American Journal of Science, Engineering and Technology
2026, Vol. 11, No. 3, pp. 212–224
https://doi.org/10.11648/j.ajset.20261103.21

Research Article

A Physics-Data Dual-Channel Collaborative Cross-Domain Transfer Learning Fault Diagnosis Method for Main Circulation Pump in Converter Valve Cooling Systems

Gao Lei, Wang Wenchao, Wu Heng, Zhang Yu*, He Biao

State Grid Shaanxi Electric Power Co., Ltd., Xi'an, China

Abstract

To address the three major challenges in fault diagnosis of main circulating pumps in converter station valve cooling systems—small-sample scenarios, cross-domain distribution shift, and lack of physical interpretability—a physics-data dual-channel collaborative cross-domain transfer learning framework is proposed. The physics channel designs a physics-constrained conditional generative adversarial network (PC-cGAN), which explicitly embeds the pump's Euler equation into the loss function to constrain the hydraulic characteristics of generated samples. The data channel constructs a multi-source signal attention fusion network (MSAF-Net), achieving cross-modal fusion of vibration, temperature, and pressure features through a physically guided mechanism. Theoretically, this paper establishes a quantitative relationship between the Wasserstein distance and the HΔH divergence, deriving an explicit upper bound for domain adaptation error. Validation on measured data from 69 main circulating pumps across five UHV converter stations shows that under the stringent condition where only 5% of target domain samples are labeled, the cross-domain diagnostic accuracy reaches 89.7%, an improvement of 13.4 percentage points over the optimal baseline. The Pearson correlation coefficient between the model's decision logic and the fault physical mechanism is 0.82 ($p < 0.001$), indicating that the diagnostic results possess clear physical interpretability.

Keywords

Converter Station, Valve Cooling System, Main Circulating Pump, Fault Diagnosis, Transfer Learning, Small-sample Learning, Physical Constraints, Cross-domain Adaptation

1. Introduction

The main circulating pump of the valve cooling system is a critical auxiliary component for ensuring the thermal stability of the converter valve, and its sudden failure may cause a lock-out of the DC transmission system [1, 2]. The urgent need for intelligent operation and maintenance in ultra-high voltage (UHV) projects is significantly at odds with the limited volume of on-site data, substantial differences in operating condition distribution, and the weak physical interpretability of existing deep learning models [3, 4]. Establishing a high-precision, physically consistent diagnostic model under small-sample and cross-domain operating conditions remains a bottleneck that urgently needs to be addressed.

Recent research on fault diagnosis of main circulating pumps in converter valve cooling systems can be categorized into three main streams. First, traditional signal-processing-

*Correspondence: Zhang Yu (517571853@qq.com)

Received: 30 July 2026; **Accepted:** 8 August 2026; **Published:** 22 September 2026

based methods, such as vibration spectrum analysis and temperature threshold monitoring, have been widely applied to pump condition monitoring [1, 2]. However, these methods rely heavily on expert experience and manual feature extraction, making them difficult to adapt to complex cross-station operating conditions. Second, data-driven intelligent diagnostic methods have been introduced to improve diagnostic automation. For instance, Zhou et al. [28] proposed an ensemble fault diagnosis model based on deep reinforcement learning for main circulation pumps in VSC-HVDC transmission systems, achieving 95.00% accuracy. Mei et al. [29] developed a lightweight fault diagnosis model based on multidimensional vibration feature graphs for ultra-high voltage converter valve cooling systems. Zhao et al. [30] further analyzed the fault causes and hidden danger management of main circulating pumps in flexible DC converter stations. Moreover, a CWT-based dilated 2DCNN method was specifically designed for the main circulation pump of VSC-HVDC valve cooling systems. Nevertheless, these methods still require a large number of labeled fault samples or focus on specific fault types, which are scarce in high-reliability UHV systems. Third, with the advancement of deep learning, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown superior feature extraction capabilities in rotating machinery fault diagnosis [5-9]. Yet, most of these studies focus on generic rotating machinery (e.g., bearings, gearboxes) rather than the specific hydraulic–mechanical coupling characteristics of main circulating pumps in valve cooling systems. In particular, existing literature rarely addresses the three critical technical bottlenecks simultaneously: (i) diagnostic reliability degradation under small-sample conditions, (ii) cross-domain distribution shift caused by differences in pump manufacturers, models, and operating conditions across converter stations, and (iii) the lack of physical interpretability in pure data-driven models. Consequently, there remains a significant gap between current research and the practical demands of intelligent operation and maintenance for UHV valve cooling systems.

Existing diagnostic models face three major constraints in this scenario. First, diagnostic reliability is insufficient in small-sample scenarios. The high-reliability design of power systems leads to a scarcity of fault samples, and traditional deep learning struggles with data imbalance [5, 6]. Although meta-learning reduces sample dependency through task-level knowledge transfer [7, 8], its generalization capability significantly declines in cross-domain scenarios [9]. Second, cross-domain transfer capability is weak. Differences in manufacturers, models, and operating conditions of main circulating pumps across stations cause shifts in signal statistical distribution [10, 11]. While domain adaptation techniques can align feature distributions [12, 13], they require extensive fine-tuning with labeled target domain samples, which severely deviates from reality in terms of scarce labeling [14, 15]. Semi-supervised domain adaptation reduces labeling dependency [16], but existing methods based on JS divergence or adversarial training risk gradient saturation when the distribution support sets do not overlap, and lack physical constraints. Third, multi-source information fusion and physical mechanism constraints are absent. Current research predominantly focuses on single vibration signals [17-20], failing to integrate multi-source information such as temperature and pressure. Although physics-informed neural networks introduce hydraulic constraints [21-23], they do not address the coupling challenges of small-sample cross-domain scenarios, and the decoupled generation and learning tasks lead to limited physical interpretability.

To address the above issues, this paper proposes a physics-constrained generation-learning dual-channel collaborative architecture that integrates physical prior constraints (Figure 1). The architecture connects the generation and learning channels through a physical constraint layer: on the generation side, PC-cGAN embeds the Euler equations into the loss function to constrain the sample augmentation process with hydraulic characteristics; on the learning side, MSAF-Net introduces physically guided multi-head cross-modal attention to integrate multi-source heterogeneous information such as vibration, temperature, and pressure; the domain adaptation module is driven by the Wasserstein distance, which utilizes its quantitative relationship with the HΔH divergence to constrain the alignment error of feature distributions. The contributions of this paper are threefold: Theoretically, it derives an explicit upper bound for domain adaptation errors by leveraging the Lipschitz duality between the Wasserstein distance and the HΔH divergence, providing gradient stability guarantees for stable transfer under non-overlapping distributions. Methodologically, it constructs a generation-learning dual-channel collaborative architecture connected by a physical constraint layer, achieving physically guided fusion of multi-source heterogeneous information through a cross-modal attention mechanism. In terms of application, under the stringent condition where only 5% of target domain samples are labeled, the cross-domain diagnostic accuracy reaches 89.7%, and the decision logic is highly correlated with the fault mechanism (Pearson $r = 0.82$), offering a highly reliable solution for UHV valve cooling operation and maintenance.

