

A Calibration-Based Hybrid Transfer Learning Framework for RUL Prediction of Rolling Bearing Across Different Machines

Yafei Deng¹, Shichang Du¹, *Member, IEEE*, Dong Wang¹, *Member, IEEE*, Yiping Shao², and Delin Huang²

Abstract—The effective remaining useful life (RUL) prediction of rolling bearings could guarantee mechanical equipment reliability and stability. The hybrid physical and data-driven prognosis model (HPDM) is recently receiving increasing attention. However, HPDM approaches suffer from two significant challenges that limit their applicability to complex prognosis scenarios: 1) the reality gap between the simulation and measurement data and 2) the limited model generality to accommodate different working conditions and machines. From the perspective of leveraging physical model inference as “teachers” for the data-driven model (DDM), this article proposes a calibrated-based hybrid transfer learning framework to improve the data fidelity and model generality. First, a five-degree-of-freedom (5-DOF) dynamic model of rolling bearing is constructed. Comprehensively considering the crack and spall behaviors of degradation evolution, the physical model could provide various failure trajectories. Second, the particle filter-based calibration is proposed to retain the high fidelity of the physical simulation. Finally, a physics-informed Bayes deep dual network (PI-BDDN) is designed. The designed network fuses the physical calibrated simulation as augmented input space to learn representative prognosis features and makes the transfer learning process interpretable by combining the physical model parameters into adversarial learning to selectively identify the most informative knowledge for RUL prediction. The effectiveness of the proposed method is verified on two representative bearing datasets, and comparative results show the superiority of the proposed method on prediction accuracy and uncertainty quantification.

Index Terms—Adversarial learning, Bayesian deep learning (BDL), bearing dynamic model, particle filters, physics-informed machine learning.

Manuscript received 17 December 2022; revised 16 February 2023; accepted 2 March 2023. Date of publication 22 March 2023; date of current version 30 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 52275499; in part by the National Key Research and Development Program of China under Grant 2022YFF0605700; in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ22E050017; and in part by Ningbo City, Zhejiang Province, Major Scientific and Technological Breakthrough and “Unveiling and Commanding” Project under Grant 2021Z093. The Associate Editor coordinating the review process was Dr. Yang Song. (*Corresponding author: Shichang Du.*)

Yafei Deng, Shichang Du, and Dong Wang are with the State Key Laboratory of Mechanical System and Vibration and the Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: phoenixdyf@sjtu.edu.cn; lovbin@sjtu.edu.cn; dongwang4-c@sjtu.edu.cn).

Yiping Shao is with the College of Mechanical Engineering, Zhejiang University of Technology, Zhejiang 310014, China (e-mail: syp123gh@zjut.edu.cn).

Delin Huang is with the College of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai 201209, China (e-mail: huangdelin2009@163.com).

Digital Object Identifier 10.1109/TIM.2023.3260283

I. INTRODUCTION

REMAINING useful life (RUL) prediction has become a critical issue in the field of prognostics and health management (PHM) since it is significant to prevent unnecessary breakdowns and catastrophic accidents [1]. As a key component in mechanical systems, the rolling element bearing has been utilized in various rotating types of machinery under complex working conditions and became an important factor affecting the performance of mechanical systems, which is reported that over 40% of failures are related to bearings [2].

Therefore, there is an urgent need to develop accurate RUL prediction techniques for rolling bearings to improve the safety and reliability of mechanical systems. Various kinds of RUL prediction approach for rolling bearings have been proposed in the last decade, which could be divided into three categories: physics-based models (PBMs), data-driven models (DDM), and hybrid physical and data-driven models (HPDMs).

PBMs: They aim to reveal the rolling bearings’ degradation process by solving a set of equations derived from engineering and science knowledge. El-Thalji and Jantunen [3] proposed a dynamic model of wear evolution that considers the topographical and tribological changes over the lifetime to enhance the prognosis of rolling bearings. Behzad et al. [4] proposed a new Paris law-based crack growth model for two-stage degradation pattern of rolling bearings. Cui et al. [5] constructed a simulated degradation dataset of rolling bearings under different health stages, and the RUL was estimated by utilizing the similarity method. Zhang et al. [6] reviewed the current existing physics spall size estimation models and evaluated the prognostic performance with different sensor technologies.

