The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Uncovering Hidden Degeneration: A Physics-Guided Bidirectional Inference Framework for Industrial Time Series Prediction

Xingwang Li[1], **Fei Teng**[1,2] *, **Xin Wu**[1], **Qiang Duan**[3]
[1] School of Computing and Artificial Intelligence, Southwest Jiaotong University
[2] Engineering Research Center of Sustainable Urban Intelligent Transportation, Ministry of Education
[3] Information Sciences and Technology Department, The Pennsylvania State University
{lixingwang, xinwu}@my.swjtu.edu.cn, fteng@swjtu.edu.cn, qxd2@psu.edu

Abstract

Hidden degenerations in industrial time series often precede observable failures, they remain undetected by standard monitoring systems until anomalies become apparent. This gap between microscopic degradation and macroscopic observation renders conventional predictors inherently reactive, as they rely on correlations in sensor data rather than uncovering the underlying, physics-consistent degradation states. Crucially, the microscopic mechanisms governing system evolution depend on macroscopic state variables—whose measurements are expectations over microscopic probability distributions—so purely data-driven “top-down” or purely physics-guided “bottom-up” approaches cannot forecast degeneration-entangled industrial faults. To address these challenges, we propose a Physics-Guided Bidirectional Inference Framework that represents hidden microscopic states from macroscopic measurements. Our approach uniquely combines: (1) bottom-up physics-based simulation using Continuum Damage Mechanics to model micro-scale damage evolution under environmental stressors, and (2) top-down probabilistic inference via maximum entropy formalism to estimate latent microstate distributions from sparse sensor data. This bidirectional mechanism enables early failure prediction by bridging observable measurements with unobservable degeneration. Validation on real-world railway infrastruc datasets demonstrates significant improvements in early fault prediction compared to state-of-the-art baselines. Our method establishes a new paradigm for safety-critical industrial applications requiring reliable prediction of hidden degeneration processes.

Code — https://github.com/benjaminlbj38-eng/PIBI

Introduction

Industrial time series, capturing vital process information, can be leveraged through analysis and forecasting to facilitate downstream production operations—such as plant scheduling and fault prediction (Sha et al. 2024). Owing to the stringent demands for reliability and continuity in industrial processes, fault prediction has garnered significant attention from both industry and academia (Chen et al. 2023b; Lohia et al. 2024). Faults predominantly stem from the gradual degradation of production equipment—a process inherently unobservable and influenced by operational environments (Xia et al. 2024; Lu et al. 2024). Current monitoring approaches, however, capture only the final stages of this process, missing the crucial early phases where preventive action remains possible (Benkabou et al. 2025; Ho, Karami, and Armanfard 2025). The fundamental challenge thus becomes detecting and modeling these hidden degeneration dynamics that precede visible failure indicators.

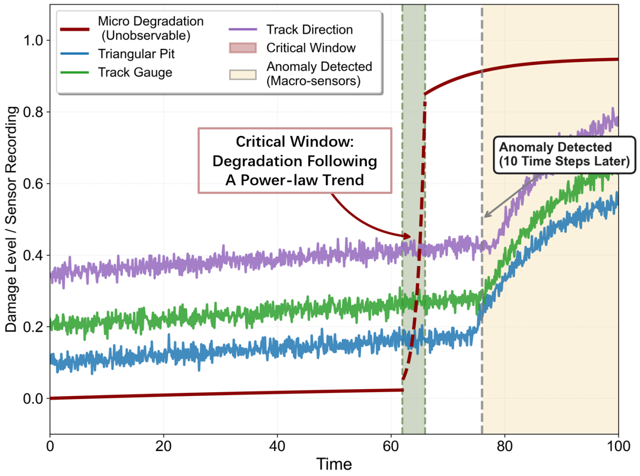

Figure 1: Fault degeneration under micro–macro asymmetry. The figure illustrates a common failure scenario in industrial systems, where micro degeneration (red line) accumulates silently before any observable changes (eg. track gauge) appear in macro-scale sensor readings.

The hidden nature of early-stage degeneration creates a critical blind spot in fault prediction. In Fig. 1, micro-level degeneration (red curve) accelerates exponentially while macro-scale sensor readings remain deceptively stable until approaching failure thresholds. Traditional detection methods identify deviations only after the *critical window* has passed—when degeneration has already progressed beyond preventive intervention. The temporal gap between actual degeneration onset and sensor-detectable anomalies renders most predictive models inherently reactive rather than truly

---

[*] Corresponding author: Fei Teng

preventive. Bridging this micro-macro disconnect requires approaches that can infer hidden degeneration states from available observations before they manifest as detectable anomalies.

Most existing fault prediction pipelines remain purely data driven (Zhang et al. 2025). Multivariate histories are fed into autoencoders, graph neural networks, and remaining useful life predictors to map past macroscopic signals to future failures (Fang et al. 2025). Implicitly, these models rest on two assumptions: (A) observability sufficiency—all precursors of failure are encoded in the measured streams; and (B) statistical continuity—the mapping from signals to degeneration is smooth and learnable from historical data. However, these assumptions are not always valid. Many damage mechanisms remain undetectable until a threshold is crossed, resulting in abrupt, discontinuous changes. Moreover, heterogeneous operating conditions can shift the underlying mapping, further undermining observability (Chen et al. 2025). Consequently, data-driven models tend to excel at identifying late-stage anomalies but fail to reconstruct the hidden progression of micro-damage (Harte et al. 2024).

