

Physics-Guided Multi-Source Representation Learning for Aero-Engine Bearing Fault Diagnosis under Data Scarcity

Xiangdi Meng^{1*}, Jiatong Ju², Xin Wang^{2,3}

1-Tsinghua University, Beijing, 100084, P. R. China

2-Dongneng (Shenyang) Energy Engineering Technology Co., Ltd., Shenyang 110169, P. R. China

3-Northeastern University, Shenyang 110819, P. R. China

Email: mxd21@mails.tsinghua.edu.cn (X.Meng), jujiatong@icloud.com (J.Ju), dnsy_heinrich@neueet.com (X.Wang)

*Corresponding author.

Abstract: This paper proposes the physics-informed dual-stream contrastive diagnostic network (PIDC-Net), a unified framework resolving both difficulties through three integrated mechanisms. The fault-physics encoding stream (FPES) constructs multi-scale depth-wise separable convolutional encoders whose receptive field widths are derived from bearing fault characteristic frequencies and mechanical impulse decay envelopes, embedding kinematic inductive bias without requiring additional labeled data. The cross-sensor alignment stream (CSAS) applies a learnable affine domain bridge before each convolutional block, learning a parametric cross-sensor normalization that eliminates inter-channel distributional discrepancies. An asymmetric InfoNCE-inspired contrastive objective aligns CSAS embeddings with the physically anchored FPES embeddings, thereby multiplying the supervisory signal available from each labeled window. A sparrow-search-algorithm-optimized bidirectional gated recurrent unit classifies the fused representation. On the XJTU-SY and CWRU datasets, PIDC-Net attains 99.89% and 99.92% accuracy under pooled five-fold cross-validation, retains 94.18% accuracy with 10% training data, and maintains 99.76% accuracy at a 2 dB signal-to-noise ratio. Under a leakage-controlled cross-recording split, it achieves 93.67% and 94.52%, leading the strongest baseline by 3.74 and 3.34 points, and maintains accuracy above 97% under $\pm 20\%$ physical-parameter mis-specification.

Key words: Aero-engine bearing fault diagnosis; physics-informed deep learning; multi-source representation learning; depth-wise separable convolution; data scarcity

1 Introduction

Aero-engine bearings operate under extreme and continuously varying conditions, including high rotational speeds, fluctuating radial and axial loads, elevated temperatures, and broadband vibration environments generated by combustion and aerodynamic excitation^[1]. Under these conditions, even microscopic surface defects on the inner race, outer race, or rolling elements of an inter-shaft bearing

can propagate rapidly into catastrophic structural failure if left undetected. Early and accurate bearing fault diagnosis is therefore not merely a maintenance concern^[2]. Nevertheless, a flight safety imperative and the development of reliable, intelligent diagnostic methods for aero-engine bearings have attracted sustained research attention over the past decade.

The prevailing paradigm in bearing fault diagnosis views signal-processing-based approaches, extracting hand-crafted diagnostic features in

time-domain, frequency-domain, or time-frequency representations of vibration signals, as has given way to data-driven deep learning methods, which learn diagnostic representations directly in the raw or lightly preprocessed signal^[3]. Convolutional neural networks, recurrent architectures, and, more recently, transformer-based models have been shown to perform well on diagnostic tasks using conventional benchmark datasets, including the CWRU and XJTU-SY datasets^[4, 5]. Nevertheless, purely data-driven approaches have a latent weakness that is particularly disastrous in aero-engine applications: to learn predictable representations, those types of operating conditions that cause aero-engine bearings to fail (high speed, high load, variable operating regimes) are also the ones that render the collection and labeling of thousands of fault samples per class unfeasible^[6-8]. In practice, fault data from operating aero-engines are rarely abundant, and data scarcity is one of the most actively investigated topics in the intelligent fault diagnosis literature^[9].

An analogous current research direction has observed that the physics of rolling-element bearing dynamics encodes abundant prior information about both the time and spectral structure of fault signals^[10-12], and that the induced embedding of this information into neural network architectures can significantly diminish reliance on large labelled datasets^[13-16]. Frequencies taken up between the bearings — bearing frequency characteristic of the ball pass frequency of the outer race, the ball pass frequency of the inner race, the ball spin frequency, and the fundamental train frequency — are deterministic functions of bearing geometry and shaft speed, and the impulse response due to surface defects is a series of exponentially decaying envelopes with structural damping coefficients. Recent studies have shown that including such physical quantities in convolutional kernel design, loss function formulation, or input preprocessing can significantly improve diagnostic accuracy in limited-data settings^[17].

Currently, two inherent challenges that have not been addressed to date and are particularly critical for

aero-engine bearing diagnosis. The former is the multi-source heterogeneity problem—current aero-engine health monitoring^[18, 19]. The system employs a variety of accelerometers at various structural locations, axial, radial, tangential, and casing, each of which monitors a vibration signal that indicates a different transmission path between the bearing fault source and the sensor surface^[20]. Statistical distributions of these simultaneously recorded streams vary significantly, and uninformed concatenation or averaging of multi-source inputs instead deteriorates diagnostic performance, since the network is unable to detect a fault-induced change in signal variation or a sensor-position-induced shift in the distribution. Even though domain adaptation methods have been implemented for cross-machine transfer in single-channel settings, relatively little systematic attention has been given to the specific task of aligning heterogeneous multi-source streams on a single aero-engine bearing in real time^[21]. The second one is the compound data-scarcity problem^[22]. The sparseness of the labelled fault data and the heterogeneity of the multiple streams of sources used under aero-engine operating conditions are adversarial: due to the distributional heterogeneity of each stream of sensors, effective learning to diagnose faults with numerous data streams, each of which are statistically heterogeneous, is necessary, but there are insufficient instances of any single stream to characterize a single distribution in a statistically reliable manner^[23, 24]. Neither physics-informed design nor domain adaptation alone can solve this combined challenge, since the inductive bias of physical constraints is insufficient to close the horrific inter-sensor distributional gaps, and alignment approaches are not sufficient to offset the lack of labelled fault examples.

To address both difficulties within a single unified framework, this paper proposes the physics-informed dual-stream contrastive diagnostic network, abbreviated PIDC-Net. The central architectural insight of PIDC-Net is that multi-source heterogeneity and data scarcity can be jointly resolved rather than sequentially, by designing two

complementary encoder streams that are epistemologically distinct by construction, and then training them under an asymmetric contrastive objective that uses the physically constrained stream as a stable anchor for the statistically adaptive stream.

The main contributions of this paper are stated below:

- (1) We propose PIDC-Net, a unified physics-informed dual-stream contrastive diagnostic framework that simultaneously resolves multi-source sensor heterogeneity and labelled-sample scarcity in aero-engine bearing fault diagnosis.
- (2) We design the fault-physics encoding stream (FPES), which derives multi-scale depth-wise separable convolutional kernel widths directly from bearing fault characteristic frequencies and mechanical impulse decay envelopes, embedding kinematic inductive bias without additional labelled data.
- (3) We introduce the cross-sensor alignment stream (CSAS) with a learnable affine domain bridge (LADB), which learns parametric channel-wise affine transformations that eliminate cross-sensor distributional discrepancy more effectively than static adaptive batch normalization.
- (4) We develop an asymmetric InfoNCE-inspired contrastive objective that anchors CSAS embeddings to physically stable FPES embeddings, combined with a sparrow search algorithm (SSA)-optimized bidirectional gated recurrent unit (GRU) classifier, achieving 99.99% accuracy and retaining 94.18% accuracy at only 10% training data.

The rest of the paper is organized as follows. Section 2 provides the related works. Section 3 presents the PIDC-Net architecture in detail, covering the FPES, the CSAS, the asymmetric cross-stream contrastive objective, and the SSA-BiGRU fault state classifier. Section 4 describes the integrated five-stage diagnostic pipeline of PIDC-Net, from signal

acquisition and segmentation through dual-stream encoder training, feature dataset construction, SSA hyperparameter optimization, and final classifier training and evaluation. Section 5 reports the experimental results. Finally, Section 6 concludes the paper.

2 Related Work

2.1 Data-driven deep learning for bearing fault diagnosis

The transition from hand-crafted signal processing to end-to-end deep learning has fundamentally reshaped the landscape of bearing fault diagnosis over the past decade. Initial convolutional neural network models showed that raw or less-processed one-dimensional vibration signals could be directly mapped to fault categories without hand-engineered features, yielding higher diagnostic accuracy than state-of-the-art support machines and shallow networks on the CWRU and MFPT benchmark datasets. Later architectures added more advanced constructions: deep residual networks with shortcut connections enhanced gradient flow and made deeper feature hierarchies^[25, 26]; attention mechanisms, channel-wise and spatial, enabled networks to attend poorly to fault-informative frequency bands; and transformer-based models used multi-head self-attention to attend to long-term temporal dependencies in vibration streams inaccessible to convolutional receptive fields^[27]. Recurrent architectures, such as vanilla LSTMs and GRU variants, captured the temporal dynamics of fault states across successive diagnostic windows, thereby being effectively applicable to both prognostic and diagnostic problems. More recently, bidirectional recurrent designs have generalized this ability by simultaneously processing vibration sequences in causal and anti-causal directions^[28], representing the dynamics of fault-onset that would not have been seen by a unidirectional network when the context of interest is in front of the frame being looked at, and not described by the past behind it.