Physics-based approaches could describe the degradation phenomena with explicit mathematic equations. However, it requires expert knowledge and repetitive experiments, leading to substantial economic costs and human labor. As a result, the widespread deployment of physics-based approaches in complex mechanical systems has been limited.

DDM: They aim at learning the hidden relationship from condition monitoring data, which can be further divided into statistics-based DDM and machine learning-based DDM. For the statistics-based DDM, Gebrael et al. [7], [8] first established a Bayesian-based approach to develop a closed-form RUL distribution for rolling bearings. Pang et al. [9] employed

the Bayesian inference to update the posterior distributions, which fused the condition monitoring data and accelerated degradation data. Ye et al. [10] proposed a random effect Wiener process model with measurement errors, in which the recursive joint distribution filtering was employed to estimate the degradation rate. For the machine learning-based DDM, Mao et al. [11] combined the Hilbert–Huang transform and contractive denoising autoencoder to extract deep representations. Huang et al. [12] proposed a deep convolutional neural network with multilayer perceptron (DCNN–MLP) to mine information simultaneously in time series and image-based features. Peng et al. [13] proposed a Bayesian deep learning (BDL) framework based on the variational inference method, which provides uncertainty quantification for bearing RUL prediction. In addition, deep transfer learning (DTL) gains more attention since it can endow the model with excellent generalization performance by adding well-designed modules into the backbone network, such as distance-based metrics [14], [15], adversarial loss [16], [17], and weighted pretraining [18], [19].

One crucial prerequisite of effective DDM is the availability of representative datasets. In real applications, failures rarely occur, and maintenance is always performed before severe faults; thus, the obtained datasets mostly have a small number of failure samples without all potential degradation modes. Consequently, the application of DDMs has been generally hindered.

HPDMs: They are hypothesized that the HPDM model can potentially lead to performance improvement by leveraging the advantages of each and avoid the limitations when applied in isolation.

Data-Driven-Assisted Physical Model: Yan et al. [20] built the physical model with Weibull failure rate function (WFRF), and a support vector machine (SVM) was employed to facilitate degradation stage classification. Wang et al. [21] proposed an integrated RUL prediction model for wind turbine bearing, in which the parameters in the physical degradation model were obtained through the Bayesian framework. Qin et al. [22] proposed a digital twin model: vibration signal was modeled through the dynamic model and a cycle generative adversarial network was designed to bridge the reality gap.

Physics-Informed (PI) Machine Learning: Yucsan and Viana [23] proposed a PI recurrent neural network (PI-RNN) for main bearing fatigue prognosis, in which the grease degradation model was embedded into the long short-term memory (LSTM) cell as physical nodes. Deng et al. [24] proposed a PI-temporal convolutional network (PI-TCN) to describe the bearing stiffness degradation process, in which the physics rules of stiffness changing were embedded into the network layers and loss functions.

While HPDM approaches have shown promising prognosis potential, several pending issues still need to be adequately addressed. The major research gaps in existing studies are summarized in the following aspects.

- 1) *The Data Reality Gap Between the Virtual Space and the Real World:* Although the synthetic data from the physical model could enrich the representativeness of

training dataset and provide model interpretability, various domain gaps between simulation and real samples could not be ignored. The biased simulation data would mislead the DDMs to learn irrelevant details, and the performance would be even worse than the model applied in isolation.

- 2) *The Limited Model Generalizability Across Domains:* Existing HPDM approaches mainly assume that predict the RUL within a fixed operation and identical machine, which may be inaccessible in real-world industries. The incompleteness of cross-domain prognosis knowledge transfer limits the applicability. Moreover, scientific insights such as which part of information should be reused and transferred are not revealed comprehensively.