Physics-constrained learning seeks to address the limitations of purely data-driven models by embedding physical laws within network architectures. Existing physics-constrained time series methods fall into two strands. The first locks governing ODE/PDE residuals into the training objective, yielding physics-informed networks (Chen, Liu, and Sun 2025; Zeng et al. 2025). While effective, these hard constraints become brittle when sensor noise, or parameter drift violate the residual tolerance, often forcing the optimiser toward degenerate minima. The second strand wraps a high-fidelity numerical simulator around a data-driven surrogate and calibrates parameters with observations (Tian et al. 2025; Liao et al. 2023). Such twin-model approaches demand an accurate, differentiable simulator, hampering deployment under tight latency or data budgets. Fundamentally, both approaches leverage physics only to constrain predictions, rather than establishing a bidirectional mapping that infers micro-level degradation from measurements.

To address these limitations, we propose a Physics Guided Micro–Macro Inference (PIBI) framework tailored to asymmetric observability in engineering degeneration. From the bottom up, a degeneration simulator based on Continuum Damage Mechanics (CDM) models the evolution of microstate variables under temperature fluctuations, load cycles, and other environmental stressors, thereby enabling forward simulations across diverse operating conditions. From the top down, a maximum entropy inference module treats the simulator as a prior and updates the latent microstate distributions using macroscopic measurements, thereby producing condition-aware, uncertainty-calibrated early warnings. The two directions interact bidirectionally: simulated trajectories regularize the inference network, while inferred posteriors adapt simulator parameters to asset specific heterogeneity, thereby bridging the micro–macro gap. Empirically, the framework captures nonlinear damage accumulation, incorporates environmental factors, and delivers earlier, more reliable alerts than purely data driven or rigid physics only baselines when validated on real railway infrastructure degeneration datasets. Our contributions can be summarized in three key aspects:

- We formalize a new class of degeneration prediction problems characterized by asymmetric observability between micro and macro scales, expose fundamental limitations in both data-driven and physics-informed approaches;
- We develop a novel bidirectional inference architecture that synergistically combines bottom-up physics-based degeneration simulation with top-down probabilistic microstate reconstruction, enabling condition-aware fault prediction.
- Extensive experiments conducted on real-world datasets show the effectiveness of PIBI compared with various state-of-the-art methods.

Related Work

Deep Learning in Fault Prediction

Recent studies on fault prediction have primarily focused on deep learning architectures that directly extract correlations from sensor data. SARAD detects anomalies by learning spatial inter-feature associations with a Transformer and identifying association descent patterns via autoencoding (Dai et al. 2024). Contrastive learning method enhances fault pattern discrimination in multivariate time series by enforcing temporal stability and sensor-wise relational consistency through graph-aware augmentation (Wang et al. 2024; Zhang et al. 2025). Hierarchical classification approach improves anomaly-sensitive representation by modeling time series as tokenized categories, capturing high-entropy patterns across granularity levels (Sun et al. 2025). Collectively, deep learning methods improve detection accuracy and robustness but still assume that the most informative precursors are observable (Chen et al. 2023a). These studies rarely address the challenges posed by the short onset windows that characterize abrupt industrial degradation.

In contrast, our framework treats the observed macroscale signals as delayed, lossy projections of rapidly evolving micro-states. By coupling a bottom-up degeneration simulator grounded in continuous damage mechanics with a top-down maximum-entropy inference layer, we explicitly infer the most plausible hidden damage distributions before macroscopic anomalies surface. This bi-directional design allows condition-aware early warning, provides interpretable physical pathways for each alarm.

Physics-Guided Modeling for Sequences

Physics-guided machine learning has increasingly been advocated to improve fidelity and generalization in scientific modeling, yet many approaches still target isolated processes and simplified regimes. Dual-branch architectures such as Air-DualODE emphasize the gap between explicit equations and latent neural representations, leveraging Neural ODEs to separate physical and residual dynamics, yet they leave unresolved the reconstruction of unobserved internal states from sparse outputs (Tian et al. 2025). Graph-based encoders such as PhyMPGN (Zeng et al. 2025) embed numerical integrators and Laplace–Beltrami operators

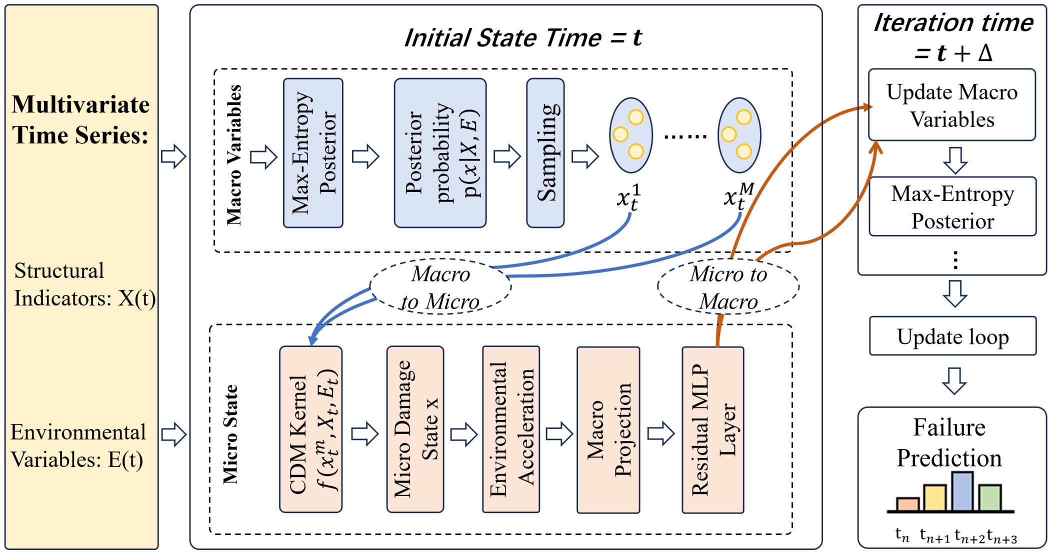

Figure 2: The framework of the proposed PIBI

to respect PDE structure, addressing mesh irregularity but relying on plentiful training data for each geometry. PINP (Chen, Liu, and Sun 2025) discretizes coupled equations directly into the architecture to gain long-horizon extrapolation, underscoring the benefit of hard physics constraints but focusing on fully observed physical fields. In industrial diagnostics, physics-informed CNNs mitigate noise and improve interpretability with filter designs derived from Fourier-domain priors, yet they remain feature extractors rather than latent-state reconstructors under non-symmetric observability. The open challenge is thus not only to "inject physics" but to reconcile multi-scale, partially observed dynamics where micro-level drivers are never directly measured, environments are non-stationary, and priors must coexist with principled uncertainty handling.