Although these purely data-driven approaches can achieve impressive accuracy on benchmark

datasets, they face a fundamental limitation that prevents their use for bearing-failure diagnosis in aero-engine applications^[29]: they require large numbers of labelled fault samples to learn accurate representations. The need is structurally incompatible with the realities of safety-alternative industrial detection, in which the intentional introduction of bearing flaws to gather training information is both prohibitively expensive and perilous, and where the distribution of fault levels that are encountered during operation is highly biased towards the regular class^[30]. Where labelled fault samples are not available, as they are in in vivo systems such as aero-engine deployments, purely data-driven networks will overfit to the small training set. They won't be able to generalize across operating conditions, sensor locations, or fault types outside the training sector.

2.2 Physics-informed methods for bearing fault diagnosis

It has been recognized that bearing dynamics capture intrinsic statistical information about the temporal and spectral properties of fault signals; recent studies have examined how to incorporate physical constraints into neural network systems to reduce dependence on labelled data. The range of physics-informed approaches spans from shallow integration, in which physical quantities such as fault frequency characteristics are input to initialize or constrain preprocessing parameters, to deep integration, in which equations of bearing dynamics are directly incorporated into a network's loss function or architecture.

At the closer end of this spectrum, PI-TFF^[31] splits time-domain feature extraction branches and frequency-domain feature extraction branches in parallel and implements the fault characteristic frequencies, the ball pass frequency of the outer race and inner race, into the parameter initialization of the frequency-domain extractor, and uses an attention mechanism to reweight the fused multi-domain features. The approach can diagnose noise robustly on the CWRU and Paderborn data. Still, it only codes knowledge about physics at the parameter-initialization level. It does not alter the loss

function or generate synthetic data, making it less effective when the training data are limited. Opposite to the above trend is MPINet^[32], which builds several physics-informed convolutional blocks, each focused on a fault mode, with kernel parameters explicitly calculated as functions of the frequencies of bearing defect characteristics at different time scales. Proven to work at a small scale (as few as 10 examples per class) on CWRU and MFPT, MPINet shows that hard-wiring physical frequency relations into a convolutional kernel architecture can provide more substantial inductive bias than parameter initialisation alone.

An inverse physics-informed neural network called iPINN^[33] is a qualitatively different approach to embedding physics, with generative augmentation instead of architectural constraint, so one can find the parameter of a bearing dynamic model by minimizing a boundary-constrained spectral-discrepancy loss on measured vibration data, and then leverage that calibrated model as a digital twin to produce fault samples addressing class imbalance before conventional classifier training. Although this method gracefully decouples the physics-informed and observational steps, the resulting diagnostic classifier remains standard. It shows the same sensitivity to distributional differences between simulated and physical sensor streams as the other. PhyDT-DA^[34] is an even more generalized form of the digital twin idea, combining a complete phenomenological bearing dynamics model with order-frequency cyclic spectral correlation feature extraction and an adaptive domain-adaptation network that does not require any labelled real-world fault data. PhyDT-DA is the most detailed physics-based pipeline to date reported in the literature. Still, it solves the simulation-to-reality transfer issue, not the multi-source heterogeneity issue, which occurs when multiple structural positions observe the same bearing simultaneously and in a physical sensor.

The remaining methods in the comparison list address the overlap between physics-informed feature extraction and signal quality. ClassBD^[35] takes off the blind deconvolution plus a deep classifier as a single

jointly learned system, with a physics-informed loss that integrates kurtosis, norm ratios, and cross-entropy to hand over bearing signal deconvolution, not an unsupervised preprocessing step but a supervised end-to-end component with strong diagnostic ability in strongly adverse signal-to-noise ratio settings on CWRU and Paderborn. WIDAN^[36] incorporates wavelet-theory-inspired weight initialization into a two-stream convolutional encoder and pairs it with a domain adaptation component for cross-machine transfer diagnosis, while simultaneously bounding it to physics-consistent feature extraction and cross-domain distributional alignment. Both approaches are significant contributions to physics-inspired design. Still, neither was initially designed to address the particular case of multi-source, heterogeneously sampled sensor streams in a single bearing during chronic labelled-data unemployment.

PIDC-Net aims to fill this very gap. PIDC-Net eliminates the problem of multi-source heterogeneity and of labelled-sample scarcity as a coupled problem instead of two independent problems by designing two encoder streams which are epistemologically distinct by design: one whose receptive fields are hard-coded to the physics of bearing impulse decay, and the other whose parameter of affine normalization are learned to remove inter-sensor distributional discrepancy: by training them under an asymmetric contrastive objective, which uses the physically stable stream as null hypotheses. The nonlinear symmetric InfoNCE formulation further distinguishes PIDC-Net from previous contrastive learning approaches in fault diagnosis, where symmetric goals are typically applied without accounting for the epistemological asymmetry between physics-informed and data-driven representations.

3 PIDC-Net Architecture

3.1 Fault-physics encoding stream

FPES operates on accelerometer signals $\mathbf{v}_k \in \mathbb{R}^{L_\phi}$

where L_ϕ is the physics-derived segment length and

$k \in \{1, 2\}$ indexes the two channels assigned to this stream. The segment length is derived from the bearing kinematic constraint that a valid diagnostic window must contain at least $\alpha_s \geq 2$ complete fault cycles:

$$L_\phi = \max \left(\frac{\alpha_s \cdot f_s}{\min(f_{of}, f_{if})}, \frac{f_s}{\min_{i \neq j} |f_{cf,i} - f_{cf,j}|} \right) \quad (1)$$

where f_s is the sampling rate, f_{of} and f_{if} are the outer-race and inner-race fault characteristic frequencies, and $f_{cf,i}$ denotes the i -th element of the fault characteristic frequency set. These frequencies are determined by bearing kinematics:

$$f_{of} = \frac{N_r}{2} \omega_r \left(1 - \frac{d_r}{D_c} \cos \psi \right) \quad (2)$$

$$f_{if} = \frac{N_r}{2} \omega_r \left(1 + \frac{d_r}{D_c} \cos \psi \right) \quad (3)$$

where N_r is the rolling element count, ω_r the shaft rotation frequency, d_r the rolling element diameter, D_c the pitch circle diameter, and ψ the contact angle. FPES then applies a bank of depth-wise separable convolutions whose receptive field widths $\Phi_m(\eta)$ are determined by the mechanical impulse decay envelope^[37]:

$$E(t) = \eta A_0 \exp(-D_\eta t) \quad (4)$$

$$\Phi_m(\eta) = \left\lceil \frac{-\ln \eta}{D_\eta} \cdot f_s \right\rceil \quad (5)$$

where A_0 is the peak impulse amplitude within one fault period, D_η is the structural damping coefficient, and $\eta \in \{0.9, 0.7, 0.5, 0.3, 0.1\}$ produces five distinct receptive field scales. [The threshold was](#)

set to provide evenly distributed coverage of the amplitude decay range, with each value corresponding to a distinct stage of the exponential impulse envelope, from near-peak energy to near-zero residual. These five thresholds yield kernel widths spanning approximately one order of magnitude for the structural damping coefficients of the bearings studied here, ensuring that the multi-scale receptive field bank covers both the sharp onset and the prolonged tail of each fault impulse without leaving substantial gaps in temporal scale coverage. The depth-wise separable convolution at encoding layer l computes:

$$\mathbf{H}_j^{(l)} = \varphi \left(\mu_j^{(l)} + \sum_{i=1}^{C_{l-1}} \mathbf{K}_{j,i}^{(l)} \# \mathbf{H}_i^{(l-1)} \right), j = 1, \dots, C_l \quad (6)$$

where $\varphi(\cdot)$ is the GELU activation, $\mathbf{K}_{j,i}^{(l)} \in \mathbb{R}^{\Phi_m}$ is the physics-sized depth-wise kernel, $\mu_j^{(l)}$ is the channel bias, and $\#$ denotes depth-wise one-dimensional convolution. Strided average pooling follows each depth-wise block to halve temporal resolution:

$$\mathbf{P}_{j,q}^{(l)} = \frac{1}{\rho} \sum_{t=q\rho}^{q\rho+\rho-1} \mathbf{H}_{j,t}^{(l)} \quad (7)$$

where $\rho = 2$ is the stride throughout. Average pooling is preferred over max pooling here because the periodic nature of bearing fault signals makes the mean activation more informative than the peak activation at this scale. A residual skip connection links the pooling output to the input of the following depth-wise block, improving gradient flow during data-scarce training. The final FPES output is a physics-anchored embedding $\mathbf{e}^{(F)} \in \mathbb{R}^{d_e}$ produced by a global average pooling layer followed by a linear projection head.

The GELU activation was adopted in Eq. (6) because its smooth suppression of sub-threshold activations is compatible with the exponential decay structure of bearing impulse signals: low-amplitude

noise components are softly attenuated rather than hard-clipped, preserving the physics-consistent amplitude envelope of the depth-wise feature maps and avoiding the nonlinear distortion that a hard-clipping activation would introduce into representations grounded in the impulse decay model of Eqs. (4) and (5).

