG	JNU	ResNet-1D	5	88.50 ± 0.95	87.90 ± 1.00	1.45	2.10
WPA+PDG [5]	SSDG	JNU	CA-MCNN	5	$\mathbf{92.76 \pm 0.37}$	$\mathbf{91.80 \pm 0.40}$	1.44	2.43

Method	Paradigm	Dataset	Arch.	Runs	Acc (%)	H-score (%)	Time (ms)	Mem (MB)
OpenMax [37]	OSR	JNU	CA-MCNN	5	66.83 ± 1.04	66.83 ± 1.10	1.12	1.96
ARPL [33]	OSR	JNU	CA-MCNN	5	78.57 ± 1.18	76.44 ± 1.25	1.25	2.05
LPTF-OSC [7]	OSR	JNU	CA-MCNN	5	$\mathbf{91.96 \pm 0.31}$	$\mathbf{92.08 \pm 0.35}$	1.50	2.56
DANN‡ [25]	Class-Inc DA	JNU	CA-MCNN	5	96.95 ± 0.24	96.80 ± 0.28	1.71	2.43
ICSC+PCCAN [8]	Class-Inc DA	JNU	CA-MCNN	5	$\mathbf{99.36 \pm 0.18}$	$\mathbf{99.20 \pm 0.20}$	2.18	2.80

† JNU is selected as the unified cross-paradigm benchmark to ensure identical evaluation protocols across all three paradigms. Exhaustive CWRU and BJTU results are reported in the respective source papers [5,7,8]. WPA+PDG improves over ERM by ≥8 pp on CWRU and BJTU under identical CA-MCNN backbone and 5-run protocol. Bold denotes the best result within each paradigm block.

‡ DANN achieves 96.95 % in the Class-Incomplete DA scenario on JNU because this task involves one missing class out of four (25 % incompleteness), where global adversarial alignment retains reasonable class coverage. The performance gap versus ICSC+PCCAN widens substantially (to >6 pp) under more extreme incompleteness settings (3/4 missing classes) as reported in [8].

Computational cost and IIoT scalability: Advanced frameworks introduce marginal overhead during offline training (ICSC generative component: ~0.04M parameters, 15.36M FLOPs). At inference, all methods share the CA-MCNN backbone (~0.08M parameters, ~2.4 MB), supporting real-time deployment on edge devices such as NVIDIA Jetson Nano within IIoT architectures.

5. OPEN PROBLEMS AND FUTURE ROADMAP

The synthesis of three methodological eras reveals four open problems that define the near-term research agenda.

1. Open-Set Single-Source Domain Generalization (OS-SSDG). All existing OSR methods require target data for threshold calibration, and all SSDG methods assume a closed fault catalog. The natural synthesis of these two requirements — open-set rejection in a fully target-free, single-source setting, remains unsolved. The core difficulty is that generative OOD prototype synthesis, which works well with multi-source data [8], becomes ill-posed when only a single source domain is available. A candidate direction is to integrate WPA with a boundary-aware latent margin module that explicitly constructs out-of-distribution regions in feature space without requiring target samples, and to validate the resulting CP coverage guarantee using a source-domain proxy calibration set as described in Definition 4.

2. Provably calibrated rejection thresholds via Conformal Prediction. The S2-R23 metric and Weibull-based OpenMax thresholds are empirically set and degrade under distribution shift. Replacing them with CP-based thresholds gives user-specified coverage at level $1 - \alpha$ with a finite-sample guarantee. The outstanding challenge is that standard CP requires exchangeability between calibration and test points — a condition violated in non-stationary machinery data. Weighted conformal prediction [41] and localized conformal methods offer theoretical frameworks for this, but their computational feasibility on embedded edge hardware has not been studied in the fault diagnosis context.

3. Physics-informed synthesis for single-source class completion. In multi-source settings, missing fault classes can be borrowed from auxiliary source domains [8]. In strictly single-source deployments, generating physically realistic missing-class samples from data alone is ill-posed. Embedding PINN constraints or Invariant Risk Minimization priors [31] into the generative pipeline offers a pathway to synthesize missing fault signatures from rotor-bearing dynamic equations rather than data statistics. This direction connects fault diagnosis research to the broader literature on physics-constrained generative models.

4. Zero-shot cross-equipment diagnosis via disentangled representations. Few-shot meta-learning approaches [2, 18] require a small target-domain support set to adapt the metric space. Achieving zero-shot transfer requires separating machine-specific dynamic invariants (geometry, bearing kinematics, stiffness) from fault-specific transient signatures (impulse periodicity, sub-band energy) at the representation level. This decomposition is related to causal representation learning, pointing toward a convergence between the fault diagnosis and causal inference communities.

