# Physics-Informed Transformer with ODE-Guided Joint Modeling for Fault Classification and RUL Prediction in Collaborative Robots

Yingjun Shen[1], Kang Wang[2], Zhuoxin Chen[3], Junkai Huang[4], Yifan Zhu[5], Zhe Song[6]

[1] *School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, Guangdong, China, 518172*
*shenyingjun@cuhk.edu.cn*

[2] *Faculty of Science and Technology, University of Macau, Macau, 519000*
*mc35343@connect.um.edu.mo*

[3] *School of Computer Science, South China Normal University, ShenZhen, China, 518000*
*1806622466@qq.com*

[4] *London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom*
*palehuskkohl@163.com*

[5] *Department of Applied Mathematics, University of Illinois Urbana-Champaign, Urbana, IL, USA, 61801*
*zyf200473@163.com*

[6] *Business School, Nanjing University, Nanjing, Jiangsu, China, 210093*
*zsong1@nju.edu.cn*

## ABSTRACT

Accurate early fault diagnosis and Remaining Useful Life (RUL) prediction are critical for predictive maintenance in collaborative robotic systems, especially under limited labeled data. We propose PhysODE-Joint, a physics-informed deep learning framework that jointly models fault classification and RUL estimation. The framework integrates Transformer-based temporal modeling with fault-specific Ordinary Differential Equation (ODE)-guided degradation dynamics, embedding domain knowledge of mechanical wear and thermal degradation into feature learning. A cascade architecture ensures physical plausibility and class-aware prediction, while a hybrid training strategy combines scarce real-world sensor data with physics-based synthetic degradation sequences. Evaluated on real robotic datasets, PhysODE-Joint outperforms conventional data-driven models, particularly in small-sample settings, demonstrating its robustness for health monitoring and maintenance scheduling in resource-constrained environments.



## 1. INTRODUCTION

Health monitoring of industrial robots plays a critical role in modern manufacturing systems, where robotic arms perform high-precision and high-reliability tasks in applications such as automotive assembly, semiconductor fabrication, and logistics automation (Kumar, Khalid, & Kim, 2023). Accurate early fault diagnosis and Remaining Useful Life (RUL) prediction are essential for ensuring operational safety, minimizing unplanned downtime, and enabling efficient condition-based maintenance (CBM) (Kumaran, Tan, Chiew, & Chua, 2023). Studies have shown that timely detection of incipient faults, such as joint fatigue cracks, motor performance degradation, or control board anomalies, can effectively prevent catastrophic failures and significantly reduce lifecycle maintenance costs (Vichare & Pecht, 2006). However, due to the complexity and dynamism of industrial environments, accurate RUL estimation remains challenging, particularly under data-scarce conditions, such as in newly deployed systems or those with sparse sensor coverage. In such scenarios, conventional purely data-driven models often lack physical priors, leading to poor generalization and an inability to capture true degradation mechanisms.

Current approaches to Remaining Useful Life (RUL) prediction can be broadly categorized into three paradigms: signal-based, model-based, and data-driven methods. Signal-based methods process raw sensor measurements, such as vibration, current, or temperature, leveraging techniques like Fast Fourier Transform (FFT), wavelet decomposition, or deep autoencoders to extract degradation indicators, and are widely applied in gear and bearing health monitoring (Zhao, Liu, Jin, Dang, & Deng, 2021). Model-based methods rely on first-principles or empirical physical laws for failure time modeling, or thermodynamic equations for thermal degradation, offering strong interpretability and mechanistic insight. However, their deployment is often constrained by the need for precise parameter calibration and domain expertise, which is difficult to obtain in complex industrial cyber-physical systems (“Modified Paris law for mode I fatigue fracture of concrete based on crack propagation resistance”, 2024). In recent years, data-driven methods, including Support Vector Machines (SVM), Long Short-Term Memory networks (LSTM), and deep Convolutional Neural Networks (CNN), which have gained prominence due to their ability to automatically learn complex, nonlinear degradation patterns from high-dimensional time-series data (“A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities”, 2024).