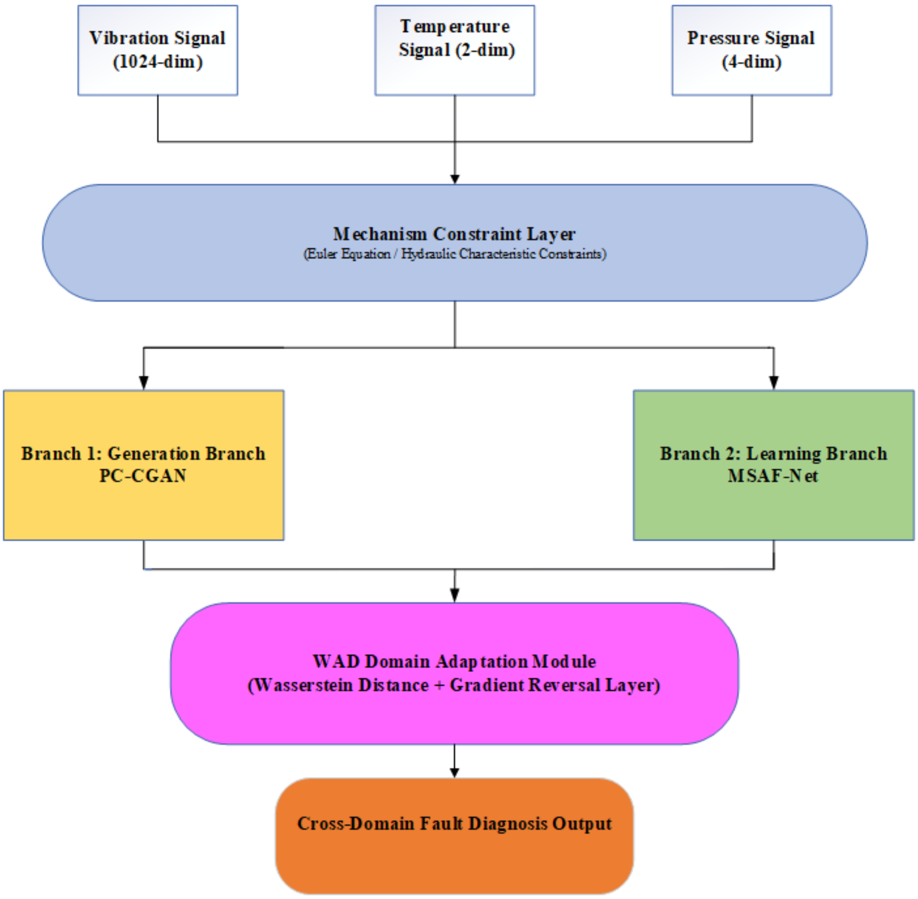

Figure 1. *Generation-Learning Collaborative Transfer Learning Architecture with Physics-Based Prior Constraints.*

2. Theoretical Foundations and Problem Modeling

2.1. Formal Definition of the Problem

Let the source domain be $D_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$, containing n_s labeled samples. The input feature $\mathbf{x}_i^s \in \mathbb{R}^{1030}$ comprises multi-domain statistics of vibration, temperature, and pressure signals; the label $y_i^s \in \{0,1,\dots,4\}$ corresponds to health status and four types of typical failures (bearing wear, mechanical seal failure, impeller cavitation, and coupling misalignment). The target domain $D_t = \{(\mathbf{x}_j^t, y_j^t)\}_{j=1}^{n_t^l} \cup \{\mathbf{x}_k^t\}_{k=1}^{n_t^u}$ contains n_t^l labeled samples and n_t^u unlabeled samples, with the labeling ratio not exceeding 5% under extreme conditions. The probability distributions of the source domain and the target domain are different, i.e., $P_s(\mathbf{x}, y) \neq P_t(\mathbf{x}, y)$, indicating the presence of domain drift.

Define the feature extraction mapping $f_\theta: \mathcal{X} \to \mathcal{Z}$, which projects the input from $\mathbb{R}^{1030}$ (vibration: 1024-dim, temperature: 3-dim, pressure: 3-dim) into a 256-dimensional feature space; the classifier $g_\phi: \mathcal{Z} \to \mathbb{R}^5$ outputs class predictions. The optimization objective is to minimize the expected risk on the target domain:

$$\min_{\theta,\phi} R_t(f_\theta, g_\phi) = \mathbb{E}_{(\mathbf{x},y)\sim P_t}\left[L_{CE}\left(g_\phi(f_\theta(\mathbf{x})), y\right)\right] \quad (1)$$

Here L_{CE} is the cross-entropy loss. Due to the scarcity of labeled samples in the target domain, practical optimization relies on source domain supervision and domain adaptation constraints, as detailed in Section 3.4.

2.2. Theoretical Boundaries of Domain Adaptation

Lemma 1 (Lipschitz Continuity of Feature Extractor [24]) Let the MSAF-Net feature extractor f_θ be an L-layer neural network, with activation function σ satisfying $\mathrm{Lip}(\sigma) \leq 1$. Then there exists a finite constant $L_f^\theta = \prod_{l=1}^{L} \sigma_{\max}(W_l)$ such that for any x_1, x_2, $\| f_\theta(x_1) - f_\theta(x_2) \|_2 \leq L_f^\theta \| x_1 - x_2 \|_2$ holds.

Theorem 1 (Upper bound on domain adaptation error) Let H be a set of measurable classification functions, $d_{H\Delta H}(P_s, P_t)$ be the $H\Delta H$ distance, $\lambda^* = \min_{h\in H}[R_s(h) + R_t(h)]$. Assume H consists of L_h-Lipschitz functions. Then for any $h \in H$:

$$R_t(h) \leq R_s(h) + \frac{1}{2} d_{H\Delta H}(P_s, P_t) + \lambda^* \tag{2}$$

In the framework of integral probability metrics, if H is composed of L_h-Lipschitz functions, based on the Lipschitz duality between integral probability metrics and the Wasserstein distance [25], it follows that:

$$d_{H\Delta H}(P_s, P_t) \leq 2L_h W_1(P_s, P_t) \tag{3}$$

Let $h = g_\phi \circ f_\theta$ denote the composite hypothesis. Substituting (3) into (2) yields the explicit upper bound:

$$R_t(h) \leq R_s(h) + L_h W_1(P_s, P_t) + \lambda^* \tag{4}$$

This bound indicates that the target domain risk is explicitly defined by the source domain risk, the Wasserstein distance, and the joint optimal error. W_1 is continuously differentiable under a weak topology, enabling effective gradient propagation even when the distribution supports have no overlap, thus avoiding the gradient saturation issue of JS divergence in the early stages of domain adaptation [26].