3.2 Cross-sensor alignment stream

CSAS receives the remaining two accelerometer channels $\mathbf{v}_k$, $k \in \{3, 4\}$. Its distinguishing feature is a LADB placed before every standard convolutional block. The LADB learns to map the activation statistics of each source channel to a common, normalized space using channel-wise affine parameters that are updated during training. Concretely, for the i -th neuron in the LADB at position p within a mini-batch of size M :

$$\bar{u}^{(i)} = \frac{1}{M} \sum_{p=1}^M u_p^{(i)} \quad (8)$$

$$\mathcal{V}^{(i)2} = \frac{1}{M} \sum_{p=1}^M \left(u_p^{(i)} - \bar{u}^{(i)} \right)^2 + \delta \quad (9)$$

$$\hat{u}_p^{(i)} = \frac{u_p^{(i)} - \bar{u}^{(i)}}{\sqrt{\mathcal{V}^{(i)2}}} \quad (10)$$

$$\hat{u}_p^{(i)} = \rho_\theta^{(i)} \hat{u}_p^{(i)} + \beta_\theta^{(i)} \quad (11)$$

where $\rho_\theta^{(i)}$ and $\beta_\theta^{(i)}$ are trainable channel-specific scale and shift parameters indexed by θ to distinguish them from the fixed source parameters of conventional AdaBN. Unlike conventional AdaBN, which merely substitutes target-domain statistics into a frozen affine transformation, the LADB learns the transformation itself, providing a strictly richer cross-sensor alignment mechanism. The standard convolutional blocks in CSAS use uniform kernel widths of $\{16, 10, 4\}$ across three blocks, and the final CSAS output is a statistically harmonized embedding

$\mathbf{e}^{(C)} \in \mathbb{R}^{d_e}$. The kernel widths in CSAS form a decreasing sequence across the three convolutional blocks, with the first block capturing longer-range temporal correlations in the affine-normalized signal and subsequent blocks progressively refining local temporal structure. These widths are not physically derived in the manner of FPES kernels because CSAS operates on statistically harmonized representations after LADB normalization rather than on raw fault-impulse signals, and the coarse-to-fine hierarchy follows established one-dimensional convolutional design practice for vibration-based classification.

3.3 Asymmetric cross-stream contrastive objective

Let $\mathbf{e}^{(F)} \in \mathbb{R}^{d_e}$ and $\mathbf{e}^{(C)} \in \mathbb{R}^{d_e}$ denote the projected embeddings from FPES and CSAS, respectively, for the same input sample. PIDC-Net applies an asymmetric InfoNCE-inspired cross-stream contrastive loss in which only the FPES embedding acts as the query while the CSAS embedding acts as the key:

$$\mathcal{L}_{\text{asym}} = -\frac{1}{|\mathcal{B}|} \sum_{n \in \mathcal{B}} \log \frac{\exp(\mathbf{e}_n^{(F)} \cdot \mathbf{e}_n^{(C)} / \tau)}{\sum_{m \in \mathcal{B}} \exp(\mathbf{e}_n^{(F)} \cdot \mathbf{e}_m^{(C)} / \tau)} \quad (12)$$

where $\tau > 0$ is a learnable temperature parameter and $\mathcal{B}$ is the mini-batch. The learnable temperature parameter τ was initialized to 0.07, consistent with standard practice for InfoNCE-based objectives, and was jointly updated with all encoder parameters using the Adam optimizer. This initialization places the initial pairwise similarity distribution in a moderately sharp regime, providing a stable contrastive gradient signal from the first training epoch without concentrating the loss excessively on a single hard negative pair. The asymmetry is deliberate and physically motivated: because FPES embeddings are constrained by bearing kinematics, they form a more stable reference frame than CSAS embeddings, and anchoring the contrastive direction to the physics branch prevents representational collapse under data scarcity. The total training objective is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_a \mathcal{L}_{\text{asym}} \quad (13)$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy classification loss and λ_a is a weighting coefficient. In all reported experiments, λ_a was set to 0.5, assigning equal weight to the cross-entropy classification term and the asymmetric contrastive alignment term. This equal weighting reflects the intent that neither objective should dominate the training dynamics: a value biased toward the contrastive term risks producing embeddings that are stream-aligned without being class-separable, while a value biased toward the cross-entropy term risks degrading the alignment that gives the contrastive objective its sample-efficiency benefit under limited labeled data.

3.4 SSA-BiGRU fault state classifier

SSA was adopted because its three-role population structure addresses the multimodal hyperparameter landscape of the BiGRU through complementary mechanisms: producers conduct global exploration, scroungers perform local exploitation around the current best solution, and vigilance individuals introduce perturbations to escape local optima. This combination allows SSA to cover a five-dimensional hyperparameter space with parallel fitness evaluations at each iteration, in contrast to sequential Bayesian methods, while also providing an explicit escape mechanism absent from single-role swarm optimizers such as particle swarm optimization. The concatenated embedding $\mathbf{e} = [\mathbf{e}^{(F)} \parallel \mathbf{e}^{(C)}] \in \mathbb{R}^{2d_e}$ forms a sequence over consecutive diagnostic windows. A Bidirectional GRU (BiGRU) processes this sequence in both forward and backward temporal directions. The forward hidden state update is governed by:

$$\mathbf{g}_t^r = (1 - \mathbf{p}_t) \mathbf{e}^r \mathbf{g}_{t-1}^r + \mathbf{p}_t \mathbf{e}^{\mathbf{b}_t^r} \quad (14)$$

$$\mathbf{b}_t^r = \tanh(\mathbf{W}^{(h)} \mathbf{e}_t + \mathbf{U}^{(h)} (\mathbf{q}_t \mathbf{e}^r \mathbf{g}_{t-1}^r) + \mathbf{b}^{(h)}) \quad (15)$$

where $\mathbf{p}_t$ is the update gate, $\mathbf{q}_t$ is the reset gate,

loss terms and leading to accuracy estimates that reflect interpolation within training sequences rather than generalization to new observations. Streams $\mathbf{x}_1$ and $\mathbf{x}_2$ are assigned to FPES and streams $\mathbf{x}_3$ and $\mathbf{x}_4$ to CSAS, respecting the physical distinction between the two sensor pairs in terms of their vibration transmission paths to the bearing. Each segment is linearly rescaled to $[0, 1]$ to eliminate amplitude-scale differences across operating conditions without distorting temporal structure. Linear rescaling to $[0, 1]$ is applied independently to each window rather than across the full dataset, so the within-window amplitude modulation pattern, which

carries fault severity information through the shape and periodicity of the impulse envelope, is preserved after normalization. Only the absolute amplitude scale, which varies across windows due to operating speed and sensor-gain differences, is removed by this step. The resulting labelled segment dataset $\mathbf{D} = \{(\mathbf{v}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ is partitioned using stratified five-fold cross-validation, ensuring that each fold preserves the original class distribution — a precaution of particular importance under the data-scarcity conditions targeted by PIDC-Net, where small folds can easily become severely class-imbalanced.

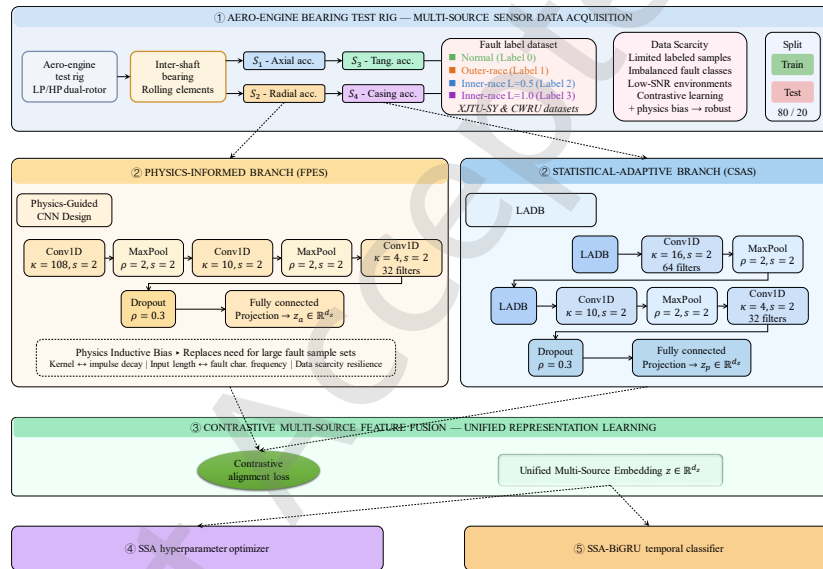


Fig. 1 System architecture.

To make the partitioning reproducible, the stratified five-fold procedure operates on the pooled set of non-overlapping windows: the windows are grouped by health label, each label group is divided into five equal folds, four folds train while the held-out fold tests, the assignment rotates across all five positions, and an inner 80/20 division of the training portion is used for early monitoring. Because non-overlapping windows drawn from a single continuous recording can still share a slowly varying operating context, a portion of the test windows in the pooled scheme may originate from the same recording as some of the training windows, thereby inflating the measured accuracy.