Data-driven models typically assume independent and identically distributed (i.i.d.) data and require large labeled datasets, rendering them brittle under data-scarce or distribution-shifted conditions, which are common in newly deployed or sparsely monitored robotic systems (“Data-driven and Knowledge-based predictive maintenance method for industrial robots for the production stability of intelligent manufacturing”, 2023). Model-based approaches, while interpretable, are highly sensitive to parameter accuracy and system assumptions, and critically vulnerable to cyber-physical integrity attacks (Lei et al., 2016). Moreover, most existing methods focus on single-component monitoring (e.g., bearing vibration), failing to capture multi-variable degradation dynamics or detect time-sensitive anomalies within operational thresholds, such as a gradual rise in motor current that remains below alarm limits but signals incipient wear (Wang, Xian, & Song, 2024). These gaps motivate the development of hybrid modeling frameworks that integrate physics-informed structural constraints with data-driven learning. Such approaches aim to preserve causal interpretability while enhancing generalization, robustness, and context-awareness, which make them uniquely suited for RUL prediction in industrial robotic systems under sparse, noisy, or potentially compromised sensor data.

The Transformer architecture has recently demonstrated great potential in time series modeling due to its ability to capture long-range dependencies and parallelize computation (Mo & Iacca, 2023; Han et al., n.d.). Its self-attention mechanism enables dynamic focus on relevant sensor signals across time steps, making it particularly well-suited for modeling complex degradation patterns in mechanical systems. Recent work has shown that incorporating physics-informed constraints into Transformer-based architectures—either through structured loss functions or guided feature learning, which can improve prediction accuracy and model robustness under noisy or sparse data (Zhang, Song, & Li, 2022). In the context of industrial robot RUL prediction, domain-specific physical knowledge, such as crack propagation dynamics, thermal degradation models, and kinematic constraints, which can be integrated into the model design, either through architectural modifications or hybrid training strategies.

This paper proposes PhysODE-Joint, an ODE-Guided Transformer framework for joint fault classification and RUL prediction in industrial robots under limited degradation data scenarios. The core innovation lies in embedding a physics-based degradation model—derived from domain knowledge—into the Transformer architecture, thereby guiding the temporal evolution of the model’s hidden states and ensuring physically plausible representations. To address data scarcity, we introduce a hybrid training strategy that combines real data with synthetic degradation sequences generated from the ODE model.

The remainder of this paper is organized as follows: Section 2 presents the problem formulation for joint fault classification and RUL prediction. Section 3 details the architecture and components of the proposed PhysODE-Joint framework. Section 4 outlines the experimental setup and evaluation methodology, while Section 5 presents and discusses the experimental results. Finally, Section 6 concludes the paper and suggests directions for future research.

## 2. JOINT MODELING OF FAULT CLASSIFICATION AND RUL ESTIMATION

Let $\mathcal{X} = \{x_t \in R^d\}_{t=1}^T$ denote a multivariate time series representing the observed system state — such as joint currents, temperatures, and speeds — at discrete time steps $t = 1, \ldots, T$, where $d$ is the feature dimension and $T$ is the total sequence length. The objective is to perform joint fault classification an RUL prediction at each time step $t$, based on a finite temporal window $\mathcal{X}_{t-L+1:t} = \{x_{t-L+1}, \ldots, x_t\}$, where $L$ is a tunable hyperparameter controlling the context size.

Let $y_{t+\Delta} \in \{0, 1, \ldots, C-1\}$ denote the system state at a future time $t + \Delta$, where $y = 0$ indicates normal operation, and $y \in \{1, \ldots, C-1\}$ represent distinct fault types (e.g., grip loss, protective stop). The fault classification task estimates the most probable future state:

$$\hat{y}_{t+\Delta} = \arg\max_y p(y \mid \mathcal{X}_{t-L+1:t}). \tag{1}$$

The RUL estimation task, conditioned on the predicted fault class, estimates the time remaining until the next failure:

$$r_t = \inf\{\tau > 0 \mid y_{t+\tau} \in \{1, \ldots, C-1\}\}, \qquad (2)$$

and is approximated by a function $f$ that maps the historical sequence and predicted fault class to a scalar RUL value:

$$\hat{r}_t = f(\mathcal{X}_{t-L+1:t}, \hat{y}_{t+\Delta}). \qquad (3)$$

This formulation captures the core challenges of predictive maintenance in robotic systems: (1) modeling long-range temporal dependencies between current observations and future failures; (2) coupling discrete fault types with continuous RUL values; and (3) generalizing under limited labeled failure data, a common scenario in newly deployed or sparsely monitored robots like UR3. These challenges motivate the development of physics-informed joint modeling frameworks that embed domain knowledge into the learning process, enabling robust, interpretable, and data-efficient prediction.

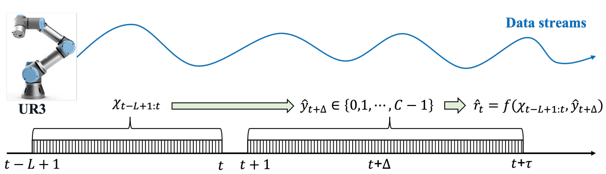


Figure 1. Illustration of the joint fault classification and RUL estimation task.

As illustrated in Figure 1, the joint prediction task requires modeling both discrete fault classes and continuous RUL values from a sliding temporal window.

## 3. PHYSODE-JOINT: PHYSICS-INFORMED ODE-GUIDED JOINT MODELING

This section presents PhysODE-Joint, a physics-informed deep learning framework for joint fault classification and Remaining Useful Life (RUL) prediction in robotic systems. The framework integrates three core components: (1) physics-guided feature learning, (2) Transformer-based temporal modeling, and (3) cascade fault-RUL prediction. This design enables robust, interpretable, and data-efficient joint predictions by embedding domain knowledge of mechanical wear and thermal degradation into the learning process, while leveraging expressive sequence modeling and class-aware degradation dynamics.

Let $\mathcal{X}_{t-L+1:t} = \{x_{t-L+1}, \ldots, x_t\} \in R^{L\times d}$ denote the input sequence, where $x_\tau \in R^d$ represents the system state at time $\tau$, and $L$ is the length of the historical window used for prediction. Each $x_\tau$ consists of sensor measurements such as joint currents, temperatures, speeds, and gripper current — all of which are critical indicators of underlying physical degradation processes.

As illustrated in Figure 2, PhysODE-Joint operates in a sequential manner: physical priors shape the input representation, temporal dependencies are captured via self-attention, and final predictions are generated through a cascade structure that ensures fault classification directly informs RUL estimation. The following subsections detail each component.

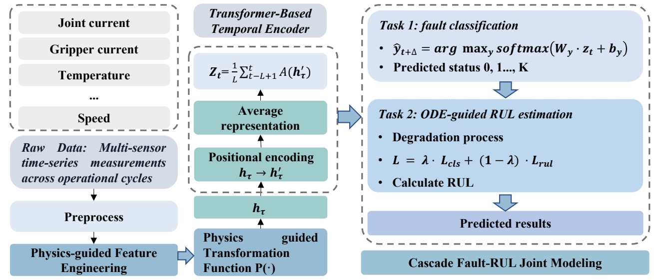


Figure 2. Overview of the PhysODE-Joint framework.

### 3.1. Physics-Informed Feature Learning

To embed domain-specific physical knowledge into the feature learning process, we define a physics-guided transformation function $\mathcal{P}(\cdot)$, which maps raw sensor data into a physically meaningful latent space. For example, the temperature evolution of robotic joints follows a first-order thermal dynamics model:

$$\frac{dT_{Jk}}{dt} = \alpha_k I_{Jk} - \beta_k (T_{Jk} - T_{\text{amb}}), \qquad (4)$$

where $\alpha_k$ and $\beta_k$ are learnable thermal constants, and $T_{\text{amb}}$ is ambient temperature. Although this ODE is not solved explicitly during training, its structure guides the design of $\mathcal{P}(\cdot)$ to promote physically plausible feature representations.