This paper adopts the Wasserstein-1 distance as a proxy measure for distribution discrepancy.

$$W_1(P_s, P_t) = \inf_{\gamma \in \Pi(P_s, P_t)} \mathbb{E}_{(x^s, x^t) \sim \gamma}[\| \mathrm{x}^s - \mathrm{x}^t \|_2] \tag{5}$$

Here, $\Pi(P_s, P_t)$ is the family of joint distributions with marginal distributions P_s and P_t, γ is the optimal transport plan, and inf denotes the infimum. $\|\cdot\|_2$denotes the Euclidean distance.

2.3. Physics-constrained Embedding Framework

Let the impeller outlet radius be r_2, the outlet width be b_2, the blade outlet angle be β_2, and the rated speed be n. Then the outlet circumferential velocity is $u_2 = \pi D_2 n/60$ ($D_2 = 2r_2$). Ignoring inlet pre-swirl, the theoretical head-flow relationship is obtained from the Euler equation:

$$H_{\text{theory}}(Q) = \frac{u_2^2}{\mathrm{g}} - \frac{u_2 \cot\beta_2}{\mathrm{g} \cdot 2\pi r_2 b_2} \; Q = a_0 - a_1 Q \tag{6}$$

In the formula, $\mathrm{g} = 9.8\mathrm{m/s^2}, a_0 = u_2^2/g, and a_1 = u_2 \cot\beta_2/(2\pi g r_2 b_2)$. Due to frictional and impact losses in actual centrifugal pumps, the head curve exhibits a downward parabolic shape. In this paper, the theoretical straight line is taken as the trend constraint baseline, aiming to limit the hydraulic characteristic trend of the generated samples rather than precise regression. Under fault conditions, the actual H-Q curve deviates from the theoretical value, and this deviation can serve as a distinguishing feature.

To embed the above physical laws into the generation process, it is necessary to invert the hydraulic parameters from the generated samples. Assuming that after decoding, the generated samples yield head estimates $\hat{H}$, flow rate estimates $\hat{Q}$, and power estimates $\hat{P}$, the physical consistency loss of the generation channel is defined as:

$$L_{\text{phy}}^{(G)} = \lambda_H^{(G)} \frac{\|\hat{H} - H_{\text{theory}}(\hat{Q})\|_1}{H_{\text{rated}}} + \lambda_P^{(G)} \frac{\|\hat{P} - P_{\text{theory}}(\hat{Q})\|_1}{P_{\text{rated}}} \tag{7}$$

where $\lambda_H^{(G)}, \lambda_P^{(G)}$ are weighting coefficients for the generation channel (determined via grid search in Section 3.2.2).

It should be noted that Equation (6) is defined in this paper as the baseline trend of head under healthy operating conditions, rather than a rigid physical equation. In fault states (such as a sudden drop in head due to cavitation in the impeller), the deviation of the measured H-Q curve from this baseline (i.e., the residual $\hat{H} - H_{theory}(\hat{Q})$) itself serves as an important distinguishing feature. Therefore, the physical loss term in Equation (7) is designed as a weak regularization term (with weight $\lambda_H^{(G)} = 0.6$ in Section 3.2.2.) Its core function is to eliminate abnormal points in the generated samples that severely violate the fundamental physical laws of pumps, such as the " positive correlation between flow and head," rather than to suppress legitimate hydraulic fluctuations caused by faults.

3. Physics-Constrained Generation-Learning Dual-Channel Collaborative Method

3.1. Method Overview

This section elaborates on three core modules in sequence: the PC-cGAN generation channel augments data under the constraints of Euler equations to address sample scarcity; the MSAF-Net learning channel targets multi-source signals and integrates vibration, temperature, and pressure information through cross-modal attention; the Wasserstein Domain Adaptation (WDA) module handles cross-domain distribution shifts by aligning source-target domain features using the Wasserstein distance. The three modules work together to achieve end-to-end cross-domain fault identification.

3.2. PC-cGAN Generation Channel

3.2.1. Network Architecture Design

PC-cGAN consists of a generator G and a discriminator D, both of which are fully connected structures with parameter configurations shown in Table 1. Input Dimension: 105 denotes the concatenated dimension of noise and condition vectors.

Table 1. Network architecture parameters of PC-cGAN.

Component	Layer Type	Input Dimensions	Output Dimension	Activation Function / Normalization
Generator G	Input layer	105	256	LeakyReLU(0.2)+BN
	Hidden Layer 1	256	512	LeakyReLU(0.2)+BN
	Hidden Layer 2	512	512	LeakyReLU(0.2) + BN
	Hidden Layer 3	512	256	LeakyReLU(0.2)+BN
	Output Layer	256	1030	tanh
Discriminator D	Input Layer	1030	512	LeakyReLU(0.2)
	Hidden Layer 1	512	256	LeakyReLU(0.2)
	Hidden Layer 2	256	128	LeakyReLU(0.2)
	Output layer	128	1	Linear

The generator input is the concatenation of 100-dimensional noise $z \sim \mathcal{N}(0,I)$ and a 5-dimensional condition vector c, outputting a 1030-dimensional sample. The discriminator takes the sample and condition vector as input, outputting a scalar score. The discriminator weights are clipped to $[-0.01,0.01]$ to ensure 1-Lipschitz continuity, stabilizing the Wasserstein distance approximation.

3.2.2. Physical Constraint Training Objectives

PC-cGAN is trained as a constrained minimax problem:

$$\min_{\theta_g}\max_{\theta_d} L_{PC-cGAN} = L_d + \lambda_{phy}^{(G)} L_{phy}^{(G)} + \lambda_{cond} L_{cond} + \lambda_{gp} L_{GP} \quad (8)$$

Adversarial Loss (Wasserstein Form):

$$L_d = \mathbb{E}_{x\sim P_s, c\sim P_c}[D(\mathrm{x}|c)] - \mathbb{E}_{z\sim P_z, c\sim P_c}[D(G(z|c))] \quad (9)$$

In the formula, P_z=$\mathcal{N}(0,I)$ and P_c are conditional vector distributions.