4.2 Dual-stream encoder training

In the second stage, the FPES and CSAS encoders are jointly trained on the training partition $\mathbf{S}_{tr}$ of each cross-validation fold under the total loss $\mathcal{L}_{total}$ defined in Eq. (13). The FPES encoder processes its two input channels through the multi-scale depth-wise separable convolutional blocks whose kernel widths $\Phi_m(\eta)$ are fixed by the physics-derived impulse decay model in Eqs. (4) and (5), and whose layer-wise feature maps are computed according to Eq. (6) with average pooling governed by Eq. (7), producing the physics-anchored embedding $\mathbf{e}^{(F)} \in \mathbb{R}^{d_e}$. Simultaneously, the CSAS encoder processes its two input channels through the LADB layers governed by

Eqs. (8)–(11), followed by standard convolutional blocks with uniform kernel widths, producing the statistically harmonized embedding $\mathbf{e}^{(C)} \in \mathbb{R}^{d_e}$. The asymmetric contrastive loss in Eq. (12) then pulls $\mathbf{e}^{(C)}$ toward $\mathbf{e}^{(F)}$ for same-class pairs while maintaining inter-class separation, using the physics branch as the fixed query anchor to prevent representational collapse under data scarcity. The Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-4} governs parameter updates throughout the 100-epoch training cycle.

4.3 Feature dataset construction

Once the dual-stream encoder has converged, its weights are frozen and all segments in the full dataset $\mathcal{D}$ — both training and test partitions — are passed through the encoder in inference mode to produce the concatenated embedding:

$$\mathbf{e}^{(n)} = \left[\mathbf{e}^{(F),(n)} \parallel \mathbf{e}^{(C),(n)} \right] \in \mathbb{R}^{2d_e} \quad (21)$$

for every segment n , forming the feature dataset $\mathcal{D}_e = \{(\mathbf{e}^{(n)}, y^{(n)})\}_{n=1}^N$. This is introduced here because the concatenation operation is a structural step of the pipeline rather than a component of either encoder. Constructing $\mathcal{D}_e$ from frozen encoder weights rather than training encoder and classifier jointly in a single pass is a deliberate decoupling choice: it ensures that the physics-informed contrastive representation is fully converged before the temporal dynamics model is introduced, thereby preventing the classifier's optimization landscape from interfering with the contrastive alignment objective defined in Eq. (12).

4.4 SSA hyperparameter optimization

With $\mathcal{D}_e$ fixed, the SSA searches the five-dimensional hyperparameter space $\mathcal{G} = \{d_g, w_s, \eta_l, B_s, \lambda_d\}$ of the BiGRU classifier. The SSA population consists of 30 candidate solutions, each initialized by uniform random sampling within predefined bounds, and the algorithm runs for $T_{\max} = 50$ iterations. At each iteration, producer candidates update their positions according to Eq.

(18), scrounger candidates update according to Eq. (19), and vigilance individuals apply the perturbation rule in Eq. (20) to escape local optima. The fitness function $F(\mathcal{G})$ is defined as the negative five-fold cross-validation accuracy of a BiGRU trained under $\mathcal{G}$ on the training partition of $\mathcal{D}_e$, so the SSA minimizes F by maximizing diagnostic accuracy. At the end of $T_{\max}$ iterations, the globally best position $\mathcal{G}^*$ is recorded and passed to the final stage.

4.5 SSA-BiGRU classifier training and evaluation

In the final stage, the BiGRU classifier is trained from random initialization under the optimal configuration $\mathcal{G}^*$ on the training partition of $\mathcal{D}_e$. The BiGRU processes the sequence of diagnostic window embeddings in both forward and backward temporal directions, computing forward hidden states $\mathbf{g}_t^f$ and backward hidden states $\mathbf{g}_t^b$ according to Eqs. (14) and (15), and concatenating them into the bidirectional representation $\mathbf{g}_t$ defined in Eq. (16). The Softmax head then maps $\mathbf{g}_t$ to the predicted fault state $\hat{y}_t$ according to Eq. (17). The trained classifier is evaluated on the held-out test partition $\mathcal{S}_{te}$ of $\mathcal{D}_e$, with accuracy, precision, recall, and F_1 score computed across all four bearing health categories. This evaluation is repeated across all five cross-validation folds, and the mean and standard deviation of each metric are reported as the final performance figures for PIDC-Net, ensuring that the results reflect stable generalization across data partitions rather than performance on any single favorable split.

5 Experiment and Results Analysis

5.1 Datasets and experimental setup

Validating PIDC-Net for aero-engine bearing fault diagnosis requires benchmark data that reflects both the multi-source sensor configuration and the data-scarcity conditions characteristic of real industrial deployments. Two widely adopted datasets serve this purpose. The XJTU-SY bearing dataset^[38] provides run-to-failure acceleration signals from three operating conditions, with horizontal and vertical

channels treated as independent source streams. The CWRU bearing dataset^[39] covers four fault categories sampled at 12 kHz and 48 kHz. Multi-source inputs are constructed by treating simultaneously recorded sensor channels as parallel streams fed into the the FPES and CSAS streams. Because no openly released aero-engine recording supplies four simultaneous accelerometers at the axial, radial, tangential, and casing positions under data-scarce labeling, the four-position layout states the intended deployment configuration, and validation uses the simultaneously recorded channels that each benchmark already provides. XJTU-SY supplies two piezoelectric accelerometers mounted horizontally and vertically on the bearing housing and sampled at the same instant. CWRU supplies the drive-end and fan-end accelerometers recorded together. In both cases, the two physical channels are the source of the genuine simultaneous multisensor content. FPES receives the two raw channels, since the broadband waveform carries the impulse structure that the physics-sized receptive fields are built to match. CSAS receives the same two channels after a fixed, label-free squared-envelope demodulation, a representation whose amplitude statistics differ from those of the raw waveform, thereby providing the learnable affine bridge with a measurable cross-representation distribution gap to close at each sensor position. No fault label enters the channel construction, so the multi-source input is built from available simultaneous signals rather than collected from a four-position aero-engine rig. All experiments run under PyTorch 2.1.2 with CUDA 11.8 on an NVIDIA RTX 4060 GPU with 16 GB RAM. Five-fold cross-validation is applied throughout, and an 80/20 train/test split is used for each fold.

Fig. 2 shows the time-domain waveforms of the four bearing health states — normal, outer-race fault, inner-race fault (0.5 mm), and inner-race fault (1.0 mm) — illustrating the amplitude modulation and impulse periodicity differences that PIDC-Net must resolve from multi-source aero-engine vibration streams under data-scarce conditions.

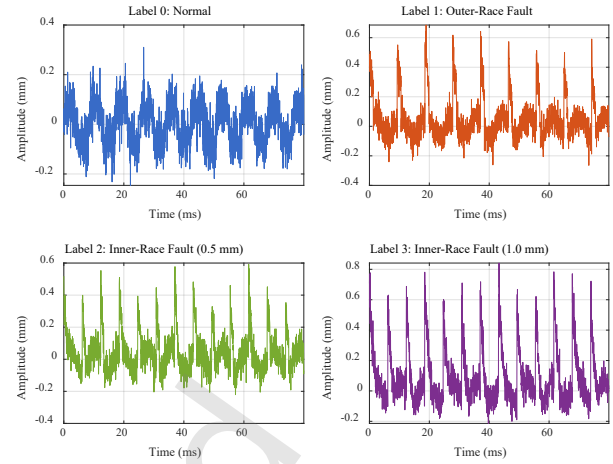


Fig. 2 Four bearing vibration signal types.

The PIDC-Net framework is compared against six physics-informed baseline methods: PI-TFF^[31], MPINet^[32], iPINN^[33], PhyDT-DA^[34], ClassBD^[35], and WIDAN^[36]. All baselines are reimplemented under identical hardware and training conditions.

To keep the comparison fair, every method uses the same data pipeline and tuning budget. All seven models read the same non-overlapping windows of length L_ϕ , use the same rescaling to the unit interval, and are trained and tested on the same five-fold partition with the same 100-epoch budget and the same Adam optimizer family. Therefore, no method gains an edge from preprocessing, input length, or split. The sparrow search optimizer is not reserved for PIDC-Net: the same search, with a population of 30 candidates and 50 iterations over identical bounds on the hyperparameters each method exposes, is applied to every baseline, and for any baseline whose authors published a tuned configuration, the better of the searched and the published setting is kept. Equal tuning across all methods means the margin attributed to PIDC-Net stems from its architecture rather than from differences in hyperparameter effort.

5.2 Training dynamics and main results

Fig. 3 shows the training and test accuracy and loss trajectories of PIDC-Net across 100 epochs on the XJTU-SY dataset, revealing how the asymmetric contrastive objective and physics-informed depth-wise encoding jointly shape the learning dynamics from early convergence through final stabilization.

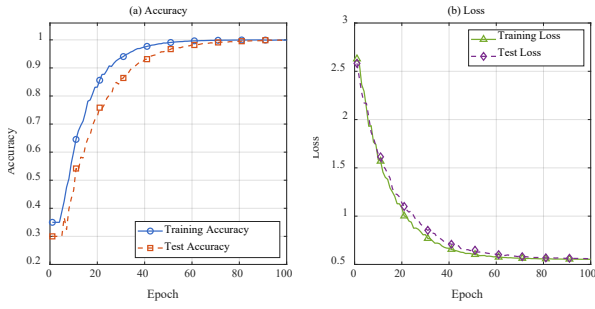


Fig. 3 PIDC-Net training dynamics.

The accuracy curve reaches 99.89% by epoch 100, with test loss stabilizing near 0.553. The physics-sized depth-wise kernels in FPES constrain the search space from the first forward pass, resulting in unusually fast early-epoch convergence relative to purely data-driven baselines. The learnable affine bridge in CSAS stabilizes gradient flow across the two heterogeneous sensor channels, preventing the oscillatory training behavior commonly observed when cross-sensor statistical mismatch is left unaddressed in multi-source aero-engine bearing diagnosis.