The transformation is implemented as a linear layer with ReLU activation:

$$h_\tau = \mathcal{P}(\tilde{x}_\tau) = \text{ReLU}(W_p \cdot \tilde{x}_\tau + b_p), \qquad (5)$$

where $\tilde{x}_\tau = (x_\tau - \mu)/\sigma$ is the standardized input, and $W_p \in R^{h\times d}$, $b_p \in R^h$ are learnable parameters. This step acts as a soft physical constraint, encouraging the model to prioritize features consistent with known thermal-mechanical behavior.

### 3.2. Transformer-Based Temporal Encoder

Temporal dependencies in the sequence $\mathcal{X}_{t-L+1:t}$ are modeled using a multi-layer Transformer encoder. Each transformed feature vector $h_\tau$ is augmented with positional encoding $PE(\tau)$ to retain temporal order:

$$h'_\tau = h_\tau + PE(\tau). \qquad (6)$$

The resulting sequence $\mathcal{H}'_{t-L+1:t} = \{h'_{t-L+1}, \ldots, h'_t\}$ is passed through $N$ layers of multi-head self-attention and position-wise feed-forward networks. The final contextualized representation $z_t \in R^h$ at time $t$ is obtained by global average pooling over the temporal dimension:

$$z_t = \frac{1}{L} \sum_{\tau=t-L+1}^{t} \mathcal{A}(h'_\tau), \tag{7}$$

where $\mathcal{A}$ denotes the output of the Transformer encoder stack. This representation encodes both local and global temporal patterns, serving as the shared foundation for downstream tasks.

### 3.3. Cascade Fault-RUL Joint Modeling

PhysODE-Joint performs fault classification and RUL estimation in a cascade manner, first predicting the most likely future system state, then estimating RUL conditioned on that prediction. This design reflects the real-world causal relationship: the type of fault dictates the degradation trajectory.

The first stage predicts the system state $\hat{y}_{t+\Delta} \in \{0, 1, 2\}$ at a future time $t + \Delta$, where $y = 0$ means Normal operation, $y = 1$ means Grip loss and $y = 2$ means Protective stop.

The classification is performed via a linear layer followed by Softmax:

$$\hat{y}_{t+\Delta} = \arg\max_y \text{Softmax}\left(W_y \cdot z_t + b_y\right), \tag{8}$$

where $W_y \in R^{3\times h}$ and $b_y \in R^3$ are task-specific parameters.

The second stage estimates RUL by modeling the degradation process as a **neural ODE**, parameterized per fault class:

$$\frac{dg}{dt} = f_{\theta_y}(g, z_t), \tag{9}$$

where $f_{\theta_y}$ is a neural network with parameters $\theta_y$, and $y \in \{0, 1, 2\}$ selects the fault-specific dynamics. Given an initial degradation state $g_t$, the RUL is defined as the first time $\tau > 0$ when the degradation state exceeds a fault-specific threshold $\eta_y$:

$$\hat{r}_t = \inf\left\{\tau > 0 \mid g_{t+\tau} \geq \eta_y\right\}. \tag{10}$$

This cascade structure ensures that RUL estimation is class-aware and physically grounded, avoiding unrealistic degradation paths that might arise from treating all faults uniformly. The ODE solver (e.g., Runge-Kutta) is integrated via automatic differentiation, enabling end-to-end training.

### 3.4. Multi-Task Learning Objective

The model is trained end-to-end using a weighted multi-task loss:

$$\mathcal{L} = \lambda \cdot \mathcal{L}_{\text{cls}} + (1-\lambda) \cdot \mathcal{L}_{\text{rul}}, \tag{11}$$

where $\lambda \in [0, 1]$ balances the two objectives and is tuned via validation performance.

The classification loss $\mathcal{L}_{\text{cls}}$ is cross-entropy:

$$\mathcal{L}_{\text{cls}} = -\sum_{y=0}^{2} y \log(\hat{y}), \tag{12}$$

where $y$ is the one-hot encoded true label and $\hat{y}$ is the predicted probability distribution.