This article proposes a novel calibration-based hybrid transfer learning framework to handle the above challenging issues. The key idea of the proposed approach is that the physical model could be leveraged as a teacher for the DDM, in which the physical virtual sensors are utilized to enhance the input space and the inference parameters are employed to transfer the most informative prognosis knowledge selectively. Specifically, a five-degree-of-freedom (5-DOF) bearing dynamic model is developed to generate massive simulation failure data. Meanwhile, the calibration is performed to bridge the reality gap with reasonable computation cost. Subsequently, calibrated simulation and measurement signal is further combined as augmented input space for deep Bayes learning network to learn representative features for RUL prediction with uncertainty quantification. Moreover, the physical model inference parameters notably collaborate with adversarial learning to adaptively determine which part of information should be activated for the current transfer task. Finally, the performance of the proposed method is validated on two representative bearing datasets.

The main contributions are summarized as follows.

- 1) A novel hybrid strategy is designed to leverage physical inference as guidance for the DDM. A bearing lifecycle dynamic model that comprehensively describes the crack and spall degradation is constructed to provide sufficient information for the DDM, without the requirement of auxiliary data such as downtime inspection or additional sensor signals.
- 2) A particle filter-based dynamic calibration is designed to achieve both reality gap elimination and unobservable parameter estimation. The proposed iterative updating strategy provides the high fidelity of simulation data while ensuring a timely response with the DDM.
- 3) A PI transfer learning model is designed based on the physics-inferred system parameters, which provides theoretical guarantees to selectively share the most informative prognosis knowledge according to the current task. The proposed PI module could effectively promote the model's generalizability on different domains without sacrificing interpretability.

The rest of this study is organized as follows. In Section II, the theoretical foundation is formally introduced. In Section III, details of the proposed method are described.

In Section IV, case studies and implementation details are given. In Section IV-F, comparative results and ablation studies are provided. The conclusion and future works are presented in Section V.

II. THEORETICAL FOUNDATION

A. Bearing Dynamic Models

Dynamic models provide a time-domain transient simulation of the vibration response, which could be calculated by solving differential equations of each part's motions in the rolling element bearing. Fukata et al. [25] first proposed a two-degree-of-freedom (DOFs) model to theoretically describe the bearing vibration response, and the nonlinear behavior was modeled through a mass–spring–damping system

$$m \frac{d^2}{dt^2} \begin{bmatrix} x \\ y \end{bmatrix} + c \frac{d}{dt} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} W_x \\ W_y \end{bmatrix} \quad (1)$$

where m is the mass of the inner race and the rotor, c is the equivalent viscous damping, W_x and W_y are radial force components loaded on the rotor shaft, and F_x and F_y are nonlinear contact force components, which could be calculated based on the Hertzian contact deformation theory.

Based on this kind of lumped-parameter model, many improvements have been achieved for different research purposes [26]. Generally, the unified form of the rolling bearing lumped-parameter model can be written as

$$\hat{X}(t) = \mathcal{F}[\theta(t)], \quad t \in [0, T] \quad (2)$$

where $\hat{X}(t)$ indicates the simulated signals, $\mathcal{F}[\cdot]$ represents a series of nonlinear differential operators, and $\theta(t)$ indicates the system parameters, including static structure information and time-varying properties of rolling bearings. In order to solve the above dynamic models, techniques with different model complexity and computational requirements have been developed, including MATLAB Simulink, bond graph model (SYMBOLS), multibody dynamics model (ADAMS), and finite-element model (FEM) [27].

B. Bayesian Calibration

It refers to an iterative process of updating the uncertainty distributions of the model's internal parameters to ensure that the simulated output of the physical model is consistent with observations. Compared with the way that directly minimizes the difference between observations $X(t)$ and simulation $\hat{X}(t)$ through multivariate optimization or machine learning method, the Bayesian calibration approach determines the most likely uncertainties for input parameters distribution $p(\theta(t)|X(t))$ to ensure the output following as $\hat{X}(t) \sim X(t)$, which retains the uncertainty quantifications inherent in the calibration parameters. Generally, the calibration could be seen as $\theta(t) = \mathcal{F}^{-1}[X(t)]$.

In order to provide a time-efficient solution, the sequential Bayesian technique, such as particle filters [28], has been developed, and the latent parameters of the physical model can be inferred through computing maximum a posteriori (MAP) estimation, which is given as [29]

$$\hat{\theta}(t) = \operatorname{argmax}_{\theta} p(\theta(t)|X(t)) \quad (3)$$

where $\hat{\theta}(t)$ indicates calibrated parameters and $p(\theta(t)|X(t))$ denotes the conditional probability distribution of parameters.