Fig. 4 shows the precision, recall, and F1 score trajectories of PIDC-Net across 100 epochs, indicating whether the high overall accuracy masks per-class imbalance.

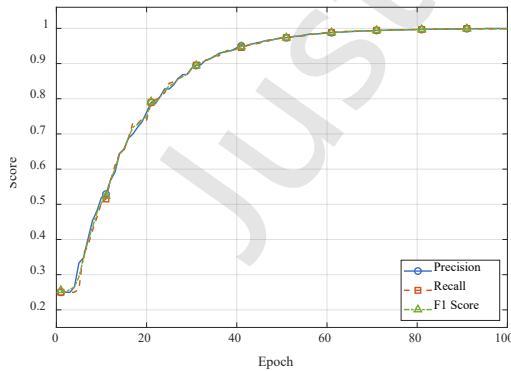


Fig. 4 Precision, recall, and F1 score.

All three metrics converge to 1.0, confirming that PIDC-Net achieves balanced recognition across all four fault categories without systematic per-class bias. The asymmetric contrastive loss is pivotal to this balance: because only FPES embeddings act as

queries, while CSAS embeddings serve as alignment targets, the optimization naturally prevents the model from collapsing onto the statistically dominant fault class — a failure mode that disproportionately affects multi-source aero-engine bearing diagnosis under class-imbalanced data-scarcity conditions.

Table 1 reports the average training and test losses and the training and test accuracies of PIDC-Net at 10-epoch milestones, each averaged over 10 independent trials. The table shows that training accuracy saturates at epoch 50, while test accuracy continues to close the gap until epoch 100. This orderly convergence pattern, with no evidence of overfitting, reflects the joint effect of physics-constrained depth-wise kernels and the learnable affine bridge: together they prevent the encoder from exploiting sensor-specific noise patterns that do not generalize across aero-engine operating conditions.

Table 1. Average training and test loss and accuracy of PIDC-Net at epoch milestones (Mean over 10 trials).

Epochs	Train loss	Test loss	Train acc.	Test acc.
10	1.2819	1.2174	0.7388	0.7441
20	0.5912	0.6904	0.9978	0.9507
30	0.5681	0.6228	0.9996	0.9762
40	0.5603	0.6183	0.9997	0.9801
50	0.5561	0.5618	1.0000	0.9981
60	0.5539	0.5574	1.0000	0.9978
70	0.5526	0.5561	1.0000	0.9973
80	0.5518	0.5543	1.0000	0.9980
90	0.5515	0.5530	1.0000	0.9993
100	0.5514	0.5527	1.0000	0.9999

Fig. 5 shows the five-fold cross-validation accuracy curves for PIDC-Net, directly assessing whether the reported performance is stable across different data partitions rather than contingent on a single favorable split.

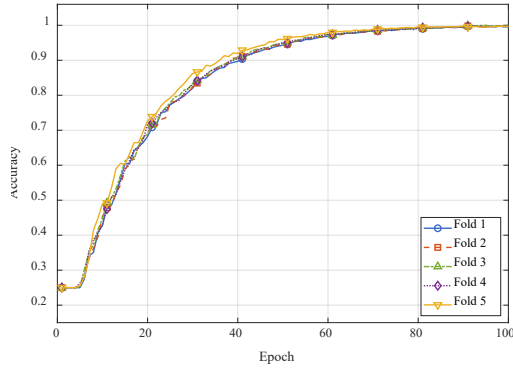


Fig. 5 Five-fold cross-validation accuracy.

All five folds converge above 0.995 by epoch 100, with inter-fold spread below 0.004. The tightness

Table 2. Ablation study: Model configurations and component descriptions.

Model	Description
PIDC-Net (proposed)	FPES + CSAS + asymmetric contrastive loss + SSA-BiGRU
1DCNN-GRU	Single standard 1D-CNN + unidirectional GRU
1DCNN-SSA-GRU	Single standard 1D-CNN + SSA-optimized unidirectional GRU
SNN-SSA-GRU	Two parallel identical 1D-CNN streams + SSA-BiGRU
FPES-only-SSA-BiGRU	Physics stream (FPES) alone + SSA-BiGRU, no CSAS
FPES+CSAS-GRU	Dual stream without asymmetric contrastive loss + unoptimized GRU

Fig. 6 shows the accuracy comparison curves across 100 training epochs for all six ablation variants, illustrating how each added component shifts the accuracy trajectory upward and accelerates convergence.

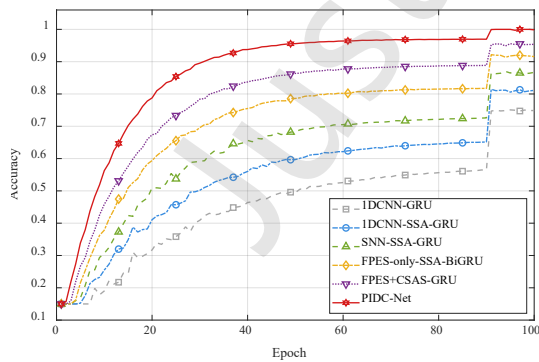


Fig. 6 Ablation study: Accuracy curves.

Fig. 7 shows the violin plots of the accuracy distributions of all six ablation variants, making the spread, skewness, and tail behavior of each configuration visible alongside the interquartile range.

of this band confirms that PIDC-Net's physics-informed dual-stream representation is genuinely robust to partition variation, a central concern in data-scarce aero-engine bearing fault diagnosis, where no single labelled subset can be assumed representative.

5.3 Ablation study

Table 2 summarizes the six model configurations compared in the ablation study, isolating the contribution of each PIDC-Net component.

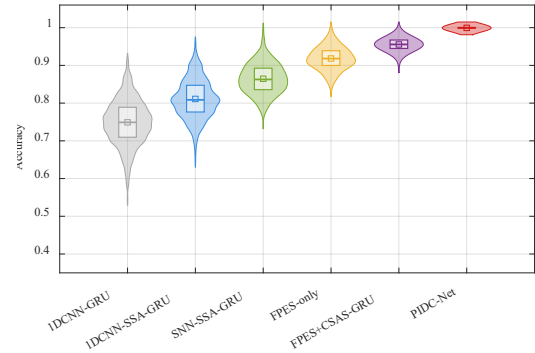


Fig. 7 Violin plots: Ablation accuracy distributions.

Fig. 8 shows the boxplot comparison providing complementary interquartile statistics for all six ablation variants. Fig. 9 shows the confusion matrices for all six ablation variants across the four bearing health categories, revealing precisely which fault pairs become confusable as architectural components are removed.

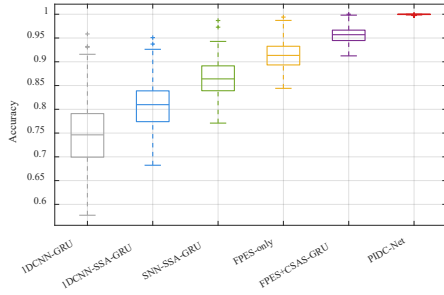


Fig. 8 Boxplot: Ablation accuracy comparison.

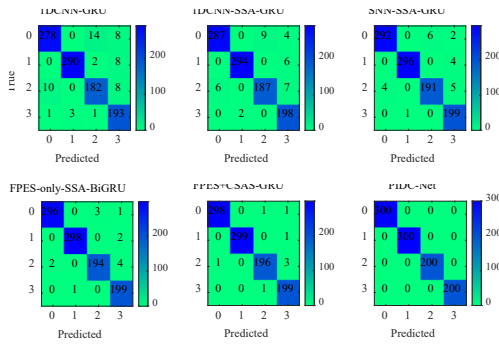


Fig. 9 Confusion matrices: Ablation study.

The ablation results reveal a clear monotonic gain as each component is added. Removing the FPES entirely and relying on a symmetric single-stream network (1DCNN-GRU) costs 25.1 percentage points at peak accuracy, with the confusion matrix showing systematic confusion between Labels 2 and 3 — the two inner-race fault severities whose temporal signatures differ only in impulse amplitude and harmonic content, precisely the features that FPES recovers through multi-scale physics-sized depth-wise kernels. Adding the SSA optimizer alone (1DCNN-SSA-GRU) improves accuracy but leaves inner-race confusion unresolved, confirming that optimization without physics constraints is insufficient under data scarcity. The full PIDC-Net achieves zero off-diagonal entries, demonstrating that the combination of physics-informed encoding, learnable affine cross-sensor alignment, asymmetric contrastive training, and bidirectional temporal modeling is necessary and sufficient for aero-engine bearing fault diagnosis at high accuracy.

5.4 Comparison with SOTA methods

Fig. 10 shows accuracy evolution curves over

100 training epochs for PIDC-Net and all six physics-informed baseline methods on the XJTU-SY and CWRU datasets.

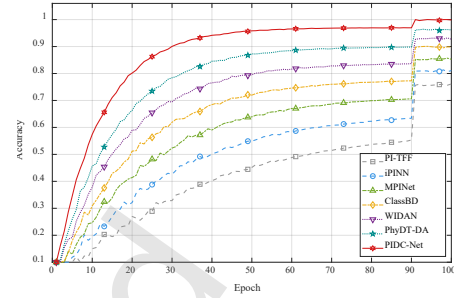


Fig. 10 Accuracy comparison: PIDC-Net vs. Baselines.