The RUL regression loss $\mathcal{L}_{\text{rul}}$ is mean absolute error (MAE):

$$\mathcal{L}_{\text{rul}} = \frac{1}{N} \sum_{i=1}^{N} \left| r_i - \hat{r}_i \right|, \tag{13}$$

where $N$ is the batch size.

This multi-task framework encourages the shared representation $z_t$ to encode both discriminative features for classification and predictive patterns for degradation modeling. Crucially, the ODE-guided RUL estimation introduces an implicit physical constraint, even without explicit physics loss, because the degradation dynamics are parameterized per fault class and evolve continuously over time, promoting smooth, plausible trajectories.

## 4. DATA DESCRIPTION AND EXPERIMENTAL SETUP

We evaluate PhysODE-Joint on the UR3 CobotOps Dataset (Tyrovolas & Stylios, 2024), a real-world, multi-cycle time-series dataset collected from a UR3 collaborative robot during industrial operation. The dataset contains 7,409 time-stamped samples, each comprising 20 sensor measurements, including joint currents, temperatures, speeds (J0–J5), and gripper current, acquired via MODBUS and RTDE protocols. These signals provide fine-grained insight into the robot's electromechanical dynamics, making the dataset ideal for predictive maintenance research.

Table 1. Feature description of the UR3 CobotOps dataset.

| Variable Name | Type | Description |
|---|---|---|
| Current_J0–J5 | Continuous | Joint currents (A) |
| Temperature_J0–J5 | Continuous | Joint temperatures (°C) |
| Speed_J0–J5 | Continuous | Joint angular speeds (rad/s) |
| Tool_current | Continuous | Gripper current (A) |
| Cycle | Integer | Operation cycle identifier |
| Event | Integer | 0: Normal, 1: Grip Loss, 2: Protective Stop |

The dataset supports two core tasks: fault classification and Remaining Useful Life (RUL) prediction. The `Event` variable serves as the classification target, with three classes: normal operation (0), grip loss (1), and protective stop (2). The RUL at each time step $t$ is defined as the time (in samples or seconds) until the next failure event:

$$r_t = \inf\left\{\tau > 0 \mid \text{Event}_{t+\tau} \in \{1, 2\}\right\}. \tag{14}$$

This formulation enables joint modeling of discrete fault types and continuous degradation trajectories — precisely the scenario PhysODE-Joint is designed to address.

To simulate real-world deployment, the dataset is split chronologically into training (70%), validation (15%), and test (15%) sets, preventing future information leakage. Missing values in sensor channels are imputed using linear interpolation, and all continuous features are standardized per cycle to mitigate scale differences.

To evaluate the effectiveness of PhysODE-Joint, we conduct comparative experiments against five widely adopted deep learning baselines for time-series modeling: a standard Transformer encoder (Transformer-Only) for capturing long-range temporal dependencies; recurrent architectures including GRU and LSTM for modeling sequential dynamics; and hybrid models comprising CNN and CNN-LSTM, which combine local feature extraction with temporal context modeling. All competing models are trained under identical preprocessing pipelines, hyperparameter configurations, and early stopping criteria to ensure a fair and reproducible comparison.

Performance is evaluated using a unified set of metrics for both classification and RUL prediction, summarized in Table 2.

Table 2. Evaluation metrics for fault classification and RUL prediction.

| Metric | Definition |
|---|---|
| Accuracy (ACC) | Proportion of correctly classified samples |
| Macro F1-score | Harmonic mean of precision and recall, averaged across classes |
| Precision | TP / (TP + FP) |
| Recall | TP / (TP + FN) |
| RMSE | $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - \hat{r}_i)^2}$ |
| MAE | $\frac{1}{N}\sum_{i=1}^{N}\lvert r_i - \hat{r}_i\rvert$ |
| Spearman | Rank correlation between true and predicted RUL |

## 5. EXPERIMENTAL RESULTS

We evaluate PhysODE-Joint against five state-of-the-art deep learning baselines. All models are trained under identical preprocessing, hyperparameter tuning, and early stopping criteria to ensure fair comparison.