5. Foundation models and large-scale self-supervised pre-training for IIoT. The success of self-supervised pre-training in natural language processing and computer vision motivates a parallel effort for industrial machinery data: training large-parameter models on heterogeneous multi-equipment vibration corpora, then fine-tuning for specific fault types with minimal labeled examples. Two challenges are unique to this setting. First, publicly available multi-equipment vibration datasets [9] are orders of magnitude smaller than image or text corpora, making direct Transformer pre-training data-hungry. Second, foundation-model representations must capture physically meaningful sub-band fault signatures rather than statistical co-occurrence patterns. Wavelet-structured tokenization [42, 43] offers a principled interface between physical signal processing and sequence-model architectures. Addressing both challenges is likely to require federated pre-training across industry partners, raising privacy-preserving machine learning as a co-design requirement [40].

6. CONCLUSIONS

The past decade of intelligent fault diagnosis research reveals a consistent pattern: as deployment conditions become more constrained, purely data-driven approaches require progressively stronger structural regularization to remain reliable. The proposed six-tier taxonomy organizes this development by making explicit which assumptions each paradigm requires and where those assumptions break down under realistic industrial conditions. Convergent results from independent research groups on attention mechanisms [42, 43] and self-supervised pre-training [9] confirm that frequency-structured inductive biases improve fault discrimination independent of whether the training objective is supervised, adversarial, or contrastive - corroborating the CA-MCNN design and supporting Category VI as a substantive methodological direction.

The empirical record shows concrete progress: WPA and PDG improved SSDG accuracy over strong baselines on the JNU benchmark, while prototype-aware OSR frameworks elevated the H-score without requiring target-domain threshold calibration. The principal lesson from this comparative analysis is that global distributional alignment is unreliable when source label spaces are incomplete - a problem that generative class completion and prototype-anchored alignment address more directly than global domain confusion.

Looking ahead, five open problems span the research agenda: OS-SSDG under conformal coverage guarantees, calibrated rejection thresholds, physics-informed class completion, zero-shot disentanglement, and foundation-model pre-training for IIoT. Across all five, the combination of Conformal Prediction, physics-informed generative models, and wavelet-structured latent spaces represents the most principled direction for replacing empirical heuristics with bounded, interpretable uncertainty estimates in industrial fault diagnosis.

ACKNOWLEDGEMENTS

This study was supported by the Serbian Ministry of Science, Technological Development and Innovation from the grant 451-03-34/2026-03/200108.

REFERENCES

[1] H. Tao, P. Wang, Y. Chen, V. Stojanovic, and H. Yang, "An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks", Journal of the Franklin Institute, Vol. 357(12), pp. 7525–7545, https://doi.org/10.1016/j.jfranklin.2020.04.024, (2020)

[2] H. Tao, L. Cheng, J. Qiu, and V. Stojanovic, "Few shot cross equipment fault diagnosis method based on parameter optimization and feature metric", Measurement Science and Technology, Vol. 33(11), p. 115005, https://doi.org/10.1088/1361-6501/ac8368, (2022)

[3] Y. Sun, H. Tao, and V. Stojanovic, "Autoregressive data generation method based on wavelet packet transform and cascaded stochastic quantization for bearing fault diagnosis under unbalanced samples", Engineering Applications of Artificial Intelligence, Vol. 138, p. 109402, https://doi.org/10.1016/j.engappai.2024.109402, (2024)

[4] Y. Sun, H. Tao, and V. Stojanovic, "Pseudo-label guided dual classifier domain adversarial network for unsupervised cross-domain fault diagnosis with small samples", Advanced Engineering Informatics, Vol. 64, p. 102986, https://doi.org/10.1016/j.aei.2024.102986, (2025)

[5] Y. Sun, H. Tao, Y. Ni, and V. Stojanovic, "A generic single-source domain generalization framework for fault diagnosis via wavelet packet augmentation and pseudo-domain generation", IEEE Internet of Things Journal, Vol. 12(15), pp. 31629-31642, https://doi.org/10.1109/JIOT.2025.3573752, (2025)

[6] Y. Sun, H. Tao, and V. Stojanovic, "Multi-domain weakly decoupled domain generalization network for fault diagnosis under unknown operating conditions", Knowledge-Based Systems, Vol. 330, p. 114452, https://doi.org/10.1016/j.knosys.2025.114452, (2025)