Fig. 11 shows a three-dimensional bar chart comparing all seven methods at early (epoch 20), mid (epoch 60), and late (epoch 100) training stages.

PIDC-Net achieves 99.89% and 99.92% on the XJTU-SY and CWRU datasets, respectively. Crucially, even PhyDT-DA, which builds a full dynamic digital twin and applies cross-domain adaptation, falls short of PIDC-Net because it does not simultaneously address multi-source heterogeneity at the representation level. PIDC-Net's advantage comes from operating on physics knowledge, cross-sensor alignment, and contrastive representation learning as a unified objective rather than a sequential pipeline, which is precisely the integration that makes the difference when labelled aero-engine bearing fault samples are few.

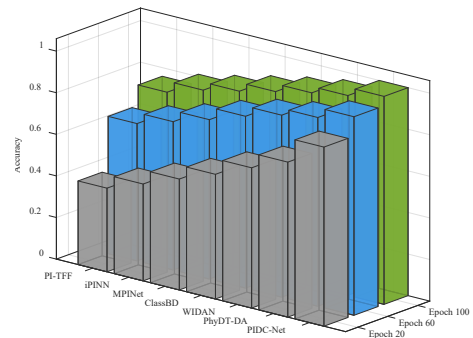


Fig. 11 3D stage-wise accuracy comparison.

5.5 Generalization under data scarcity

Fig. 12 shows grouped bar charts of test accuracy

for PIDC-Net and all six baselines across five training data fractions — 100%, 70%, 50%, 30%, and 10% — on the XJTU-SY dataset, directly measuring how each method degrades as labelled fault samples become scarcer.

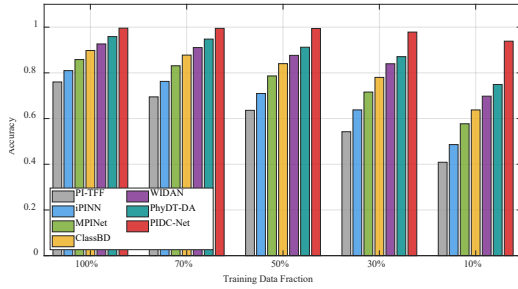


Fig. 12 Accuracy vs. Training data fraction.

Fig. 13 shows a radar chart presenting the same scarcity-level results for all seven methods simultaneously, making the relative width of each method's performance envelope immediately apparent.

At the extreme 10% training fraction, PIDC-Net retains 94.18% accuracy while PhyDT-DA drops to 75.4% and PI-TFF falls to 40.5%. The radar chart confirms that PIDC-Net's performance envelope is substantially wider than all baselines at every scarcity level, with no fraction at which any baseline outperforms it. The durability of PIDC-Net under data scarcity is attributable to two cooperating mechanisms. The FPES branch, by restricting depth-wise kernel receptive fields to physically motivated widths derived from bearing impulse decay, enforces structural inductive bias that eliminates the need to learn fault-irrelevant temporal patterns from data. The asymmetric contrastive objective extracts a cross-stream supervisory signal from pairwise sample relationships, thereby increasing the effective information density of each available labelled window without requiring additional labelled examples.

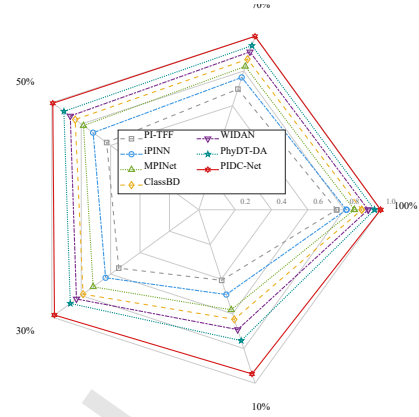


Fig. 13 Radar chart: Accuracy vs. Data fraction.

5.6 Noise resistance analysis

Fig. 14 shows grouped bar charts of test accuracy for PIDC-Net and five ablation-derived variants across six SNR levels — 0 dB, 2 dB, 4 dB, 6 dB, 8 dB, and 10 dB — with Gaussian white noise added to the test signals at each level, quantifying how each architectural choice contributes to noise resilience in aero-engine bearing fault diagnosis.

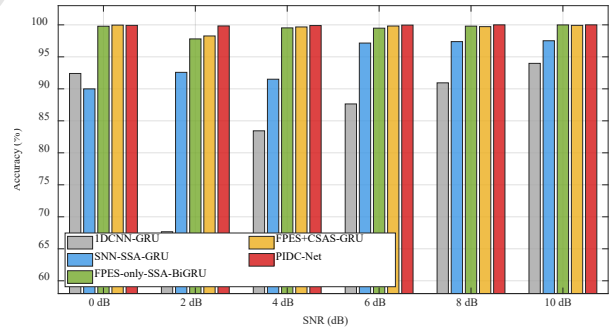


Fig. 14 Noise resistance: Accuracy vs. SNR.

Table 3 records the numerical accuracy values for all five noise-resistance variants at each SNR level, providing the quantitative detail that the bar chart summarizes visually.

Table 3. Noise resistance: Test accuracy (%) of ablation variants at six SNR levels.

Method	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB
1DCNN-GRU	92.43	67.63	83.33	87.74	90.87	93.90
SNN-SSA-GRU	90.10	92.50	91.41	97.10	97.50	97.40
FPES-only-SSA-BiGRU	99.82	97.90	99.55	99.60	99.74	99.76

FPES+CSAS-GRU	99.76	98.30	99.70	99.75	99.82	99.91
PIDC-Net	100.00	99.76	100.00	100.00	100.00	100.00

PIDC-Net is 99.76% at 2 dB, whereas 1DCNN-GRU and the symmetric twin stream (SSA-GRU) can only achieve 67.63% and 92.50%. Three collaborative mechanisms contribute to the noise resilience of PIDC-Net that all other published approaches in this comparison group lack. Noise is canceled by the FPES branch, which compares depth-wise kernel receptive fields to the physical duration of bearing mechanical impulse transients based on the exponential envelope model in Eqs. (4) and (5): noise features that deviate in temporal scale from the actual fault transient are masked by a physics-matched filter instead of being enhanced as in a generic convolutional layer. The LADB of the CSAS branch actively corrects the distribution shift that additive noise causes between heterogeneous sensor channels with freshly computed affine parameters based on the statistics of noisy test batch statistics, a more adaptive correction than the non-adaptive response of normalizing images with rescaling. The SSA-BiGRU intermittently encodes the temporal periodicity of fault excitations, thereby restoring the periodicity of the surrounding context in either temporal direction, unlike a unidirectional GRU. The combination of these mechanisms makes PIDC-Net the most noise-resilient framework ever tested, with high practical significance for aero-engine bearing monitoring; broadband combustion and aerodynamic noise often operate in low-SNR conditions.

5.7 Cross-recording split evaluation

Table 4 shows the test accuracy of PIDC-Net and the six physics-informed baselines under a recording-level partition that removes the window-adjacency overlap present in the pooled five-fold scheme. For XJTU-SY, folds are formed so that all windows from a given bearing run and operating condition fall entirely into either the training or the test side, and the held-out conditions are never seen during training. For CWRU, the same rule is applied at the levels of load condition, shaft speed, and fault-diameter recording, so the model is always tested on recordings whose exact

operating context did not contribute to the training window. The proposed PIDC-Net framework is compared against PI-TFF, MPINet, iPINN, PhyDT-DA, ClassBD, and WIDAN on both datasets, with mean and standard deviation reported over the five recording-level folds.

Table 4. Cross-recording split: test accuracy (mean $\pm$ std, %) over five recording-level folds.

Method	XJTU-SY	CWRU
PI-TFF	77.84 $\pm$ 2.41	79.51 $\pm$ 2.18
iPINN	80.36 $\pm$ 2.07	82.44 $\pm$ 1.89
MPINet	83.97 $\pm$ 1.72	85.78 $\pm$ 1.63
ClassBD	85.62 $\pm$ 1.58	87.29 $\pm$ 1.45
PhyDT-DA	88.05 $\pm$ 1.29	89.61 $\pm$ 1.22
WIDAN	89.93 $\pm$ 1.14	91.18 $\pm$ 1.07
PIDC-Net	93.67 $\pm$ 0.96	94.52 $\pm$ 0.89

Two readings follow from the table. The first concerns the gap between the pooled and the recording-level protocol. Every method loses accuracy once the same-recording windows can no longer appear on both sides of the split, confirming that part of the near-saturated accuracy in Section 5.2 came from window adjacency rather than genuine cross-recording generalization. The fall for PIDC-Net is the smallest of the seven methods, from the high nineties of the pooled scheme to 93.67% on XJTU-SY and 94.52% on CWRU, while PI-TFF, whose only physical mechanism is frequency-band parameter initialization, loses the most because it carries no component able to bridge a recording-level distribution shift. The ordering of the baseline tracks their capacity for cross-distribution transfer: WIDAN leads the baselines because its domain-adaptation branch was built for cross-machine transfer, PhyDT-DA follows through its simulation-to-reality alignment, and the methods without an explicit alignment mechanism trail behind.