Benefitting from the sequential Bayesian calibration leveraging point value estimations with maximum likelihood rather than computing the whole distribution, a satisfactory balance between computational complexity and simulation fidelity could be achieved for the dynamic model calibration process. Consequently, the run time of calibration is shorter than the time interval between the acquisitions of new measurement data. Meanwhile, there are sufficient data points to characterize the unobservable parameter space.

C. Adversarial Transfer Learning

The adversarial-based transfer learning approach mainly achieves domain adaptation through a maximum–minimum two-player game between a feature generator $G(\cdot)$ and a domain discriminator $D(\cdot)$, which can be described as [30]

$$V(G, D) = \mathbb{E}_{x \sim \mu_{\text{data}}} [\log D(G(x))] + \mathbb{E}_{z \sim \mu_{\text{noise}}} [\log(1 - D(G(z)))] \quad (4)$$

where $G(x)$ and $G(z)$ denote the extracted feature by $G(\cdot)$ and u_{data} and u_{noise} denote the distributions of real data and the noise, respectively. The discriminator adjusts its parameters to ensure that the generated $G(z)$ from latent space can be distinguished from the generated source feature $G(x)$. Conversely, the generator is designed to confuse the discriminator by generating fake samples with similar distribution to the source domain.

The alternating optimization process could be formulated as

$$\begin{cases} \nabla_{\theta_G} = \frac{1}{b} \nabla_{\theta_G} \sum_{i=1}^b \log(1 - D(G(z^i))) \\ \nabla_{\theta_D} = \frac{1}{b} \nabla_{\theta_D} \sum_{i=1}^b [\log D(x^i) + \log(1 - D(G(z^i)))] \end{cases} \quad (5)$$

where b denotes the batch size, ∇_{θ_G} denotes a gradient descent step on the generator parameters θ_G , and ∇_{θ_D} denotes a gradient ascent step on the discriminator parameters.

For the prognosis issue of mechanical systems, the adversarial-based approach is utilized to promote the generality of the predictive model to cover the divergence of working conditions and the variation within component family types.

III. PROPOSED METHOD

The overall procedures of the proposed hybrid prognosis framework are shown in Fig. 1. The framework is divided into three steps and each step is briefly introduced as follows.

- 1) Data acquisition in both real space and virtual space. For the virtual space, a 5-DOF dynamic model is constructed to generate the physical simulated dataset \hat{D}_S and \hat{D}_T .
- 2) Calibrate the physical model parameters $\hat{\theta}$ with particle filters, which further pushes the distribution of simulations $P(\hat{X}(t))$ to be consistent with measurements $P(X(t))$.
- 3) Transfer the prognosis knowledge based on the PI-BDDN, which achieves selective learning by physical inference.