Our Physics-Guided Micro–Macro Inference closes this gap by unifying a forward CDM-based simulator (micro → macro) with a backward, entropy-principled estimator (macro → micro). The simulator captures nonlinear accumulation, environmental accelerants, and structural idiosyncrasies; the inference module, constrained by the same physics, yields distributions over latent damage consistent with sparse, delayed observations. This reciprocity transforms physics from a post hoc regularizer into an operational bridge between scales, enabling earlier, more reliable, and explainable warnings in safety-critical industrial settings.

Methodology

Model Architecture Overview

Fig. 2 illustrates the proposed PIBI's dual-stream architecture. A bottom-up Continuum Damage Mechanics (CDM) simulator forward-integrates latent micro-damage $x(t)$ under load, temperature and humidity, projecting the result onto measurable macro indicators $X(t)$. In parallel, a top-down maximum-entropy network treats the simulator as a Bayesian prior, inverting sparse sensor windows to obtain posteriors over $x(t)$ and asset-specific material parameters. The two streams share latent samples, physics-consistent moment constraints, and a multi-task objective combining binary-cross-entropy with a dual-scale consistency loss that aligns simulated and observed trajectories. End-to-end differentiation closes a bidirectional loop, yielding micro–macro trajectories, calibrated uncertainty, and earlier fault alarms than data-only or one-way physics-informed baselines.

Problem Formulation

We consider an industrial infrastructure system composed of S spatial segments (e.g., kilometer units along a railway track), each continuously monitored by a set of D sensors over time. For each segment $s \in \{1, \dots, S\}$, we collect a multivariate historical time series $\mathbf{X}^s = \{\mathbf{m}_1^s, \mathbf{m}_2^s, \dots, \mathbf{m}_t^s\} \in \mathbb{R}^{t \times D}$, where t is the number of time steps and $\mathbf{m}_t^s \in \mathbb{R}^D$ contains physical measurements such as geometry, vibration, and alignment at time t. Our goal is to predict whether segment s will exhibit structural damage within the future horizon τ, based on the historical sensor readings from a fixed look-back window of length w. The early warning target is defined as a binary label $y_s^{(t+\tau)} \in \{0, 1\}$, where y=1 denotes the damage in the future timing $t + \tau$. Formally, the task is to learn a classifier: $f_\theta : \mathbb{R}^{w \times D} \to \{0, 1\}$. Notably, in contrast to standard multivariate time series classification (MTSC), our problem involves predicting future events rather than assigning class labels to observed sequences. For clarity, in what follows, we fix a specific region s and omit the corresponding spatial subscript.

General Notation

Our framework operates on four core physical concepts that capture the complete degeneration process:

Micro-scale Damage State: $x(t) \in [0,1]$ represents the micro state at time t, where $x(t) = 1$ indicates no damage and $x(t) = 0$ denotes complete failure. This variable is the fundamental damage parameter in CDM theory, capturing the nonlinear degeneration of a material's internal stiffness and strength. Although the damage state is not directly observable, it dictates the entire future trajectory of material degeneration.

Macro-scale Indicators: $\mathbf{X}(t) = (X_1, X_2, \ldots, X_d)$ encompasses track geometry parameters (alignment, gauge) and other condition indicators recorded by inspection systems. These measurements reflect the cumulative manifestation of underlying damage processes but provide indirect evidence of the true health state. Environmental Variables: $\mathbf{E}(t) = [T(t), P(t), L(t)]$ captures external environmental conditions that influence degeneration kinetics: - $T(t)$: ambient temperature (°C) affecting corrosion-fatigue processes. $P(t)$: cumulative rainfall (mm) influencing moisture-related degeneration. $L(t)$: operational load intensity representing traffic-induced stress cycles.

Physics-Based Transition Function: The $f(x_t, \mathbf{X}_t, T(t), P(t), L(t))$ models the damage evolution mechanism based on established physics principles, transforming theoretical damage accumulation into quantifiable degeneration rates.

Failure Prediction Target: $S(t)$ represents the cumulative failure count over specified time horizons, serving as the prediction objective for maintenance planning.

Maximum Entropy Micro-State Inference

Purpose. Given a macro–level observation $X(t) \in \mathbb{R}^D$ and environmental variables $E(t) = [T(t), P_{\text{rain}}(t), L(t)]$, we seek the *least–biased* distribution of the latent micro-damage variable $x \in [0,1]$. Theoretically, the posterior probability does not have to be uniquely identifiable. Therefore, we explore a more practical question, whether the posterior variance is small enough to support reliable decision making under the available sampling density. We adopt the maximum-entropy form.

$$\begin{aligned} p(x \mid X(t), E(t)) = \frac{1}{Z} \exp\Big[&-\lambda_1(t)\, x \\ &-\lambda_2(t)\, f\big(x, X(t), E(t)\big)\Big], \end{aligned} \tag{1}$$

where Z is the partition function, x is the micro-damage score and $f(\cdot)$ is the CDM-based degeneration kernel, and $\lambda_{1,2}$ are Lagrange multipliers enforcing the constraints.