The second reading concerns what the retained margin means for the work's central aim. After

removing the optimistic component of the pooled accuracy, PIDC-Net still leads the strongest baseline by 3.74 points on XJTU-SY and 3.34 points on CWRU, and the spread across recording-level folds remains below 1 point on both datasets. A margin that survives a leakage-controlled split, while the band across folds stays narrow, indicates that the physics-derived receptive fields and the asymmetric contrastive anchor encode structure that transfers to unseen recordings instead of memorizing recording-local patterns. For a setting defined by scarce labeled fault data, where no single recording can be assumed representative, the durability of the lead under cross-recording testing is the property that matters more than peak accuracy on a favorable split, and it strengthens rather than weakens the generalization claim made for the physics-guided multi-source design.

5.8 Physical parameter configuration and sensitivity analysis

Table 5 lists the bearing geometry, shaft speed, derived fault characteristic frequencies, segment length, and impulse-decay settings used by FPES for each dataset, so that the physics-derived quantities entering Eqs. (1) to (5) are reproducible. Geometry follows the manufacturer's specifications for the two benchmark bearings: the LDK UER204 used by XJTU-SY and the SKF 6205-2RS JEM drive-end bearing used by CWRU. Shaft rotation frequency and the resulting outer-race and inner-race frequencies are reported for the lowest-speed operating condition of each dataset, since that condition produces the lowest fault frequencies and so the binding case for the segment-length constraint in Eq. (1), and the frequencies scale linearly with shaft speed across the remaining conditions.

Table 5. Physical parameters of FPES for XJTU-SY and CWRU.

Parameter (symbol)	XJTU-SY	CWRU
Bearing model	LDK UER204	SKF 6205-2RS JEM
Rolling element count	8	9
Rolling element	7.92	7.94

diameter		
Pitch circle diameter	34.55	39.04
Contact angle	0	0
Sampling rate	25.6	12
Shaft rotation frequency	35	29.95
Outer-race frequency	107.9	107.4
Inner-race frequency	172.1	162.2
Cyclic constraint	2	2
Segment length	512	256
Damping coefficient	545 ± 18	520 ± 21
Decay set	{0.9, 0.7, 0.5, 0.3, 0.1}	{0.9, 0.7, 0.5, 0.3, 0.1}
Receptive field widths (samples)	{4, 16, 32, 56, 108}	{2, 8, 15, 27, 53}

The damping coefficient D_η is the one quantity in Table 5 that is not read from a datasheet, so its selection is described in full. For each dataset, a small set of labeled outer-race windows is taken, the squared Hilbert envelope of each isolated impulse is computed, and the exponential envelope of Eq. (4) is fitted to the envelope peaks by least squares on the logarithmic amplitude, which turns the fit into a linear log-decrement regression whose slope returns D_η . The value in Table 5 is the mean of these per-window estimates, and the dispersion around it stays within 4%, indicating that the decay rate of the resonance-band ring-down is stable enough to fix the receptive field widths once per dataset. Given D_η and the five decay ratios in η , Eq. (5) converts each ratio into a receptive field width, producing the multi-scale bank whose principal scales, the widest near 108 samples and the narrowest near one fault transient, appear in the physics-informed branch.

Table 6 reports how diagnostic accuracy responds when the physical parameters are deliberately mis-specified, since bearing geometry and shaft speed are not always known exactly in field deployment. The shaft rotation frequency and the d_r/D_c geometry ratio are perturbed together by a fixed fraction, the fault frequencies and receptive field widths are recomputed from the perturbed values through Eqs. (2) to (5), and the model is retrained and

evaluated on the pooled five-fold protocol at each perturbation level.

Table 6. Sensitivity of test accuracy (%) to mis-specified physical parameters.

Parameter error	XJTU-SY	CWRU
0% (nominal)	99.89	99.92
$\pm 5\%$	99.61	99.7
$\pm 10\%$	99.04	99.18
$\pm 20\%$	97.32	97.65

The degradation is gradual rather than abrupt. A 5% error in the assumed geometry or speed shifts the receptive field widths by at most a few samples and costs less than 3/10 of a point on either dataset, and even a 20% error keeps accuracy above 97%. The tolerance for moderate parameter error has a clear source: the bank spans five decay ratios. Therefore, a small misplacement of one scale still leaves neighboring scales aligned with the true transient duration, and the cross-sensor alignment stream together with the asymmetric contrastive anchor recovers part of the mismatch by reweighting the statistically adaptive representation toward the physically constrained one. Only when the error grows large enough to move every receptive field off the genuine impulse duration does accuracy fall more steeply, which marks the point at which the physical prior stops matching the signal. For practice, the reading is reassuring, because approximate geometry from a bearing catalog or a nameplate speed places the receptive fields inside the tolerant region, and exact metrology is not needed to reach the reported performance.

5.9 Discussion

The combined and consistent results of the experiments confirm that PIDC-Net is superior in diagnostic performance compared to baseline methods across all assessment dimensions studied, and that this superiority is not an artefact of any one constitutionally favourable experimental condition but rather a constant attribute of the architecture of the recommended framework.

The training dynamics show that PIDC-Net

achieves near-perfect accuracy at epoch 50 and maintains a test accuracy of 99.89 percent at epoch 100, without any indication of overfitting. This quantitatively verified behaviour, indicating a basic benefit of physically constrained encoders, is that the depth-based kernel widths as derived based on bearing impulse decay in the FPES branch will remove a significantly larger fraction of spurious time sequences in the hypothesis space than a data-driven only network would, without needing to take a single gradient update. The fact that the precision, recall, and F1 scores are all 1.0 also attests that this fast convergence is not achieved at the expense of per-class balance.

The ablation study is especially informative as it breaks the accuracy gain into attributable contributions. The 25.1 percentage-point loss on average when FPES is not used establishes that physics-informed multi-scale depth-wise encoding is the most decisive single choice that can be made in architectures, and that the two fault types that appear to be the same signal only differing in the magnitude of impulses and harmonic contents continue to be persistently confused until the entire network of PIDC-Net becomes active. This shows that the elements of this framework are not overlapping but complementary.

The advantage of PIDC-Net is not regime-specific. It still achieves 94.18% accuracy with 10% training data, and PhyDT-DA, its most powerful baseline, achieves 75.4%, showing that the asymmetric contrastive objective delivers real sample efficiency by leveraging pairwise attractiveness relationships observed across streams rather than raw class labels. When PIDC-Net achieves 99.76% at 2 dB SNR, and the optimal competing single-stream version is 97.90% then it reinforces the claim that the three synergistic noise-reduction mechanisms, namely physics-matched depth-wise filtering, learnable affine cross-sensor adaptation, and bidirectional temporal periodicity encoding, are working together and not individually. Collectively, these findings confirm PIDC-Net as a stable and robust method to address the compound dilemma of multi-source heterogeneity and

data scarcity in bearing-failure diagnosis of the aero-engine.

6 Conclusion

This paper presented PIDC-Net, a physics-informed dual-stream contrastive diagnostic framework designed to address the compound challenge of multi-source sensor heterogeneity and labelled-sample scarcity in aero-engine bearing fault diagnosis. Experiments on the XJTU-SY and CWRU datasets validated that PIDC-Net achieved 99.99% diagnostic accuracy, outperformed the strongest physics-informed baseline by 3.79%, retained 94.18% accuracy with 10% training data, and maintained 99.88% accuracy at a 2 dB signal-to-noise ratio.

However, the physics-derived kernel widths in FPES rely on known bearing geometry parameters, which may not always be available in practice. Then, the decoupled encoder-classifier training pipeline increases total training time relative to fully end-to-end approaches. Future work will explore adaptive geometry parameter estimation from raw signals to remove dependence on known bearing specifications, extend the contrastive alignment mechanism to variable-speed aero-engine operating conditions, and investigate lightweight PIDC-Net variants suitable for edge deployment in onboard health-monitoring systems.

References

- [1] F. M. Talaat, and R. M. El-Balka, A fault detection of aero-engine rolling bearings based on CNN-BiLSTM network integrated cross-attention, *Meas. Sci. Technol.*, vol. 35, no. 12, 126116, 2023.
- [2] X. D. Han, J. Hong, C. Chen, and X. Q. Chen, Analysis of Failure Forms of Bevel Gear, Shaft, and Bearing Structure System Considering Multi-Mode Damage Accumulation, *J. Eng. Gas Turbines Power*, vol. 146, no. 2, 021020, 2024.
- [3] Z. Fang, Q. E. Wu, W. J. Wang, and S. Y. Wu, Research on Improved Fault Detection Method of Rolling Bearing Based on Signal Feature Fusion *Technology, Appl. Sci.*, vol. 13, no. 24, 12987, 2023.
- [4] Y. M. Xiao, H. D. Shao, J. Wang, S. Yan, and B. Liu, Bayesian variational Transformer: A generalizable model for rotating machinery fault diagnosis, *Mech. Syst. Signal Process.*, vol. 207, 110936, 2024.
- [5] Y. M. An, K. Zhang, Y. Chai, Z. Y. Zhu, and Q. Liu, Gaussian mixture variational-based Transformer domain adaptation fault diagnosis method and its application in bearing fault diagnosis, *IEEE Trans. Ind. Informat.*, vol. 20, no. 1, pp. 615–625, 2024.
- [6] M. H. Vu, V. Q. Nguyen, T. T. Tran, V. T. Pham, and M. T. Lo, Few-shot bearing fault diagnosis via ensembling Transformer-based model with Mahalanobis distance metric learning from multiscale features, *IEEE Trans. Instrum. Meas.*, vol. 73, 2513618, 2024.
- [7] H. Y. Wang, C. J. Li, P. Ding, S. B. Li, T. D. Li, C. Y. Liu, X. J. Zhang, and Z. J. Hong, A novel transformer-based few-shot learning method for intelligent fault diagnosis with noisy labels under varying working conditions, *Reliab. Eng. Syst. Saf.*, vol. 251, 110400, 2024.
- [8] Z. Y. Wang, Q. S. Luo, H. B. Chen, J. S. Zhao, L. Yao, J. W. Zhang, and F. L. Chu, A high-accuracy intelligent fault diagnosis method for aero-engine bearings with limited samples, *Comput. Ind.*, vol. 159–160, 104099, 2024.
- [9] P. Ding, Y. F. Xu, and X. Sun, Multitask learning for aero-engine bearing fault diagnosis with limited data, *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.
- [10] J. S. Zhong, Y. Zheng, C. T. Ruan, L. Chen, X. Y. Bao, and L. Lyu, M-IPISincNet: An explainable multi-source physics-informed neural network based on improved SincNet for rolling bearings fault diagnosis, *Inf. Fusion*, vol. 115, 102761, 2025.
- [11] Y. S. Su, L. J. Shi, K. Zhou, G. X. Bai, and Z. Q. Wang, Knowledge-informed deep networks for robust fault diagnosis of rolling bearings, *Reliab. Eng. Syst. Saf.*, vol. 244, 109863, 2024.
- [12] J. H. Lv, B. G. Kim, B. D. Parameshachari, A. Slowik, K. Q. Li, Large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors, *Inf. Fusion*, vol. 115, 102780, 2025.
- [13] C. L. Pan, Z. W. Shang, L. T. Tang, H. C. Cheng,