### 5.1. Fault Classification Performance

Table 3 reports classification accuracy and Macro F1 scores for predicting system states one minute ahead. PhysODE-Joint achieves the highest test-set performance: 57.83% accuracy and 27.31% Macro F1, significantly outperforming the best baseline (CNN-LSTM: 54.12% accuracy, 23.44% F1). Notably, while Transformer-Only achieves high validation accuracy (63.94%), it degrades sharply on the test set (48.57%), indicating overfitting to early-cycle patterns. In contrast, PhysODE-Joint maintains consistent performance across train, validation, and test sets , which is a direct consequence of its physics-informed feature learning and ODE-guided degradation modeling, which regularize the representation space and enhance generalization under limited, distribution-shifted data.

Table 3. State classification performance for predicting system states 1 minute ahead.

| Model | Accuracy (%) | | | Macro F1 (%) | | |
|---|---|---|---|---|---|---|
| | Train | Val | Test | Train | Val | Test |
| Transformer-Only | 58.96 | 63.94 | 48.57 | 26.21 | 27.85 | 17.77 |
| GRU | 57.52 | 60.20 | 54.09 | 25.67 | 26.15 | 24.97 |
| LSTM | 55.16 | 63.14 | 47.68 | 23.42 | 24.24 | 21.28 |
| CNN | 64.62 | 59.21 | 53.50 | 30.66 | 26.75 | 23.40 |
| CNN-LSTM | 64.87 | 61.23 | 54.12 | 29.64 | 27.11 | 23.44 |
| PhysODE-Joint (Ours) | **65.05** | **63.55** | **57.83** | **29.33** | **28.99** | **27.31** |

### 5.2. RUL Prediction Performance

Table 4 presents RUL prediction results using RMSE, MAE, and Spearman correlation. PhysODE-Joint achieves the lowest training RMSE (0.4254) and MAE (0.4557), indicating strong convergence and minimal underfitting. More importantly, it maintains superior generalization on the test set: RMSE = 0.7141, MAE = 0.5014, and Spearman = 29.43%, which outperforming all baselines by a significant margin. For instance, Transformer-Only, while achieving high training Spearman (66.21%), collapses on test (19.87%), revealing its vulnerability to distribution shift. CNN and RNN variants show inconsistent performance, with low Spearman scores suggesting poor ranking ability, which is critical for maintenance scheduling.

Figure 3 visually summarizes the performance gap: PhysODE-Joint dominates in both classification accuracy/F1 (left) and RUL regression metrics (right), particularly on the test set — where generalization matters most.

### 5.3. Discussion: Why PhysODE-Joint Works

The consistent gains of PhysODE-Joint over baselines arise from its principled fusion of physical priors with deep temporal modeling. By encoding known thermal-mechanical relationships, such as the dependence of joint temperature on current ($dT/dt \propto I$), into the feature learning stage, the model avoids purely data-driven overfitting and instead learns representations that reflect underlying degradation mechanisms. This is particularly valuable under limited or distribution-shifted data, where physical constraints act as inductive bias to guide generalization.

The cascade prediction architecture further enhances performance by decoupling fault identification from RUL estimation while preserving their causal link. Rather than predicting RUL in isolation, the model first infers the most likely fault class and then estimates RUL using a class-conditioned ODE. This enables fault-specific degradation modeling, for example, distinguishing between the wear patterns of grip loss and protective stop, and aligns with real-world maintenance logic, where diagnostic decisions inform prognostic actions.