Physics consistency loss: The generated samples are inverted to obtain,, and, which are constrained to follow the Euler equation trend. An H-Q curve fitting term is added, where a_0 and a_1 are the Euler equation coefficients defined in (6):

$$L_{\mathrm{HQ}} = \frac{1}{N}\sum_{i=1}^{N} \frac{|\hat{H}_i - (a_0 - a_1\hat{Q}_i)|}{H_{\mathrm{rated}}} \quad (10)$$

Through grid search, the weights are determined as $\lambda_H^{(G)} = 0.6, \lambda_P^{(G)} = 0.3$, and $\lambda_{curve} = 0.5$, where a_0 and a_1 correspond to the coefficients in (6).

Conditional Consistency Loss: The auxiliary classifier C ensures that the generated samples are consistent with the specified category.

$$L_{cond} = -\mathbb{E}_{z\sim P_z, y\sim P_c}[y^{\top}\log(C(G(z|y)))] \quad (11)$$

Gradient Penalty (Lipschitz Constraint):

$$L_{GP} = \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

$$\hat{\mathrm{x}} = \epsilon \mathrm{x} + (1-\epsilon)G(z,c), \epsilon \sim U[0,1] \quad (12)$$

Training parameters: batch size 64, Adam optimizer (learning rate 0.0002, $\beta_1 = 0.5, \beta_2 = 0.999$), gradient penalty coefficient $\lambda_{gp} = 10$. 500 samples are generated for each type of fault, and the source domain training set is expanded from 1260 to 3260.

3.2.3. Comparative Analysis with Conventional cGAN

Conventional cGAN relies solely on the adversarial loss (Eq. (9)) and conditional consistency loss (Eq. (11)) to drive the generator, with its optimization objective focusing only on the goodness of distribution fitting of samples on the data manifold, without imposing any inductive bias on the inherent hydraulic characteristics of the pump. This means that the generator may produce fake samples through "shortcut learning" that appear statistically reasonable but severely violate the trends of Euler's equations (e.g., generated samples exhibit the unphysical property of "head rising as flow rate increases"). Although such samples can deceive the discriminator, they contaminate the physical consistency of downstream classifiers, leading to degraded cross-domain generalization performance.

In contrast, PC-cGAN explicitly embeds the H-Q and P-Q curve constraints (Equation (10)) into the generator loss function, forcing the generated samples to maintain a monotonically decreasing physical trend between head, power, and flow rate. This constraint is not a hard restriction but exists as a weak regularization term (weight $\lambda_H^{(G)} = 0.6$), allowing the generated samples to have reasonable fluctuations around the theoretical baseline, thereby achieving a balance between fidelity and diversity.

3.3. MSAF-Net Learning Channel

3.3.1. Multi-source Feature Encoding

To achieve deep fusion of heterogeneous signals (vibration, temperature, and pressure), MSAF-Net adopts a hierarchical feature encoding architecture, with its structural configuration detailed in Table 2.

Table 2. *Feature encoding module architecture of MSAF-Net.*

	Encoding method	Key Parameters	Output dimension
Vibration signal	Three-layer 1D-CNN + Global Average Pooling	Number of convolution kernels: 16, 32, 64; kernel size: 3; pooling size: 2	256
Temperature/Pressure Signal	Statistical Feature Extraction + two-layer MLP	Hidden layers: 16; Statistical feature dimension: 6	32→256 (after Projection)

The vibration signal extracts multi-scale features through three layers of one-dimensional convolution, and the initialization of the first convolutional kernel incorporates the prior knowledge of fault characteristic frequency.

$$W_1^{vib(init)} = W_{Xavier} + \gamma \cdot F_{BPFO}, \gamma = 0.1 \tag{13}$$

Here F_{BPFO} is the bandpass filter response matrix constructed based on the outer race fault characteristic frequency, with the same shape as W_{Xavier}. The convolution output is projected into a 256-dimensional space z_{vib} after global average pooling. The temperature and pressure signals extract 6-dimensional statistics such as mean and variance, which are encoded into a 32-dimensional representation z_{TP} through two fully connected layers, and then projected into 256 dimensions z'_{TP} to align the dimensions.

3.3.2. Cross-modal Attention Fusion

MSAF-Net employs physically guided multi-head cross-modal attention with $h = 4$ heads, each with dimension $d_h = 64$. For the -th head, the query, key, and value are computed as $q_i = W_Q^{(i)}, k_i = W_K^{(i)} z'_{TP} z_{vib}$, and $v_i = W_V^{(i)} z'_{TP}$, respectively. The projection matrix initialization introduces a physically guided vector m_i:

$$W_Q^{(i,init)} = W_{Xavier}^{(i)} + \gamma \cdot m_i 1_{256}^T \tag{14}$$

Here, the guiding vector $m_i \in \mathbb{R}^{d_h}$, with its kth element $m_{i,k} = \cos(k \cdot \omega_i) \cdot \exp(-k^2/(2\sigma^2))$, ω_i corresponds to BPFO, BPFI, BPF, and rotational frequency, $\sigma = d_h/4$. The design of the guiding vector aims to enhance the prior sensitivity of the projection matrix to specific fault characteristic frequencies and their sidebands, with its frequency-domain envelope characteristics resembling those of a bandpass filter response. The attention weights and output are computed as follows:

$$\alpha_i = \sigma\left(\frac{q_i \odot k_i}{\sqrt{d_h}}\right), o_i = \alpha_i \odot v_i \tag{15}$$

Here $\odot$ denotes element-wise multiplication (Hadamard product), and $\sigma(\cdot)$ denotes the sigmoid function.

Multi-head Fusion and Output:

$$O = \mathrm{Concat}\,[o_1; \ldots; o_h], z_{cross} = W_O \cdot O,$$
$$f_{fuse} = \mathrm{LayerNorm}(z_{cross} + z_{vib}) \tag{16}$$

Where W_O is the output projection matrix, z_{cross} is the attention-weighted output of temperature-pressure features, and it is fused with the original vibration feature z_{vib} through a residual connection to ensure the dominance of vibration information. This mechanism achieves adaptive cross-modal weighting at the feature level, and the multi-head structure can capture differentiated correlations under different fault modes.

3.4. WDA Domain Adaptation Module

3.4.1. Adversarial Domain Adaptation Architecture

The WDA module consists of a gradient reversal layer (GRL) and a Wasserstein domain discriminator. The GRL performs identity mapping during forward propagation and negates the gradient during backpropagation.

$$R_\lambda(f) = f, \frac{\partial R_\lambda}{\partial f} = -\lambda(t) I \tag{17}$$

Time-varying weights $\lambda_{dom}(t) = \lambda_{\max} \cdot [\frac{2}{1+\exp(-10t/t_{\max})} - 1]$, $\lambda_{\max} = 0.5$. The domain discriminator is a three-layer fully connected network (256→128→64→1), with weights clipped to $[-0.01,0.01]$ to maintain the 1-Lipschitz condition.