[7] Y. Sun, H. Tao, and V. Stojanovic, "Open-set classification method via latent representation prompt and time-frequency fusion toward unknown fault recognition", Advanced Engineering Informatics, Vol. 68, p. 103779, https://doi.org/10.1016/j.aei.2025.103779, (2025)

[8] B. Wang, Y. Sun, H. Tao, and V. Stojanovic, "Collaborative completion and alignment for class-incomplete multi-source domain adaptation in rotating machinery fault diagnosis", Nonlinear Dynamics, Vol. 114, p. 324, https://doi.org/10.1007/s11071-025-12203-y, (2026)

[9] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap", Mechanical Systems and Signal Processing, Vol. 138, p. 106587, https://doi.org/10.1016/j.ymssp.2019.106587, (2020)

[10] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring", Mechanical Systems and Signal Processing, Vol. 115, pp. 213–237, https://doi.org/10.1016/j.ymssp.2018.05.050, (2019)

[11] M. Chen, J. Zheng, H. Pei, L. Yang, Q. Zhang, Q. Han, and H. Luo, "Noise-robust fault diagnosis of rotating machinery: a systematic review and methodological roadmap", Measurement Science and Technology, Vol. 37, p. 192002, https://doi.org/10.1088/1361-6501/ae65b7, (2026)

[12] T. T. Feisa, H. S. Gebremedhen, F. Kibrete, D. E. Woldemichael, S. D. Ageze, and H. D. Nega, “A comprehensive review of deep transfer learning for intelligent fault diagnosis in rotating machines”, Engineering Reports, Vol. 7(11), p. e70494, https://doi.org/10.1002/eng2.70494, (2025)

[13] W. Fu, S. Li, B. Wen, B. Zheng, W. Liao, and C. Tan, “A review of rolling bearing fault diagnosis: data preprocessing and model optimization”, Measurement Science and Technology, Vol. 36(6), p. 062002, https://doi.org/10.1088/1361-6501/add7fb, (2025)

[14] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals”, Sensors, Vol. 17(2), p. 425, https://doi.org/10.3390/s17020425, (2017)

[15] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, “Real-time motor fault detection by 1-D convolutional neural networks”, IEEE Transactions on Industrial Electronics, Vol. 63(11), pp. 7067–7075, https://doi.org/10.1109/TIE.2016.2582729, (2016)

[16] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, “Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data”, Mechanical Systems and Signal Processing, Vol. 72–73, pp. 303–315, https://doi.org/10.1016/j.ymssp.2015.10.025, (2016)

[17] Mingyue Y., Yongpeng L., Guanglei M., and Y. Wang, “An intelligent fault diagnosis for rotating machine under strong noise based on cross-attention-driven spatial-temporal feature fusion and duplexing time sequence convolution optimization”, Engineering Applications of Artificial Intelligence, Vol. 153, p. 110897, https://doi.org/10.1016/j.engappai.2025.110897, (2025)

[18] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks”, Proceedings of the 34th International Conference on Machine Learning, Vol. 70, PMLR, pp. 1126–1135, (2017)

[19] L. Wen, X. Li, L. Gao, and Y. Zhang, “A new convolutional neural network-based data-driven fault diagnosis method”, IEEE Transactions on Industrial Electronics, Vol. 65(7), pp. 5990–5998, https://doi.org/10.1109/TIE.2017.2774777, (2018)

[20] J. Snell, K. Swersky, and R. Zemel, “Prototypical networks for few-shot learning”, Advances in Neural Information Processing Systems, Vol. 30, pp. 4080-4090, https://doi.org/10.48550/arXiv.1703.05175, (2017)

[21] G. Douzas, F. Bacao, and F. Last, “Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE”, Information Sciences, Vol. 465, pp. 1–20, https://doi.org/10.1016/j.ins.2018.06.056, (2018)

[22] X. Wang, C. Zhang, Y. Wang, Y. Xu, Z. Liu, and H. Jiang, “Fault diagnosis of rolling bearings with imbalanced samples using conditional temporal denoising diffusion generative adversarial network”, Measurement, Vol. 283, p. 122036, https://doi.org/10.1016/j.measurement.2026.122036, (2026)

[23] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei, and F. Zhou, “A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data”, IEEE Transactions on Industrial Informatics, Vol. 20(5), pp. 7820–7831, https://doi.org/10.1109/TII.2024.3366991, (2024)

[24] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models”, Advances in Neural Information Processing Systems, Vol. 574, pp. 6840 - 6851, (2020)