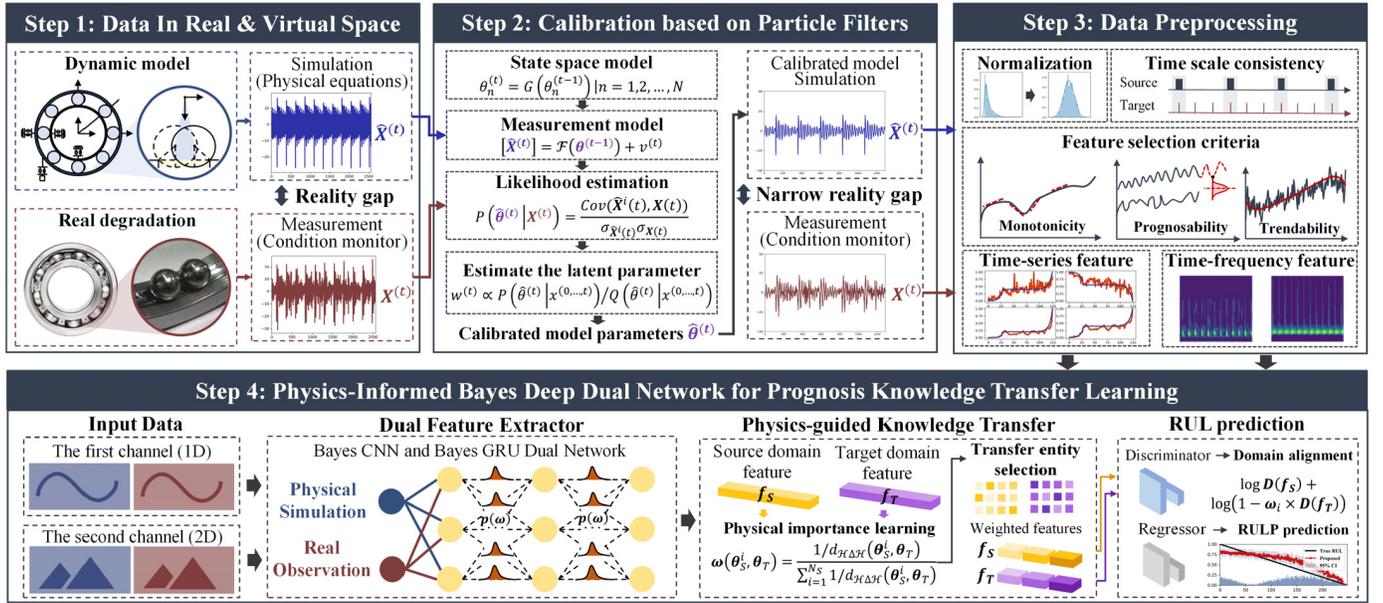


Fig. 1. Overall flowchart of the proposed calibration-based hybrid transfer learning framework.

The physical information enhances the model's prognostic performance from two aspects: simulation data could enrich the informative features of input space and the unobservable model parameters could measure the importance of source bearings based on their similarities with the current task in degradation behaviors. Furthermore, Bayes propagation is embedded into the deep dual network to provide prognostic uncertainty quantification. Each part is elaborated as follows.

A. 5-DOF Dynamic Model

In this section, a 5-DOF dynamic model considering both the translational and the rotational motions is established, referring to the bearing pedestal model [5]. The schematic of the proposed 5-DOF model is shown in Fig. 2(a), where the displacements of inner and outer races are modeled through four pairs of mass–spring–damper, and high-frequency resonant behavior is described through an extra small sprung mass with relatively high damping.

According to the research emphasis in this article, some necessary simplifications of the dynamic model are listed.

- 1) The nonlinear behavior of elastohydrodynamic lubrication is out of consideration.
- 2) The translational motions are in the xy plane, and the rotational motions are about the z -axis.
- 3) The mass and the inertia of rolling elements are not considered, and the rolling elements are assumed to be arranged with an identical angular interval in bearing.
- 4) The mechanisms of surface morphology and the complicated behaviors of bearing elements (such as cage slippage and dynamic ring misalignment) are reasonably ignored [31].
- 5) The results of race defects can be used to comprehensively characterize the vibration trajectories of bearing long-term damage evolution process [5].

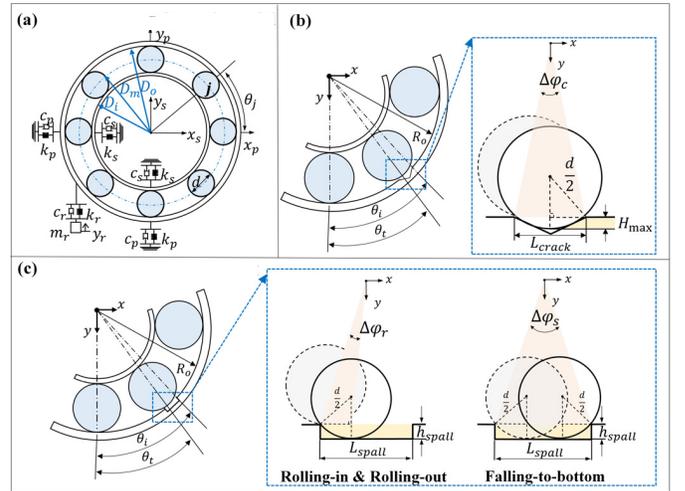


Fig. 2. Schematic of the rolling bearing dynamic model: (a) whole schematic, (b) specified description of the crack stage, and (c) specified description of the damage stage.