3.4.2. Training Objective Function

Total loss function:

$$L_{total} = L_{cls} + \lambda_{dom}(t) \cdot L_{dom} + \lambda_{phy}^{(F)} \cdot L_{phy}^{(F)} \quad (18)$$

The Superscripts(G) and (F) denote the generation and feature learning channels, respectively.

Classification loss (using only labeled samples):

$$L_{cls} = -\frac{1}{n_s^{(l)}+n_t^{(l)}} \sum_{i=1}^{n_s^{(l)}+n_t^{(l)}} \sum_{c=1}^{5} \mathbb{I}(y_i = c)\log(p_{ic}) \quad (19)$$

In the formula, p_{ic} represents the class prediction probability.

Domain discrimination loss (Wasserstein form, using all samples):

$$L_{dom} = \frac{1}{n_s}\sum_{i=1}^{n_s} D_w(f_i^s) - \frac{1}{n_t}\sum_{j=1}^{n_t} D_w(f_j^t) \quad (20)$$

In the formula, $D_w(\cdot)$ is the output of the Wasserstein domain discriminator.

Feature Physical Consistency Loss (using all samples): $\phi_{phy}: \mathbb{R}^5 \to \mathbb{R}^{256}$ is a fully connected projection network from physical parameters to feature space, which inverts physical quantities and reconstructs features through an autoencoder, constrained as follows:

$$L_{phy}^{(F)} = \frac{1}{256} \parallel f_{fuse} - \phi_{phy}(x_{phy}) \parallel_2^2 + \lambda_{eq} \cdot L_{eq} + \lambda_{freq} \cdot L_{freq} \quad (21)$$

Here, $x_{phy} = [\hat{H}, \hat{Q}, \hat{P}, \hat{f}_{BPFO}, \hat{f}_{BPFI}]^\top$ and L_{eq} are power equation constraints, L_{freq} is a frequency consistency constraint, and the weight is $\lambda_{phy}^{(F)} = 0.3, \lambda_{eq} = 0.2, \lambda_{freq} = 0.1$.

3.4.3. Training Strategy

A three-stage progressive training approach is adopted:

Phase 1 (Pre-training, 50 epochs): The domain discriminator is frozen, while the feature extractor and classifier are optimized with a learning rate of 0.001, using only L_{cls}.

Stage 2 (Adversarial Training, 100 epochs): All parameters are jointly optimized, $\lambda_{dom}(t)$ is gradually increased to 0.5, and the learning rate is set to 0.0005 with cosine annealing.

Phase 3 (Target Domain Fine-Tuning, 20 epochs): The feature extractor is frozen, and only the classifier is optimized, with a learning rate of 0.0001 and an early stopping patience of 5 epochs.

The batch size is 64. In phases 1 and 2, all source domain samples and unlabeled target domain samples are used; in Phase 3, only labeled target domain samples are used.

4. Experimental Verification and Result Analysis

4.1. Experimental Setup

4.1.1. Dataset Construction

The dataset was collected from 69 main circulation pumps across 5 UHV converter stations (Stations A to E), covering different voltage levels, power ratings, and rotational speeds. The source domain includes 43 devices from Stations A and B, while the target domain comprises 26 devices from Stations C, D, and E. The technical parameters and sample distribution are shown in Table 3.

Table 3. Technical parameters and original segment distribution of main circulation pumps in converter stations.

Converter Station	Voltage Level / kV	Rated Power / kW	Rated Flow Rate / (m^3/h)	Rated Head / m	Rotational Speed / ($r \cdot min^{-1}$)	No. of Devices (H/F)	No. of Samples (H/F)	Domain
Station A	±800	30	250	55	1450	15/5	300/100	Source
Station B	±800	45	400	70	2900	14/9	280/180	Source
Source Subtotal	—	—	—	—	—	29/14	580/280	860
Station C	±800	22	200	45	1450	8/3	80/30	Target
Station D	±1100	37	320	60	2900	6/2	60/20	Target
Station E	±1100	30	250	55	1450	5/2	50/20	Target

Converter Station	Voltage Level / kV	Rated Power / kW	Rated Flow Rate / (m^3/h)	Rated Head / m	Rotational Speed / ($r\cdot min^{-1}$)	No. of Devices (H/F)	No. of Samples (H/F)	Domain
Target Subtotal	—	—	—	—	—	19/7	190/70	260
Total	—	—	—	—	—	48/21	770/350	1120

Note: H/F denotes Healthy/Faulty. The source domain comprises 43 devices (29 healthy, 14 faulty) with 860 samples, while the target domain comprises 26 devices (19 healthy, 7 faulty) with 260 samples. The sampling density in the source domain is 20 samples per device, whereas in the target domain it is 10 samples per device, reflecting differences in monitoring durations across stations.

The target domain comprises 260 original segments (190 healthy and 70 faulty samples) collected from Stations C, D, and E. Under the extreme condition of a 5% labeling ratio, 13 samples are selected for supervised fine-tuning. To eliminate statistical bias from random sampling, a stratified sampling strategy is adopted. The 13 labeled samples are allocated across five categories (one healthy state and four fault types) under two explicit constraints: (i) Minimum occupancy: each of the four fault categories is guaranteed at least two samples to ensure diagnostic coverage for rare failures, accounting for 8 of the 13 labels; (ii) Proportional allocation: the remaining 5 samples are distributed between the healthy state and the most prevalent fault categories according to their natural prevalence in the target domain. For fault categories whose target-domain proportion falls below this sampling threshold, category weight compensation is applied during classifier training to mitigate the resulting class imbalance. This sampling process is repeated for five independent trials with the random seed fixed at 2024, and the reported performance metrics (accuracy and macro-F1) are presented as mean ± standard deviation across the five runs, minimizing statistical bias from a single random realization.

Preprocessing: The vibration signal is sampled at 20 kHz, filtered through a fourth-order Butterworth bandpass filter (0.1 to 8 kHz), and then normalized using Z-score. Samples are constructed using a sliding window (window length 4096, 50% overlap). Temperature and pressure signals are sampled at 1 Hz, aligned via linear interpolation, and then 6-dimensional features are extracted: mean, variance, peak value, kurtosis, skewness, and root mean square. A total of 860 raw segments are collected from the source domain, and after segmentation with overlapping sliding windows (window length 4096, 50% overlap), 1260 training samples are obtained. PC-cGAN generates 500 samples for each of the 4 fault types, expanding the source domain training set to 3260 samples. The target domain retains its original distribution.