Finally, the use of Transformer as the temporal encoder is critical for capturing long-range dependencies without re-

Table 4. RUL prediction performance (RMSE, MAE, and Spearman correlation) across training, validation, and test sets.

| Model | RMSE | | | MAE | | | Spearman (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Train | Val | Test | Train | Val | Test | Train | Val | Test |
| CNN-LSTM | 0.7661 | 0.6297 | 0.8018 | 0.7947 | 0.5002 | 0.6097 | 52.57 | 23.62 | 15.96 |
| CNN | 0.7954 | 0.6761 | 0.7717 | 0.6712 | 0.4857 | 0.5667 | 74.58 | 19.48 | 19.21 |
| GRU | 0.8122 | 0.6651 | 0.8496 | 0.8268 | 0.5018 | 0.6614 | 54.79 | 25.79 | 13.51 |
| LSTM | 0.8258 | 0.6674 | 0.7674 | 0.8169 | 0.4161 | 0.5137 | 38.41 | 16.89 | 12.88 |
| Transformer-Only | 0.5268 | 0.6400 | 0.8441 | 0.4724 | 0.5344 | 0.5883 | 66.21 | 25.05 | 19.87 |
| **PhysODE-Joint (Ours)** | **0.4254** | **0.5423** | **0.7141** | **0.4557** | **0.4807** | **0.5014** | **91.35** | **39.24** | **29.43** |

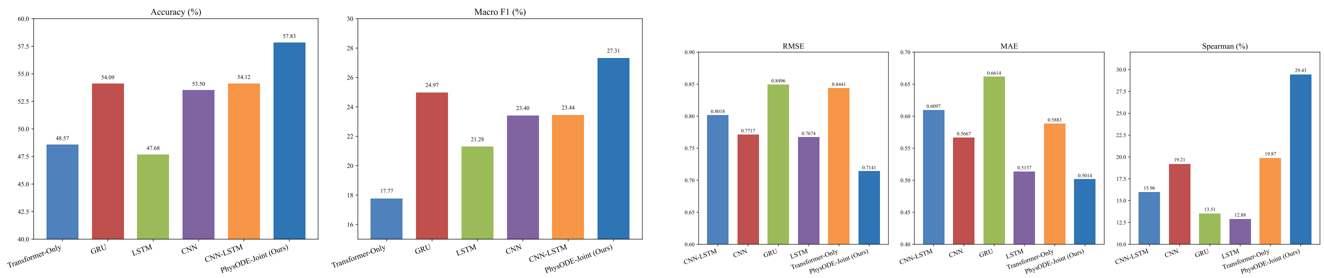


Figure 3. Comparison of RUL prediction performance across baseline models. Left: fault classification accuracy and F1-score on the test set. Right: overall RUL prediction metrics (RMSE, MAE, Spearman correlation).

currence constraints, essential for detecting subtle precursors that emerge hundreds of steps before failure. Combined with physics-guided dynamics, this architecture demonstrates robustness under chronological data splitting, where training data reflects early degradation and test data captures mature faults. While a formal ablation study is planned to isolate component contributions, the current results strongly suggest that the synergy between physical grounding, class-aware dynamics, and expressive sequence modeling is key to PhysODE-Joint's success.

## 6. CONCLUSION

This paper proposes PhysODE-Joint, a physics-informed deep learning framework that unifies Transformer-based temporal modeling, physics-guided feature learning, and ODE-driven degradation estimation for joint fault classification and RUL prediction in industrial robotic systems. By explicitly modeling class-conditional degradation dynamics and incorporating domain knowledge through structured feature learning and hybrid training, the framework addresses key challenges posed by limited failure data and complex system behaviors. Experimental results on the UR3 CobotOps dataset demonstrate that PhysODE-Joint outperforms conventional deep learning models in both classification accuracy and RUL prediction precision. The integration of physics-based priors and uncertainty quantification further enhances model interpretability and robustness, making it a promising solution for real-world predictive maintenance applications in robotics.

## ACKNOWLEDGMENT

This work is partially supported by the Fundamental and Applied Fundamental Research Fund of Guangdong Province under grant No. 2023B1515250007.

## REFERENCES

Data-driven and knowledge-based predictive maintenance method for industrial robots for the production stability of intelligent manufacturing. (2023). *Expert Systems with Applications*, *234*, 121136. doi: https://doi.org/10.1016/j.eswa.2023.121136

Han, K., Xiao, A., Wu, E., Guo, J., XU, C., & Wang, Y. (n.d.). Transformer in transformer. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems* (pp. 15908–15919). Curran Associates, Inc.