Moment constraints. Ideally, the Lagrange multipliers $\lambda_{1,2}(t)$ are determined by the coupled constraints:

$$\mathbb{E}_p[x] = \hat{x}(t), \quad \mathbb{E}_p[f(\cdot)] = \hat{r}(t), \tag{2}$$

where $\hat{x}(t)$ is the *macro–to–micro* estimate produced by a Transformer-based inference network that consumes the most recent window $\{X(t-\tau{:}t), E(t-\tau{:}t)\}$ and outputs a scalar health estimate. $\hat{r}(t) = f\big(\hat{x}(t), X(t), E(t)\big)$ is the *physics-consistent* average degeneration rate obtained by evaluating the same kernel f at $\hat{x}(t)$. Instead of enforcing $E_p[f(x)] = \hat{r}(t)$ as a strict equality, we adopt the modeling approximation $E_p[f(x)] \approx f(E_p[x])$ and softly regularize the residuals through a penalty term:

$$\mathcal{L}_{\text{mom}} = \mu_1\big(E_p[x] - \hat{x}\big)^2 + \mu_2\big(E_p[f(x)] - \hat{r}\big)^2, \tag{3}$$

where μ_1, μ_2 are small coefficients (empirically $10^{-2} \sim 10^{-1}$). This ensures numerical stability without requiring inner Newton iterations.

Physics-Consistent Forward Simulation

Purpose. This module propagates each micro-damage sample forward in time and projects it back to observable macro indicators.

Micro-level time stepping. From (1) we draw M i.i.d. samples $\{x_t^{(m)}\}_{m=1}^{M}$ and propagate each according to the CDM kernel:

$$x_{t+\Delta}^{(m)} = x_t^{(m)} - \int_{s=t}^{t+\Delta} f\big(x_s^{(m)}, X_s, E_s\big)\, \mathrm{d}s\,, \tag{4}$$

where the integration variable is indicated by s. We adopt a single-step fourth-order Runge–Kutta scheme for numerical stability. To ensure $x_t \in [0,1]$ during integration, we reparameterize $x = \sigma(\eta) = 1/(1 + e^{-\eta})$ and integrate $\dot{\eta} = -f(\sigma(\eta), X, E)/[\sigma(\eta)(1 - \sigma(\eta))]$. This logit transformation guarantees boundedness while keeping the dynamics differentiable.

Macro projection. Updated micro states are aggregated via a differentiable operator $G : [0,1] \to \mathbb{R}^D$, which projects the ensemble-averaged micro damage $\bar{x}$ onto a physics-guided stress-gradient constructed from an analytic kernel plus a two-layer residual MLP calibrated during training.

$$\begin{gathered} \frac{\mathrm{d}X_j}{\mathrm{d}t} = G_j\big(\bar{x}, X, E\big), \quad j = 1, \ldots, D, \\ \bar{x} := \tfrac{1}{M} \sum_{m=1}^{M} x^{(m)} \\ G(x, X, E) = K_\gamma * x + \text{MLP}_\theta([x, X, E]), \end{gathered} \tag{5}$$

where $K_\gamma(r) = \exp(-\gamma r^2)$ is a Gaussian kernel and MLP_θ has two hidden layers of 64 units with ReLU activations. The convolution $K_\gamma * x$ captures spatial smoothing, while the residual term models nonlinear corrections.

Dual-scale consistency loss. Simulated macro trajectories $\tilde{X}_{t+\Delta}^{(m)} = X_t + \Delta\, G\big(x_t^{(m)}, X_t, E_t\big)$ are aligned with real observations by

$$\mathcal{L}_{\text{cons}} = \frac{1}{M} \sum_m \left\| \tilde{X}_{t+\Delta}^{(m)} - X_{t+\Delta} \right\|_2^2 \tag{6}$$

gradients propagate to *all* parameters in Eqs. (1)–(6), thereby closing the micro–macro training loop.

Physics-Guided Damage Evolution Mechanism

CDM-Based Kinetics. Following Continuum Damage Mechanics (Voyiadjis and Kattan 2025), the micro-damage state $x \in [0, 1]$ deteriorates under mechanical stress and environmental accelerants:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -f(x, \mathbf{X}, \mathbf{E}(t)) \tag{7}$$

where $X(t)$ denotes the current macro-level indicators. The degeneration kernel f is factorised into three multiplicative constituents that reflect material degeneration, mechanical solicitation and environmental acceleration, respectively:

$$f(x, \mathbf{X}, \mathbf{E}) = \alpha(x)^n \sigma_{\text{eff}}(\mathbf{X}, L) \phi_{env}(\mathbf{E}). \tag{8}$$

Material-degeneration term: the factor $(x)^n$ captures the classical CDM observation that once damage initiates, subsequent degradation accelerates with the remaining intact fraction (x). Mechanical-stress term: $\sigma_{\text{eff}}(\mathbf{X}, L)$ represents the stress that blends geometry-derived stresses from $\mathbf{X}$ with the external load $L(t)$. In the general case we set

$$\sigma_{\text{eff}}(X, E) = \beta_1 \, \|\tilde{X}\|_2 + \beta_2 \, \tilde{L}(t), \tag{9}$$

where β_1 and β_2 are coefficients that are learned together with the network weights. The quantities $(\tilde{X}, \tilde{L})$ are normalized by their nominal reference scales $(X_{\text{ref}}, L_{\text{ref}})$ to ensure unit consistency. For the railway-rail degeneration prediction, the $|\mathbf{X}\|_2$ is incorporating rail-specific geometric variables (track alignment, gauge, etc). Alternative infrastructures can be seamlessly integrated without modifying the core framework.