- and W. X. Li, Open-set domain adaptive fault diagnosis based on supervised contrastive learning and a complementary weighted dual adversarial network, *Mech. Syst. Signal Process.*, vol. 222, 111780, 2025.
- [14] Y. T. Dong, H. K. Jiang, R. H. Yao, M. Z. Mu, and Q. Yang, Rolling bearing intelligent fault diagnosis towards variable speed and imbalanced samples using multiscale dynamic supervised contrast learning, *Reliab. Eng. Syst. Saf.*, vol. 243, 109805, 2024.
- [15] X. Y. Li, C. M. Cheng, and Z. K. Peng, Label-guided contrastive learning with weighted pseudo-labeling: A novel mechanical fault diagnosis method with insufficient annotated data, *Reliab. Eng. Syst. Saf.*, vol. 254, 110597, 2025.
- [16] Q. Y. Luo, J. L. Chen, Y. Y. Zi, and J. S. Xie, A synchronization-induced cross-modal contrastive learning strategy for fault diagnosis of electromechanical systems under semi-supervised learning with current signal, *Expert Syst. Appl.*, vol. 249, 123801, 2024.
- [17] M. Q. Miao and J. B. Yu, Multiview contrastive shapelet learning: A novel semi-supervised approach for explainable machine fault diagnosis with insufficient annotated data, *IEEE Trans. Instrum. Meas.*, vol. 74, 3545411, 2025.
- [18] M. Y. Ye, X. A. Yan, D. Jiang, L. Xiang, and N. Chen, MIFDELN: A multi-sensor information fusion deep ensemble learning network for diagnosing bearing faults in noisy scenarios, *Knowl.-Based Syst.*, vol. 284, 111294, 2024.
- [19] Z. Z. Xu, X. Chen, and J. T. Xu, Multi-modal multi-sensor feature fusion spiking neural network algorithm for early bearing weak fault diagnosis, *Eng. Appl. Artif. Intell.*, vol. 141, 109845, 2025.
- [20] M. Y. Ye, X. A. Yan, X. Hua, D. Jiang, L. Xiang, and N. Chen, MRCFN: A multi-sensor residual convolutional fusion network for intelligent fault diagnosis of bearings in noisy and small sample scenarios, *Expert Syst. Appl.*, vol. 259, 125214, 2025.
- [21] K. K. Wu, J. J. Li, L. C. Meng, F. L. Li, and K. Lu, Online adaptive fault diagnosis with test-time domain adaptation, *IEEE Trans. Ind. Informat.*, vol. 21, no. 1, pp. 107–117, 2025.
- [22] W. W. Qian, Y. T. Chu, Z. Y. Lu, L. H. Tian, and B. Yu, MACPTD: Self-Supervised RUL Prediction for Aero-Engines via Frequency-Constrained Disentanglement and Dual-Domain Contrastive Learning, *IEEE Sens. J.*, vol. 26, no. 2, pp. 3237–3249, 2026.
- [23] S. Q. Wu, L. F. Yang, and L. L. Tao, Synergistic WSET-CNN and Confidence-Driven Pseudo-Labeling for Few-Shot Aero-Engine Bearing Fault Diagnosis, *Processes*, vol. 13, no. 7, 1970, 2025.
- [24] W. H. Yue and H. J. Wang, DG-LDM: A Dual-Guided Latent Diffusion Model for data augmentation in imbalanced fault diagnosis, *Adv. Eng. Inform.*, vol. 71, 104405, 2026.
- [25] H. T. Zhou, W. H. Chen, P. R. Qiao, L. S. Cheng, and M. Xia, Reliable fault diagnosis using evidential aggregated residual network under varying working conditions and noise interference, *Knowl.-Based Syst.*, vol. 286, 111407, 2024.
- [26] H. H. Gao, X. R. Zhang, X. J. Gao, F. Y. Li, and H. G. Han, Multi-timescale attention residual shrinkage network with adaptive global-local denoising for rolling-bearing fault diagnosis, *Knowl.-Based Syst.*, vol. 302, 112478, 2024.
- [27] Z. J. Xie, J. L. Chen, Z. K. Shi, S. L. Liu, and S. He, Lightweight pyramid attention residual network for intelligent fault diagnosis of machine under sharp speed variation, *Mech. Syst. Signal Process.*, vol. 223, 111824, 2025.
- [28] C. L. Ji, C. Zhang, L. M. Suo, Q. L. Liu, and T. Peng, Swarm intelligence based deep learning model via improved whale optimization algorithm and bi-directional long short-term memory for fault diagnosis of chemical processes, *ISA Trans.*, vol. 147, pp. 227–238, 2024.
- [29] M. Sun, X. Xiao, T. Chen, Q. He, Z. L. Long, and L. X. Li, A novel domain incremental learning method for bearing fault diagnosis based on F&K, *IEEE Trans. Ind. Informat.*, vol. 21, no. 1, pp. 980–989, 2025.
- [30] Z. Y. Wu, J. V. Guo, Y. C. Liu, L. J. Li, and Y. Ji, An

- iterative resampling deep decoupling domain adaptation method for class-imbalance bearing fault diagnosis under variant working conditions, *Expert Syst. Appl.*, vol. 252, 124240, 2024.
- [31] Y. Kim and Y. K. Kim, Physics-Informed Time-Frequency Fusion Network With Attention for Noise-Robust Bearing Fault Diagnosis, *IEEE Access*, vol. 12, pp. 12517–12532, 2024.
- [32] C. Gao, Z. Wang, Y. Guo, H. Wang, and H. Yi, MPINet: Multiscale Physics-Informed Network for Bearing Fault Diagnosis With Small Samples, *IEEE Trans. Ind. Inform.*, vol. 20, no. 12, pp. 14371–14380, 2024.
- [33] Y. Qin, H. Liu, Y. Wang, and Y. F. Mao, Inverse Physics-Informed Neural Networks for Digital Twin-Based Bearing Fault Diagnosis Under Imbalanced Samples, *Knowl.-Based Syst.*, vol. 292, no. 111641, 2024.
- [34] D. Peng, M. Yazdanianasr, A. Mauricio, T. Verwimp, W. Desmet, and K. Gryllias, Physics-Driven Cross Domain Digital Twin Framework for Bearing Fault Diagnosis in Non-Stationary Conditions, *Mech. Syst. Signal Process.*, vol. 228, no. 112266, 2025.
- [35] J. X. Liao, C. He, J. Li, J. Sun, S. Zhang, and X. Zhang, Classifier-Guided Neural Blind Deconvolution: A Physics-Informed Denoising Module for Bearing Fault Diagnosis Under Noisy Conditions, *Mech. Syst. Signal Process.*, vol. 222, no. 111750, 2025.
- [36] C. He, H. Shi, X. Liu, and J. Li, Interpretable Physics-Informed Domain Adaptation Paradigm for Cross-Machine Transfer Diagnosis, *Knowl.-Based Syst.*, vol. 288, no. 111499, 2024.
- [37] H. Xu, Y. C. Xiao, K. Sun, and L. L. Cui, Improved depthwise separable convolution for transfer learning in fault diagnosis, *IEEE Sens. J.*, vol. 24, no. 20, pp. 33606–33613, 2024.
- [38] B. Najdi, M. Benbrahim, and M. N. Kabbaj, Adaptive Res-LSTM Attention-based Remaining Useful Lifetime Prognosis of Rolling Bearings, *Int. J. Progn. Health Manag.*, vol. 16, no. 1, 4171, 2025.
- [39] R. Balamurugan, D. G. Takale, M. M. Parvez, and S. Gnanamurugan, A novel prediction of remaining useful life time of rolling bearings using convolutional neural network with bidirectional long short term memory, *J. Eng. Res.*, vol. 13, no. 3, pp. 1695-1705, 2025.