4.1.2. Evaluation Metrics

Accuracy, macro-average F1 score, confusion matrix, and Physical Consistency Index (PCI) are used as evaluation metrics. PCI is defined as $1 - \tilde{\epsilon}_{fit}$, where $\tilde{\epsilon}_{fit}$ is the normalized fitting error.

4.2. Horizontal Comparison Experiment

4.2.1. Overall Performance Comparison

The comparative experiment consists of three groups. *Group A* compares the proposed method with 13 baseline methods, including traditional machine learning, deep learning, few-shot learning, domain adaptation, and multi-source fusion approaches. *Group B* is a module ablation study: PC-cGAN is consistently used for data augmentation, and each module is removed sequentially to quantify its contribution. *Group C* evaluates the effectiveness of the complete framework against the strongest baselines. The performance of each method under 5% labeled target domain samples is shown in Table 4.

Table 4. *Cross-domain diagnostic performance with 5% labeled target samples.*

Method Category	Specific methods	Accuracy/%	Macro-F1/%
Traditional Machine Learning	SVM-RBF	62.3±3.2	58.7±4.1
	Random Forest	65.8±2.9	61.2±3.8
Deep Learning	1D-CNN	68.4±3.5	64.5±4.2
	ResNet-18	71.2±3.1	67.8±3.6
	ViT-Tiny	69.5±4.2	65.3±5.1

Method Category	Specific methods	Accuracy/%	Macro-F1/%
Few-shot Learning	MAML	66.7±4.8	62.1±5.3
	Prototypical Nets	64.2±5.1	59.8±5.6
Domain Adaptation Methods	DANN	74.5±3.6	70.2±4.3
	CDAN	76.3±3.4	72.5±4.0
	MMD-based	72.8±3.9	68.9±4.5
Multi-source Fusion Method	Early Fusion	70.5±3.8	66.4±4.4
	Cross-Attention	75.2±3.3	71.8±3.9
Physical Constraint Method	Physics-informed CNN	73.6±3.5	70.8±4.1
Ablation Study (Group B)	MSAF-Net only (original data)	65.3±3.5	61.8±4.1
	MSAF-Net + PC-cGAN	78.2±2.8	75.4±3.2
	MSAF-Net + WDA	75.6±3.1	72.3±3.6
Proposed Method (Group C)	Complete Method (PC-cGAN+MSAF-Net+WDA)	*89.7±2.1*	*87.3±2.5*

The complete method achieves an accuracy of 89.7% at a 5% labeling rate, which is a 13.4 percentage point improvement over the best baseline CDAN (76.3%). Ablation experiments show that introducing PC-cGAN increases accuracy from 65.3% to 78.2% (+12.9%), and further to 89.7% (+11.5%) after adding WDA, demonstrating significant contributions from each module with superadditive synergy.

4.2.2. Performance Under Different Proportions of Labeled Samples

Figure 2 illustrates the performance evolution of various methods under different labeling ratios. The proposed method maintains an accuracy of 68.4% even with only 1% labeled samples, outperforming MAML (52.1%) and DANN (58.7%). When the labeling ratio reaches 10% or higher, the accuracy exceeds 85% and begins to converge. Traditional methods are highly sensitive to the amount of labeled data (improving by about 20% from 1% to 50%), whereas the proposed method shows only a 15% improvement, confirming its weak dependence on labeled samples.

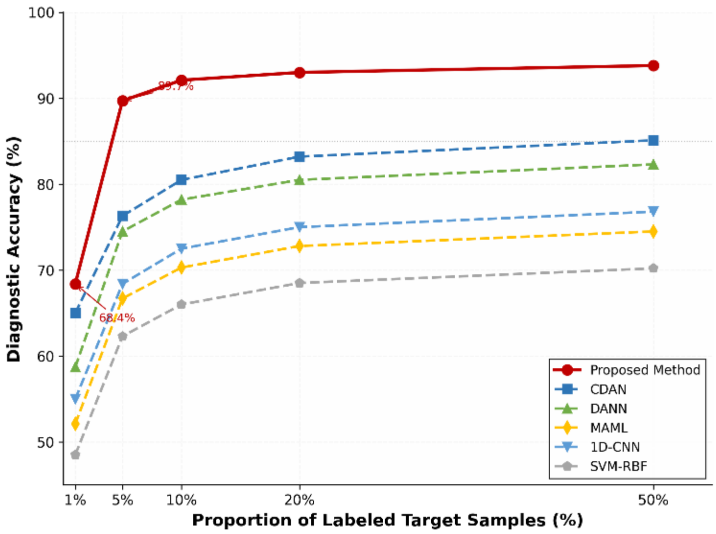

Figure 2. *Cross-domain diagnostic performance under different target-domain labeling ratios.*

4.3. Longitudinal Ablation Experiment

4.3.1. Quantification of Module Contribution

The control variable method is used to quantify the contribution of each module, and the absolute performance drop is defined as the accuracy decrease relative to the complete method. The results are shown in Table 5 and Figure 3(a).

Table 5. Ablation Configuration and Results.

Ablation Configuration	Module Composition	Accuracy/%	Performance Degradation/%	Relative Contribution/%
Complete Method	PC-cGAN+MSAF-Net+WDA	89.7±2.1	—	—
No WDA	PC-cGAN+MSAF-Net	75.6±3.1	14.1	57.8
No MSAF	PC-cGAN+WDA (Single Vibration)	76.2±3.0	13.5	55.3
No PC-cGAN	MSAF-Net+WDA (Original Data)	74.3±3.5	15.4	63.1
Baseline only	MSAF-Net (Original Data, No Domain Adaptation)	65.3±3.5	24.4	Reference group

Note: Relative contribution is defined as the ratio of individual performance degradation to the total improvement (24.4%). The sum exceeds 100% due to super-additive synergy among modules.

The WDA module contributes 57.8%, PC-cGAN contributes 63.1%, and MSAF-Net contributes 55.3% in relative terms. The sum of these relative contributions significantly exceeds 100%, indicating a strong super-additive synergy between the modules.