Environmental-acceleration term

Theorem 1. *(Arrhenius Kinetics Law) For thermally activated processes, the reaction rate follows an exponential temperature dependence: $k(T) = A \exp(-E_a/RT)$, where $k(T)$ is the rate constant, A is the pre-exponential factor, E_a is the activation energy, R is the gas constant, and T is absolute temperature (Celina et al. 2019).*

The Arrhenius-type factor

$$\phi_{\text{env}}(\mathbf{E}) = \exp\left[-\frac{Q}{R(T + 273.15)}\right](1 + \text{RH} + P_{\text{rain}}) \tag{10}$$

modulates the intrinsic kinetics by temperature T (through the activation energy Q and the gas constant R and by moisture variables, RH refers to the average relative humidity, and P_{rain} denotes the accumulated rainfall. All scalar physicochemical coefficients $\ln\alpha, \ln n, \ln Q, \beta_1, \beta_2$ are instantiated as parameters. Specifically, each kinetic coefficient is treated as a latent random variable with an informative prior (see Table 2), and is learned jointly with network weights via reparameterised gradients. During back-propagation, gradients flow through Eqs.7–10, ensuring that the learned kinetics remain physically plausible.

Failure Prediction and Risk Quantification

Failure forecasting is framed as a probabilistic classification problem over a user-defined horizon $H = K\Delta$. At test time we draw M Monte-Carlo micro-trajectories $\left\{x^{(m)}_{t+k\Delta}\right\}_{k=0}^{K}$ using the Runge–Kutta integrator. Each trajectory is aggregated into macro responses $\tilde{\mathbf{X}}^{(m)}_{t+k\Delta} = G\left(x^{(m)}_{t+k\Delta}\right)$. A failure indicator is then computed per step by

$$F^{(m)}_{t+k\Delta} = 1\left\{\min_{s \in \{t, t+\Delta, \ldots, t+k\Delta\}} x^{(m)}_s \le x_c\right\} \tag{11}$$

where x_c is material-specific safety limits established by domain rules. The step-wise failure probability results from the Monte-Carlo frequency

$$p_{t+k\Delta} = \frac{1}{M} \sum_{m=1}^{M} F^{(m)}_{t+k\Delta}, \quad k = 1, \ldots, K, \tag{12}$$

and the cumulative risk of experiencing at least one failure within the horizon equals $r_{t \to t+H} = 1 - \prod_{k=1}^{K}(1 - p_{t+k\Delta})$. During training, the ground-truth label $y_{t+k\Delta} \in \{0, 1\}$ indicates whether a defect is verified at time $t + k\Delta$. We supervise the predictive distribution with a binary-cross-entropy term

$$\begin{aligned} \mathcal{L}_{\text{BCE}} = -\frac{1}{K} \sum_{k=1}^{K} \Big[& y_{t+k\Delta} \log p_{t+k\Delta} \\ & + (1 - y_{t+k\Delta}) \log(1 - p_{t+k\Delta}) \Big] \end{aligned} \tag{13}$$

The kinetic coefficients (α, β, n) are treated as random parameters with informative log-normal priors $p(\theta) = \text{LN}(\mu_\theta, \sigma_\theta^2)$. During training we perform MAP estimation by adding the log-prior term:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{cons}}. \tag{14}$$

This preserves the intended physical regularization without introducing variational inference overhead. By grounding $p_{t+k\Delta}$ in physically consistent micro-trajectories rather than relying on a standalone classifier—the framework gains intrinsic interpretability: analysts can trace each high-risk forecast back to its underlying stress pathways and environmental drivers.

Experiment

Experimental Settings

Datasets. We evaluate our approach on two real-world structural monitoring datasets. The first is a long-term heavy-haul railway dataset collected monthly from 2018 to 2025 over a 581 km freight corridor in China, encompassing geometric, load, gradient, and environmental variables. The second is the KW51 (Maes and Lombaert 2021) bridge dataset from Belgium, comprising high-frequency vibration and strain signals recorded over 15 months across pre−, during−, and post-retrofit stages. Both datasets reflect diverse degeneration patterns under real operational and environmental conditions. The final dataset contains no personal or sensitive information.