The dynamic differential equations of the proposed 5-DOF model are given as follows:

$$\begin{aligned}
 m_s \ddot{x}_s + c_s \dot{x}_s + k_s x_s &= -F_x + em_s \omega_s^2 \cos \omega_s t \\
 m_s \ddot{y}_s + c_s \dot{y}_s + k_s y_s &= -F_y + F_s + m_s g + em_s \omega_s^2 \sin \omega_s t \\
 m_p \ddot{x}_p + c_p \dot{x}_p + k_p x_p &= -F_x \\
 m_p \ddot{y}_p + (c_p + c_r) \dot{y}_p + (k_p + k_r) y_p - k_r y_r - c_r \dot{y}_r &= m_p g + F_y \\
 m_r \ddot{y}_r + c_r (\dot{y}_r - \dot{y}_p) + k_r (y_r - y_p) &= 0
 \end{aligned} \quad (6)$$

where m , k , and c denote the term of mass, stiffness, and damping, respectively; x and y represent the horizontal and vertical response, respectively; the subscripts s , p , and r represent the inner race, the outer race, and the unit resonator, respectively; F_s is the external radial force loaded on the inner race and shaft; e is the eccentric distance; ω_s represents

the shaft angular frequency, for a shaft with RPM as n , $\omega_s = (2\pi n/60)$; and F_x and F_y are contact forces caused by Hertzian deformation.

The contact forces for the N-ball bearings can be given as follows:

$$\begin{cases} F_x = \sum_{j=1}^{n_b} K \delta_j^{1.5} \cos \theta_j H(\delta_j) \\ F_y = \sum_{j=1}^{n_b} K \delta_j^{1.5} \sin \theta_j H(\delta_j) \end{cases} \quad (7)$$

where n_b is the number of rolling elements in a bearing, K is the equivalent contact stiffness between the rolling elements and the raceways and a detailed calculative process can be referred to [5], δ_j is the relative contact deformation of the j th rolling element, $H(\delta_j)$ is a piecewise function as $H(\delta_j) = 1, \delta_j > 0$ or $H(\delta_j) = 0, \delta_j \leq 0$ considering only when the contact deformation is positive, and the compression behavior will occur. θ_j is the angular position of the j th rolling element at time t , which can be calculated as

$$\theta_j = \frac{2\pi(j-1)}{n_b} + \omega_c t \quad (8)$$

$$\omega_c = \frac{1}{2} \left(1 - \frac{d}{D_m} \cos \alpha \right) \omega_s \quad (9)$$

where ω_c is the angular velocity of the ball through the outer ring, d is the diameter of the rolling element, and D_m is the pitch diameter of the bearing.

When rolling bearings are running under a healthy state, there have no cracks or spall damages on the raceway, the corresponding deformation can be written as

$$\delta_r = (x_s - x_p) \cos \theta_j + (y_s - y_p) \sin \theta_j + C_r \quad (10)$$

where C_r is the radial clearance of the bearing. By substituting (7) and (10) into (6), the 5-DOF dynamic model for rolling bearing in the normal condition is established.

The entire degradation process of rolling bearing can be divided into the dentation stage, the cracking propagation stage, the spall damage occurrence stage, and the spall damage growth stage [3]. Since the micromorphology-related behaviors have been simplified, the dynamic model of microdefect caused by the surface dent is not considered. The damage propagation model (corresponding to the cracking propagation stage) and the damage growth model (corresponding to the spall damage occurrence stage and the spall damage growth stage) are established to reveal the vibration response with the bearing failure evolution process.

1) Damage Propagation Model of the Crack Propagation: The damage propagation originates from the surface crack, and a detailed description of the crack-induced defect is shown in Fig. 2(b). As shown in Fig. 2(b), a small amount of deformation will occur when each rolling element passes through the cracked area. Because the initial crack width is smaller than the diameter of the rolling element, each rolling element will not fully contact the bottom of the crack. Correspondingly, the relative of contact deformation in the damage propagation stage can be rewritten as $\delta_r + \eta_{\text{crack}}$.