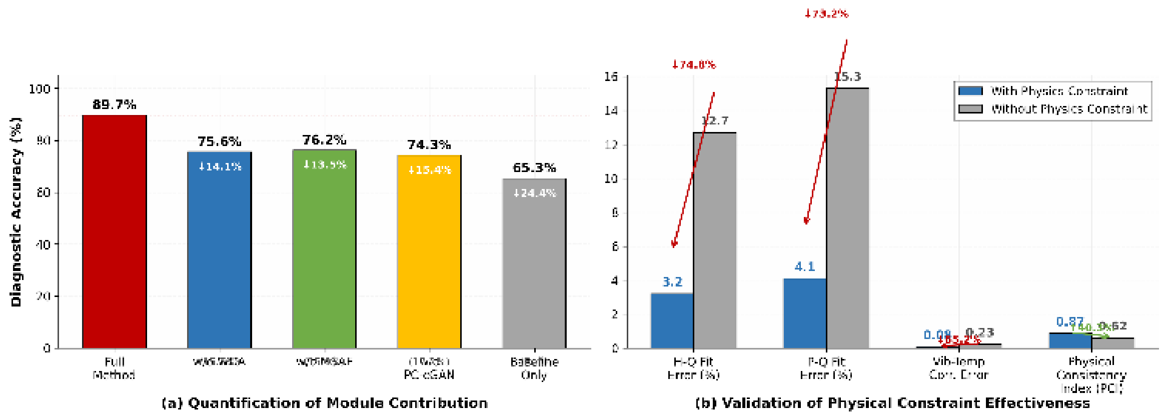

Figure 3. Ablation experiment and validation of physical constraint effectiveness.

4.3.2. Verification of Physical Constraint Effectiveness

Table 6 compares the physical consistency metrics of generated samples with and without physical constraints. After embedding constraints, the H-Q curve fitting error decreased from 12.7% to 3.2% (a reduction of 74.8%), the P-Q curve fitting error dropped from 15.3% to 4.1% (a reduction of 73.2%), and the PCI improved from 0.62 to 0.87.

Table 6. *Comparison of physical consistency of generated samples.*

Evaluation Metrics	With Physical Constraints	Without Physical Constraints	Relative Change
H-Q curve fitting error/%	3.2±0.8	12.7±2.3	−74.8%
P-Q curve fitting error/%	4.1±1.1	15.3±3.1	−73.2%
Vibration-temperature correlation error	0.08±0.02	0.23±0.05	−65.2%
Physical Consistency Index (PCI)	0.87	0.62	+40.3%

To further verify the physical credibility of the generated samples, this study inputs them into a one-dimensional pump simulation model built on AMESim (with the same rotational speed and outlet valve opening set) for physics-based validation. The results show that the average relative error between the head and flow parameters derived from the generated samples and those output by the simulation model is 5.7%. Moreover, none of the samples exhibit characteristic combinations that violate the positive-slope instability criterion, such as "head increasing with flow rate," confirming that the proposed physical constraints effectively eliminate unphysical samples and ensure consistency between the generated data and real fault mechanisms.

4.3.3. Hyperparameter Sensitivity Analysis

When $\lambda_{phy}^{(G)} \in [0.3, 0.7]$, PCI > 0.80; when $\lambda_{dom} \in [0.3, 0.5]$, the best time-domain adaptation performance is achieved. Exceeding 0.6 leads to classification performance degradation. $\lambda_{phy}^{(G)} = 0.5, \lambda_{dom} = 0.5$ are recommended.

4.4. Interpretability Analysis

The global SHAP (SHapley Additive exPlanations) [27] feature importance ranking indicates that the model's attention to bearing fault characteristic frequencies (BPFO, BPFI) and their sidebands accounts for 47.2%, significantly higher than other time-domain statistical features. Furthermore, a pointwise regression of the model's SHAP values on 500 test samples against the corresponding physical parameters (temperature gradient, vibration severity) reveals a strong linear positive correlation (Pearson $r = 0.82, p < 0.001$), validating the model's effective capture of the thermal-mechanical coupling failure mechanism at the decision logic level.

5. Conclusion

This paper addresses the challenge of small-sample cross-domain fault diagnosis for main circulating pumps in valve cooling systems. A transfer learning framework integrating physical priors with data-driven methods is proposed. Theoretically, a quantitative relationship between the Wasserstein distance and the HΔH divergence is established, providing gradient stability guarantees for domain adaptation under non-overlapping distributions. Methodologically, a dual-channel synergistic mechanism combining PC-cGAN and MSAF-Net is constructed, propagating Euler equation constraints from the data generation side to the feature learning side. In application, under the stringent condition where only 5% of target domain samples are labeled, the cross-domain diagnosis accuracy reaches 89.7%, and the model's attention mechanism aligns closely with physical failure modes (Pearson *r=0.82*). This framework offers a feasible technical pathway for small-sample cross-domain maintenance of power equipment. Limitations. Despite the promising diagnostic performance, several limitations of the proposed approach should be acknowledged. First, the physical constraints derived from the Euler equation serve as weak regularization terms; although they effectively eliminate unphysical generated samples, they may not fully capture complex fault-induced hydraulic deviations under extreme operating conditions (e.g., severe cavitation or multi-fault coupling). Second, the proposed framework relies on accurate geometric parameters of the impeller, including the outlet radius r_2, outlet width b_2, and blade outlet angle β_2; measurement uncertainties in these parameters may propagate into the physical consistency loss and degrade the quality of generated samples. Third, the experimental validation is conducted on 69 pumps across five UHV converter stations; while this represents a real-world scenario, the dataset scale is still limited, and the generalization to other pump types or stations with significantly different configurations requires further investigation. Fourth, the computational complexity of the dual-channel architecture (PC-cGAN + MSAF-Net + WDA) is relatively high, which may pose challenges for real-time deployment on resource-constrained edge devices. Finally, the current framework assumes a closed-set scenario where the fault types in the target domain are known a priori; its robustness against unknown fault categories (i.e., open-set domain adaptation) remains to be explored in future work. Future work will focus on introducing fluid-structure interaction multi-physics constraints and deploying lightweight models on edge devices for real-time valve cooling system monitoring.

Abbreviations

BPFI	Ball Pass Frequency of Inner Race
BPFO	Ball Pass Frequency of Outer Race
CDAN	Conditional Domain Adversarial Network
CNN	Convolutional Neural Network
DANN	Domain Adversarial Neural Network
GRL	Gradient Reversal Layer
H Q	Head Flow Curve
MAML	Model Agnostic Meta-Learning
MMD	Maximum Mean Discrepancy
MSAF Net	Multi Source Signal Attention Fusion Network
PC cGAN	Physics Constrained Conditional Generative Adversarial Network
PCI	Physical Consistency Index

Author Contributions

Gao Lei: Conceptualization, Methodology, Software, Writing – Original Draft, Visualization

Wang Wenchao: Data curation, Formal Analysis, Investigation, Resources

Wu Heng: Software, Validation, Resources, Visualization

Zhang Yu: Conceptualization, Supervision, Writing – review & editing, Project Administration, Funding Acquisition

He Biao: Validation, Formal Analysis, Investigation, Data Curation

Funding

This work was supported by the Science and Technology Project of State Grid Shaanxi Electric Power Co., Ltd.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Zhang G H, Ma G P. Analysis of main pump switching logic in valve cooling system under voltage faults [J]. *Power System Technology*, 2021, 45(1): 189-197.