Methods	Heavy-haul Railway Tracks							
	F1-score	Precision	Recall	MCC	Accuracy	ROC-AUC	PR-AUC	G-mean
AttnPINN (Adv. Eng. Informatics 2023)	67.83±0.91	80.54±0.62	63.21±0.57	68.01±0.44	86.18±0.88	80.07±0.35	70.84±0.37	76.21±0.31
PINP (ICLR 2025)	63.47±1.02	77.92±0.73	58.03±0.66	64.15±0.51	84.97±1.05	78.32±0.49	68.45±0.46	72.04±0.42
Air-DualODE (ICLR 2025)	66.19±0.84	79.81±0.59	61.47±0.55	66.93±0.40	86.92±0.81	79.64±0.38	70.12±0.35	74.98±0.36
PhyMPGN (ICLR 2025)	65.87±0.78	79.43±0.57	61.02±0.53	67.16±0.39	86.56±0.79	81.01±0.33	71.10±0.29	**77.02±0.27**
TS-GAC (AAAI 2024)	60.41±1.21	74.36±0.95	59.28±0.82	66.92±0.68	83.91±1.14	76.25±0.61	66.17±0.55	69.77±0.51
SARAD (NeurIPS 2024)	62.54±0.97	76.12±0.71	**67.73±0.64**	63.48±0.47	84.22±0.93	78.91±0.41	69.03±0.44	73.61±0.39
MENTORPDM (KDD 2025)	64.73±0.89	78.85±0.66	60.51±0.59	65.42±0.45	85.11±0.87	79.22±0.36	**71.92±0.32**	74.36±0.34
HCAN (AAAI 2025)	67.18±1.05	75.07±0.82	56.43±0.74	61.02±0.57	84.06±0.99	77.38±0.53	66.88±0.48	70.52±0.45
PIBI	**70.21±0.68**	**82.32±0.45**	66.04±0.43	**71.42±0.32**	**87.47±0.96**	**82.21±0.27**	71.68±0.23	76.95±0.28
	KW51							
	F1-score	Precision	Recall	MCC	Accuracy	ROC-AUC	PR-AUC	G-mean
AttnPINN (Adv. Eng. Informatics 2023)	82.83±0.63	89.52±0.67	77.13±0.61	79.04±0.46	95.21±0.69	91.12±0.29	82.03±0.34	85.86±0.28
PINP (ICLR 2025)	80.64±0.72	87.29±0.74	74.91±0.66	76.18±0.52	94.63±0.77	89.96±0.35	80.54±0.39	84.69±0.31
Air-DualODE (ICLR 2025)	83.07±0.59	89.23±0.65	78.41±0.60	79.77±0.44	**96.86±0.68**	92.03±0.27	82.72±0.33	86.53±0.27
PhyMPGN (ICLR 2025)	84.01±0.57	90.01±0.63	81.02±0.58	81.51±0.42	95.92±0.70	92.39±0.26	83.88±0.32	**88.02±0.24**
TS-GAC (AAAI 2024)	78.46±0.81	86.09±0.79	74.24±0.73	73.41±0.57	93.76±0.88	89.37±0.41	79.75±0.45	83.12±0.36
SARAD (NeurIPS 2024)	81.72±0.66	88.78±0.70	75.52±0.63	77.91±0.48	95.04±0.73	91.44±0.30	83.02±0.36	85.23±0.29
MENTORPDM (KDD 2025)	83.57±0.61	87.01±0.66	78.01±0.60	80.19±0.45	95.78±0.71	92.01±0.27	**85.41±0.28**	86.92±0.25
HCAN (AAAI 2025)	79.81±0.75	90.22±0.76	76.47±0.69	75.06±0.53	94.12±0.80	90.47±0.37	81.04±0.41	84.01±0.33
PIBI	**86.32±0.55**	**91.46±0.61**	**82.27±0.58**	**82.93±0.41**	96.12±0.72	**93.21±0.24**	84.66±0.31	87.20±0.26

Table 1: Comparisons with State-of-the-Art methods on two datasets under eight evaluation metrics (%). The best experimental results are highlighted in bold, while the second-best results are underlined.

Symbol	Description	Prior
α	Damage-rate scale	$\log\mathcal{N}(\mu{-}8,\ \sigma{=}0.5)$
n	Damage exponent	$\mathcal{N}(2.0,\ 0.2^2)$
m	Fatigue exponent	$\mathcal{N}(3.0,\ 0.2^2)$
Q	Activation energy	$\mathcal{N}(1.2 \times 10^5,\ 10^{4^2})$
β_1	Geometry stress coeff.	$\log\mathcal{N}(\mu{-}1,\ \sigma{=}0.3)$
β_2	Load stress coeff.	$\log\mathcal{N}(\mu{-}1,\ \sigma{=}0.5)$
σ_c	Ultimate stress	fixed 350 MPa

Table 2: Kinetic coefficients and priors.

Baselines. We compare our model against two categories of recently advanced and highly relevant approaches. To highlight our performance, we select four **physics-informed methods** and four **time series analysis methods**: Physics-informed methods: AttnPINN (Liao et al. 2023), PINP (Chen, Liu, and Sun 2025), Air-DualODE (Tian et al. 2025), PhyMPGN (Zeng et al. 2025). Time series analysis methods: TS-GAC (Wang et al. 2024), SARAD (Dai et al. 2024), MENTORPDM (Zhang et al. 2025), HCAN (Sun et al. 2025).

Experiments details. Following prior studies, we adopt ADAM (Kingma and Ba 2015) as the default optimizer for all experiments. We evaluate models using F1 score, precision, recall, MCC, accuracy, and G-mean (threshold-dependent metrics), which jointly capture complementary aspects of performance—including positive-class detection, false-alarm control, and balanced consideration of both classes. Among these, MCC and G-mean are particularly robust, as they incorporate all four cells of the confusion matrix. In addition, we report ROC-AUC and PR-AUC as threshold-independent ranking metrics.

The initial learning rate is selected via grid search from $\{5 \times 10^{-3}, 10^{-3}, 5 \times 10^{-4}, 10^{-4}, 5 \times 10^{-5}, 10^{-5}\}$,tuned separately for each dataset. All runs begin with the same prior hyper-parameters (Table 2). The experiments were repeated five times with fixed random seeds, and we reported the average performance. HCAN was implemented by PyTorch and trained on a single NVIDIA RTX 4090 24GB GPU.