[25] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks”, Journal of Machine Learning Research, Vol. 17(1), pp. 1–35, https://doi.org/10.48550/arXiv.1505.07818, (2016)

[26] M. Long, Y. Cao, J. Wang, and M. Jordan, “Learning transferable features with deep adaptation networks”, Proceedings of the 32nd International Conference on Machine Learning, Vol. 37, PMLR, pp. 97–105, (2015)

[27] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, “Maximum classifier discrepancy for unsupervised domain adaptation”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3723–3732, https://doi.org/10.1109/CVPR.2018.00392, (2018)

[28] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, “A theory of learning from different domains”, Machine Learning, Vol. 79(1-2), pp. 151–175, https://doi.org/10.1007/s10994-009-5152-4, (2010)

[29] J. Liang, D. Hu, and J. Feng, “Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation”, International Conference on Machine Learning, ICML, Vol. 560, pp. 6028–6039, (2020)

[30] D. Krueger, E. Caballero, J.-H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. L. Priol, and A. Courville, “Out-of-distribution generalization via risk extrapolation”, Proceedings of the 38th International Conference on Machine Learning, Vol. 139, pp. 1-30, ICLR, https://doi.org/10.48550/arXiv.2003.00688, (2021)

[31] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, “Invariant risk minimization”, arXiv preprint, pp. 1-31, https://doi.org/10.48550/arXiv.1907.02893, (2019)

[32] Y. Zhang, M. Li, R. Li, K. Jia, and L. Zhang, “Exact feature distribution matching for arbitrary style transfer and domain generalization”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8025–8035, https://doi.org/10.1109/CVPR52688.2022.00787, (2022)

[33] G. Chen, P. Peng, X. Wang, and Y. Tian, “Adversarial reciprocal points learning for open set recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 44(11), pp. 8065–8081, https://doi.org/10.1109/TPAMI.2021.3106743, (2022)

[34] Z. Wang, X. Yang, T. Li, L. She, X. Guo, and F. Yang, “Intelligent Fault Diagnosis for Rotating Machinery via Transfer Learning and Attention Mechanisms: A Lightweight and Adaptive Approach”, Actuators, Vol. 14(9), p. 415, https://doi.org/10.3390/act14090415, (2025)

[35] P. Ding, Y. Shan, Z. Fei, and X.-M. Sun, “A Novel Generative Domain Generalization Network for Intelligent Fault Diagnosis Under Unknown Operating Conditions”, IEEE Transactions on Instrumentation and Measurement, Vol. 74, p. 3562613, https://doi.org/10.1109/TIM.2025.3614879, (2025)

[36] X. Yu, Z. Zhao, X. Zhang, Q. Zhang, Y. Liu, and C. Sun, “Deep-learning-based open set fault diagnosis by extreme value theory”, IEEE Transactions on Industrial Informatics, Vol. 18(1), pp. 185–196, https://doi.org/10.1109/TII.2021.3070324, (2022)

[37] A. Bendale and T. E. Boult, “Towards open set deep networks”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1563–1572, https://doi.org/10.1109/CVPR.2016.173, (2016)

[38] M. Mehralian and B. Karasfi, “RDCGAN: Unsupervised Representation Learning With Regularized Deep Convolutional Generative Adversarial Networks”, Conference on Artificial Intelligence and Robotics, 8769811, https://doi.org/10.1109/AIAR.2018.8769811, (2018)

[39] C. Li, S. Zhang, Y. Qin, and E. Estupinan, “A systematic review of deep transfer learning for machinery fault diagnosis”, Neurocomputing, Vol. 407, pp. 121–135, https://doi.org/10.1016/j.neucom.2020.04.045, (2020)

[40] D. Yang, W. Zhang, and Y. Jiang, “Mechanical fault diagnosis based on deep transfer learning: a review”, Measurement Science and Technology, Vol. 34(11), p. 112001, https://doi.org/10.1088/1361-6501/ace7e6, (2023)

[41] I. Gibbs and E. Candes, “Adaptive conformal inference under distribution shift”, Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS), Vol. 128, pp. 1660 - 1672, https://doi.org/10.48550/arXiv.2106.00170, (2021)

[42] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141, https://doi.org/10.1109/CVPR.2018.00745, (2018)

[43] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, “Deep residual shrinkage networks for fault diagnosis”, IEEE Transactions on Industrial Informatics, Vol. 16(7), pp. 4681–4690, https://doi.org/10.1109/TII.2019.2943898, (2020)