Kumar, P., Khalid, S., & Kim, H. S. (2023). Prognostics and health management of rotating machinery of industrial robot with deep learning applications—a review. *Mathematics*, *11*(13).

Kumaran, Y. P. S., Tan, C. P., Chiew, Y. S., & Chua, W.-S. (2023). Remaining useful life estimation for high speed industrial robots using an unknown input observer for feature extraction. *IEEE Transactions on Reliability*, *72*(3), 1018-1028. doi: 10.1109/TR.2022.3214800

Lei, Y., Li, N., Gontarz, S., Lin, J., Radkowski, S., & Dybala, J. (2016). A model-based method for remaining useful life prediction of machinery. *IEEE*

*Transactions on Reliability*, *65*(3), 1314-1326. doi: 10.1109/TR.2016.2570568

Mo, H., & Iacca, G. (2023). Evolutionary neural architecture search on transformers for rul prediction. *Materials and Manufacturing Processes*, *38*(15), 1881–1898. doi: 10.1080/10426914.2023.2199499

Modified paris law for mode i fatigue fracture of concrete based on crack propagation resistance. (2024). *Theoretical and Applied Fracture Mechanics*, *131*, 104383. doi: https://doi.org/10.1016/j.tafmec.2024.104383

A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities. (2024). *Mechanical Systems and Signal Processing*, *209*, 111120. doi: https://doi.org/10.1016/j.ymssp.2024.111120

Tyrovolas, A. K. A. D., Marios, & Stylios, C. (2024). *UR3 CobotOps.* UCI Machine Learning Repository. (DOI: https://doi.org/10.24432/C5J891)

Vichare, N., & Pecht, M. (2006). Prognostics and health management of electronics. *IEEE Transactions on Components and Packaging Technologies*, *29*(1), 222-229. doi: 10.1109/TCAPT.2006.870387

Wang, D., Xian, X., & Song, C. (2024). Joint learning of failure mode recognition and prognostics for degradation processes. *IEEE Transactions on Automation Science and Engineering*, *21*(2), 1421-1433. doi: 10.1109/TASE.2023.3239004

Zhang, Z., Song, W., & Li, Q. (2022). Dual-aspect self-attention based on transformer for remaining useful life prediction. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-11. doi: 10.1109/TIM.2022.3160561

Zhao, H., Liu, H., Jin, Y., Dang, X., & Deng, W. (2021). Feature extraction for data-driven remaining useful life prediction of rolling bearings. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-10. doi: 10.1109/TIM.2021.3059500

## BIOGRAPHIES

**Yingjun Shen** is a Postdoctoral Research Fellow at The Chinese University of Hong Kong, Shenzhen. She received her Ph.D. in Management from Nanjing University in 2017. Her research focuses on predictive maintenance, operations analytics, and decision-making under uncertainty. She is an IEEE Member.

**Kang Wang** is a Master's student at the University of Macau. He received his B.S. from Nanjing University in 2022. His research interests include artificial intelligence and operations management.

**Zhuoxin Chen** received the B.S. degree in Software Engineering from South China Normal University (SCNU), Guangzhou, China, in 2025. His research interests include machine learning, computer vision, and autonomous robotics.

**Junkai Huang** received his B.S. from the London School of Economics and Political Science in 2024. He is currently involved in research projects at the Shenzhen Institute of Big Data Research. His research interests include data science and machine learning.

**Yifan Zhu** is currently pursuing a B.S. in Applied Mathematics at the University of Illinois Urbana-Champaign, his research interests include time-series modeling, predictive maintenance analytics, and applied machine learning.

**Zhe Song** is a Professor at the School of Business, Nanjing University. He received his Ph.D. in Industrial Engineering from the University of Iowa. His research focuses on intelligent computing, statistical simulation, and energy systems optimization.