Main Results

On the long-span, monthly-sampled railway dataset, our physics-guided bidirectional inference achieves the best performance on five out of eight metrics. Compared to the strongest baseline (AttnPINN, PhyMPGN) our method raises F1 from 67.83 to 70.21 (+3.5% relative) and MCC from 68.01 to 71.42 (+5.0%), while also delivering the highest ROC-AUC (82.21) and Accuracy (87.47). The gain stems from explicitly reconstructing latent micro-damage distributions from sparse macroscopic measurements: the physics prior constrains the degeneration in poorly observed intervals, and the maximum-entropy top-down inference prevents overconfident false alarms, yielding the highest Precision (82.32). Interestingly, our Recall (66.04) is second-best to SARAD (67.73), revealing a principled precision–recall trade-off induced by conservative, physics-consistent thresholds. Consistently, our PR-AUC ranks second (71.68 vs. 71.92), and the G-mean is slightly below the best physics-informed competitor (76.95 vs. 77.02), indicating that under severe class imbalance our model prioritizes precision and

Method	F1	Precision	Recall	MCC	Acc.	ROC-AUC	PR-AUC
Ours	**70.21** ± 0.08	**82.32** ± 0.16	**66.04** ± 0.09	**71.42** ± 0.14	**87.47** ± 0.17	**82.21** ± 0.05	**71.68** ± 0.07
w/o CDM physics	65.35 ± 0.14	78.51 ± 0.10	59.42 ± 0.15	64.12 ± 0.13	82.40 ± 0.07	74.10 ± 0.18	65.75 ± 0.16
w/o MaxEnt	67.81 ± 0.11	80.45 ± 0.07	61.20 ± 0.12	68.05 ± 0.09	85.10 ± 0.06	<u>80.31</u> ± 0.11	<u>69.72</u> ± 0.08
w/o Env. term	<u>68.90</u> ± 0.10	<u>81.72</u> ± 0.07	<u>62.80</u> ± 0.11	<u>69.70</u> ± 0.08	<u>86.35</u> ± 0.06	79.40 ± 0.05	68.35 ± 0.09

Table 3: Ablation study results on *Heavy-haul Railway Tracks* dataset. Mean ± standard deviation over five runs. The best experimental results are highlighted in bold, while the second-best results are underlined.

calibration over aggressive recall.

On the high-frequency bridge dataset (KW51), richer observability sharpens the probabilistic posterior, allowing our model to simultaneously keep the physics prior as a regularizer. We attain the best F1 (86.32), Precision (91.46), MCC (82.93), and ROC-AUC (93.21), improving F1 by +2.31 absolute over the best baseline (PhyMPGN). Accuracy (96.12), PR-AUC (84.66), and G-mean (87.20) are second-best, reflecting a deliberate threshold that slightly increases false positives to maximize early recall (82.27). PhyMPGN attains the highest G-mean, suggesting it maintains a higher true negative rate, whereas MentorPDM—benefiting from the dense sampling—edges on PR-AUC.

These results confirm that our bidirectional mechanism generalizes across sampling regimes: when data are scarce it relies on physics to stabilize inference; when data are dense it leverages physics to regularize and avoid overfitting, still delivering the best overall discrimination.

Ablation Study

Removing any component degrades performance, but the largest drop appears when the CDM constraint is removed (F1 6.9% on Heavy-haul), together with noticeable decreases in ROC-AUC and MCC. This indicates that purely data-driven method cannot capture the nonlinear progression of rail damage, and the physics prior is essential to keep the trajectory of micro-states physically plausible. Discarding the maximum-entropy inference yields a moderate decline (F1 3.4%). Removing the *Env. term* leads to a smaller yet consistent degeneration, particularly on Heavy-haul where temperature–humidity variability is higher, confirming that thermo-corrosion effects modulate the damage rate.

Robustness Evaluation

To evaluate the robustness of PIBI under sparse detection scenarios, this section adopts the KW51 dataset as the benchmark and systematically increases its sampling interval $\Delta t \in \{Raw, 1, 3, 7, 15, 30\}$ days. For each Δt, we compare PIBI, Data-based (HCAN), and Physics-based (PhyMPGN) models under the same training protocol.

Fig. 3(a) reports the Consistency Gap (CG)—a standardized forward-backward discrepancy that quantifies the cycle from observation to latent degraded representation and back, by computing the normalized reconstruction difference. CG grows with Δt; however, PIBI demonstrates the smallest increase, confirming that the bidirectional physics constraints reconstruct the degeneration process of the observed entities. Fig. 3(b) illustrates that PR-AUC consistently declines as Δt increases, yet PIBI exhibits the slowest degeneration.

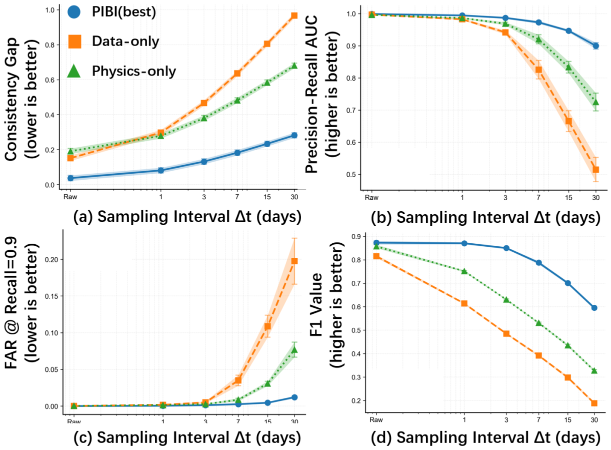

Figure 3: Robustness to Sparse Sampling (Δt) on KW51.

In Fig. 3(c), data-driven approaches incur a rapid surge in false alarms, whereas PIBI maintains better control over False Alarm Rate (FAR).

Collectively, these results demonstrate that as temporal resolution decreases, PIBI not only preserves higher classification quality but also delivers reliable self-diagnostic signals (consistency gap and coverage), thereby enhancing model robustness under sparse observation conditions—an advantage of significance for real-world industrial.