[2] Yu J Y, Qiao Z Q, Wang Y J, et al. Research on fault diagnosis technology of converter valve cooling system based on CVA [J]. *High Voltage Apparatus*, 2023, 59(11): 32-39.

[3] Wang L Y, Yang Y Q, Shi L, et al. Parameter prediction of converter station valve cooling system based on long short-term memory method [J]. *Science, Technology and Engineering*, 2022, 22(2): 547-555.

[4] Wei Y, Li X W, Zhang Y L, et al. Research on fault location method for converter station valve area based on metric learning and knowledge reasoning [J]. *Electrical Measurement and Instrumentation*, 2025, 62(6): 134-142.

[5] Hu J W. Research on intelligent diagnosis algorithms for rotating machinery equipment in small-sample cross-domain complex scenarios [D]. *China Wuhan: Wuhan University of Science and Technology*, 2024.

[6] Wang S, Shuai H Q, Hu H H, et al. Few-shot fault diagnosis of axial piston pump based on prior knowledge-embedded meta learning vision transformer under variable operating conditions [J]. *Expert Systems with Applications*, 2025, 269: 126452.

[7] Jiang C X, Chen H, Xu Q F, et al. Few-shot fault diagnosis of rotating machinery with two-branch prototypical networks [J]. *Journal of Intelligent Manufacturing*, 2023, 34(4): 1667-1681.

[8] Liu S, Li J, Zhou N, et al. Intelligent fault diagnosis of rotating machine via expansive dual-attention fusion transformer enhanced by semi-supervised learning [J]. *Expert Systems with Applications*, 2025, 260: 125398.

[9] Ye M, Yan X, Hua X, et al. MRCFN: A multi-sensor residual convolutional fusion network for intelligent fault diagnosis of bearings in noisy and small sample scenarios [J]. *Expert Systems with Applications*, 2025, 259: 125214.

[10] Zhang S, Ma X, Nie Z, et al. Integration of multi-relational graph oriented fault diagnosis method for nuclear power circulating water pumps [J]. *Measurement*, 2025, 242: 115811.

[11] Zhao C, Zio E, Shen W. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study [J]. *Reliability Engineering and System Safety*, 2024, 245: 109964.

[12] Liang P, Wang X, Ai C, et al. SRSGCN: A novel multi-sensor fault diagnosis method for hydraulic axial piston pump with limited data [J]. *Reliability Engineering and System Safety*, 2025, 253: 110563.

[13] Zhou K, Lu N, Jiang B, et al. FEV-Swin: Multi-source heterogeneous information fusion under a variant Swin Transformer framework for intelligent cross-domain fault diagnosis [J]. *Knowledge-Based Systems*, 2025, 310: 112982.

[14] Ren L, Mo T, Cheng X. Meta-learning based domain generalization framework for fault diagnosis with gradient aligning and semantic matching [J]. *IEEE Transactions on Industrial Informatics*, 2024, 20: 754-764.

[15] Jia L, Chow T W S, Yuan Y. Causal disentanglement domain generalization for time-series signal fault diagnosis [J]. *Neural Networks*, 2024, 172: 106099.

[16] Su Z, Lian P, Shang P, et al. Semi-supervised source-free domain adaptation method via diffusive label propagation for rotating machinery fault diagnosis [J]. *Reliability Engineering and System Safety*, 2024, 252: 110408.

[17] Lin T, Ren Z, Zhu L, et al. A systematic review of multi-sensor information fusion for equipment fault diagnosis [J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 3507848.

[18] Jiang X, Li X, Wang Q, et al. Multi-sensor data fusion-enabled semi-supervised optimal temperature-guided PCL framework for machinery fault diagnosis [J]. I*nformation Fusion*, 2024, 101: 102005.

[19] Yan X, Yan W J, Xu Y, et al. Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network [J]. *Mechanical Systems and Signal Processing*, 2023, 202: 110664.

[20] Huo Z, Martínez-García M, Zhang Y, et al. A multisensor information fusion method for high-reliability fault diagnosis of rotating machinery [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-12.

[21] Zhong J, Zheng Y, Ruan C, et al. M-IPISincNet: An explainable multi-source physics-informed neural network based on improved SincNet for rolling bearings fault diagnosis [J]. *Information Fusion*, 2025, 115: 102761.

[22] Sun D, Li Y, Jia S, et al. Physical knowledge-driven feature fusion and reconstruction network for fault diagnosis with incomplete multisource data [J]. *Mechanical Systems and Signal Processing*, 2025, 225: 112222.

[23] Wang Z, Zhou Z, Xu W, et al. Physics informed neural networks for fault severity identification of axial piston pumps [J]. *Journal of Manufacturing Systems*, 2023, 71: 421-437.

[24] Yang J F, Zhang N, He Y L, et al. Novel dual-network autoencoder based adversarial domain adaptation with wasserstein divergence for fault diagnosis of unlabeled data [J]. *Expert Systems with Applications*, 2024, 238: 122393.

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, Sydney, Australia, 2017, pp. 214-223.

[26] Ben-David S, Blitzer J, Crammer K, et al. A theory of learning from different domains [J]. *Machine Learning*, 2010, 79: 151-175.

[27] Lundberg S M, Lee S I. A unified approach to interpreting model predictions [C]//*Advances in Neural Information Processing Systems*, 2017: 4765-4774.

[28] Zhou S, Qin L, Yang Y, Wei Z, Wang J, Wang J, Ruan J, Tang X, Wang X, Liu K. A novel ensemble fault diagnosis model for main circulation pumps of converter valves in VSC-HVDC transmission systems [J]. Sensors, 2023, 23(11): 5082.

[29] Mei F, Zhang X G, Li J W, Lu J H, Feng T T. A lightweight fault diagnosis model for the main circulating pump of an ultra high voltage converter valve based on a multidimensional vibration feature graph [J]. Power System Protection and Control, 2024, 52(16): 83-96.

[30] Zhao Z M, Gao H D, Teng M F, Zhang Q Y, Liu Y. Fault cause analysis and hidden danger management of main circulating pump for valve cooling system in a flexible DC converter station [J]. Hunan Electric Power, 2025, 45(3): 89-94.