Conclusion

This work introduces a Physics-Guided Bidirectional Inference framework that bridges the long-standing micro–macro observability gap in industrial predictive maintenance. By coupling a Continuum Damage Mechanics simulator with a maximum-entropy inverse module, the model reconstructs latent micro-damage trajectories and projects them forward under environmental stressors. Extensive evaluation on heavy-haul railway and bridge datasets reveals consistent gains over physics-informed and data-centric baselines across eight metrics, demonstrating earlier warnings, calibrated uncertainty, and robustness to sparse sampling. Ablation studies confirm complementary roles of CDM priors, entropy-based inference and Arrhenius kinetics, while qualitative analyses highlight graceful degeneration as observation intervals widen, establishing a new paradigm for safety-critical infrastructure monitoring.

Acknowledgments

This work is supported by the grants from the National Natural Science Foundation of China (No.62272398) and Sichuan Science and Technology Program (No.2024NSFJQ0019).

References

Benkabou, S.; Benabdeslem, K.; Mansouri, D. E. K.; Chaib, S.; Mesmoudi, A.; and Hadjali, A. 2025. A temporal dependency preserving approach for anomaly detection on multivariate time series. *World Wide Web (WWW)*, 28(3): 37.

Celina, M.; Linde, E.; Brunson, D.; Quintana, A.; and Giron, N. 2019. Overview of accelerated aging and polymer degradation kinetics for combined radiation-thermal environments. *Polymer Degradation and Stability*, 166: 53–78.

Chen, H.; Liu, Y.; and Sun, H. 2025. PINP: Physics-Informed Neural Predictor with latent estimation of fluid flows. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28*.

Chen, L.; Fang, J.; Yu, Z.; Tong, Y.; Cao, S.; and Wang, L. 2023a. A Data-driven Region Generation Framework for Spatiotemporal Transportation Service Management. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, 3842–3854. ACM.

Chen, Y.; Chen, Z.; Guo, S.; Zhao, Y.; Liu, Z.; Wu, P.; Yang, C.; Li, Z.; and Yu, H. 2023b. Efficient training of large-scale industrial fault diagnostic models through federated opportunistic block dropout. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 15485–15493.

Chen, Z.; Luo, X.; Wang, W.; Zhao, Z.; Su, F.; and Men, A. 2025. Filter or compensate: Towards invariant representation from distribution shift for anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 2420–2428.

Dai, Z.; He, L.; Yang, S.; and Leeke, M. 2024. Sarad: Spatial association-aware anomaly detection and diagnosis for multivariate time series. *Advances in Neural Information Processing Systems*, 37: 48371–48410.

Fang, Q.; Su, Q.; Lv, W.; Xu, W.; and Yu, J. 2025. Boosting Fine-Grained Visual Anomaly Detection with Coarse-Knowledge-Aware Adversarial Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 16532–16540.

Harte, J.; Brush, M.; Umemura, K.; Muralikrishnan, P.; and Newman, E. A. 2024. Dynamical theory of complex systems with two-way micro–macro causation. *Proceedings of the National Academy of Sciences*, 121(50): e2408676121.

Ho, T. K. K.; Karami, A.; and Armanfard, N. 2025. Graph anomaly detection in time series: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Liao, X.; Chen, S.; Wen, P.; and Zhao, S. 2023. Remaining useful life with self-attention assisted physics-informed neural network. *Advanced Engineering Informatics*, 58: 102195.

Lohia, P.; Boué, L.; Ranganath, S.; and Agneeswaran, V. 2024. High Significant Fault Detection in Azure Core Workload Insights. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 22779–22787.

Lu, N.; Huang, S.; Li, Y.; Jiang, B.; Kaynak, O.; and Zio, E. 2024. Dynamic weight-based accelerated test modeling for fault degradation and lifetime analysis. *Reliab. Eng. Syst. Saf.*, 252: 110405.

Maes, K.; and Lombaert, G. 2021. Monitoring railway bridge KW51 before, during, and after retrofitting. *Journal of Bridge Engineering*, 26(3): 04721001.

Sha, Y.; Gou, S.; Liu, B.; Faber, J.; Liu, N.; Schramm, S.; Stoecker, H.; Steckenreiter, T.; Vnucec, D.; Wetzstein, N.; et al. 2024. Hierarchical knowledge guided fault intensity diagnosis of complex industrial systems. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5657–5668.

Sun, Y.; Xie, Z.; Chen, D.; Eldele, E.; and Hu, Q. 2025. Hierarchical classification auxiliary network for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 20743–20751.

Tian, J.; Liang, Y.; Xu, R.; Chen, P.; Guo, C.; Zhou, A.; Pan, L.; Rao, Z.; and Yang, B. 2025. Air Quality Prediction with Physics-Guided Dual Neural ODEs in Open Systems. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*.

Voyiadjis, G. Z.; and Kattan, P. I. 2025. Interaction of damage processes in continuum damage mechanics. *Acta Mechanica*, 236(6): 3565–3585.

Wang, Y.; Xu, Y.; Yang, J.; Wu, M.; Li, X.; Xie, L.; and Chen, Z. 2024. Graph-aware contrasting for multivariate time-series classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 15725–15734.

Xia, T.; Xing, X.; Yan, T.; Wang, D.; Pan, E.; and Xi, L. 2024. Interpretable temporal degradation state chain based fusion graph for intelligent bearing fault detection. *Adv. Eng. Informatics*, 59: 102342.

Zeng, B.; Wang, Q.; Yan, M.; Liu, Y.; Chengze, R.; Zhang, Y.; Liu, H.; Wang, Z.; and Sun, H. 2025. PhyMPGN: Physics-encoded Message Passing Graph Network for spatiotemporal PDE systems. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*.

Zhang, S.; Wang, T.; Adams, S.; Bhattacharya, S.; Tiyyagura, S. R.; Bowen, E.; Veeramani, B.; and Zhou, D. 2025. MentorPDM: Learning Data-Driven Curriculum for Multi-Modal Predictive Maintenance. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, 2837–2847.