When the rolling element passes through the crack area, the displacement excitation η_{crack} generates as follows:

$$\eta_{\text{crack}} = \begin{cases} \frac{\sin(\theta_j - \theta_0)}{\sin(\Delta\varphi_c/2)} H_{\text{max}}, & 0 \leq \theta_j - 2n_j\pi \\ & -\theta_0 \leq \frac{\Delta\varphi_c}{2} \\ \frac{\sin(\Delta\varphi_c + \theta_0 - \theta_j)}{\sin(\Delta\varphi_c/2)} H_{\text{max}}, & \frac{\Delta\varphi_c}{2} < \theta_j - 2n_j\pi \\ & -\theta_0 \leq \Delta\varphi_c \end{cases} \quad (11)$$

where $\Delta\varphi_c$ represents the circumferential span angle of the crack area, L_{crack} indicates the circumferential width of the crack area, and θ_0 is the initial angular position of the crack area. H_{max} is the corresponding maximum displacement excitation as the rolling element locates at the bottom of the crack area, which can be calculated as

$$H_{\text{max}} = \frac{d}{2} - \sqrt{\left(\frac{d}{2}\right)^2 - \left(\frac{L_{\text{crack}}}{2}\right)^2}. \quad (12)$$

By introducing (11) into (10), the dynamic model in the damage propagation stage can be obtained.

2) Damage Growth Model of the Spall Growth: The detailed description of spall-induced defect is shown in Fig. 2(c). From Fig. 2(c), the contact deformation behavior between the rolling element and the spalled area can be represented through three phases: the rolling-in phase, falling-to-bottom phase, and rolling-out phase. Specifically, the relative of contact deformation in the spall stage can be rewritten as $\delta_r + \eta_{\text{spall}}$.

η_{spall} represents the displacement excitation when each rolling element passes through the spalled area, which is given by a piecewise function as

$$\eta_{\text{spall}} = \begin{cases} \frac{\sin(\theta_j - \theta_0)}{\sin(\Delta\varphi_r)} h_{\text{spall}}, & 0 < \theta_j - \theta_0 - 2n_j\pi \\ & \leq \Delta\varphi_r \\ \frac{\sin(\Delta\varphi_s + \theta_0 - \theta_j)}{\sin(\Delta\varphi_r)} h_{\text{spall}}, & \Delta\varphi_s - \Delta\varphi_r < \theta_j \\ & -\theta_0 - 2n_j\pi \leq \Delta\varphi_s \end{cases} \quad (13)$$

where $\Delta\varphi_s$ represents the circumferential span angle of the spalled area, L_{spall} indicates the circumferential width of the spalled area, $\Delta\varphi_r$ represents the circumferential span angle of the rolling element when it is just at the edge of the spalled area (rolling-in phase and rolling-out phase), and h_{spall} is the damage depth of the spall. When the rolling element falls into the spall area (falling-to-bottom stage), the displacement excitation $\eta_{\text{spall}} = h_{\text{spall}}$. By introducing (13) into (10), the dynamic model in the spall damage stage can be obtained.

B. Particle Filters Calibration

The particle filter-based calibration approach is applied to infer the values of the model-correcting parameters, which provides a reasonable compromise between the computational complexity and simulation fidelity. Due to the page limits,

the key procedures of particle filters are provided in this section, and detailed discussions about particle filters are not described in this article, which can be referred to [28]. In this article, the state-space vectors comprise the damage properties of bearings, including the damage width and the damage depth. According to the evolution behaviors of crack propagation and spall growth, the damage width state and the damage depth state are modeled as exponential models, and the corresponding state equations can be expressed as follows:

$$\theta_t^\# = C_{t-1}^\# \theta_{t-1}^\# m_{t-1}^\# dt + \theta_{t-1}^\#, \quad \# = L, h \quad (14)$$

where $\theta_t^\#$ denotes the value of defect due to cracking or spalling at the moment t and the subscripts L and h represent the defect width and defect depth, respectively.

The measurement equation is readily available by solving the proposed 5-DOF bearing dynamic model $\mathcal{N}(\cdot)$ using (6), which can be formulated as follows:

$$\mathbf{X}(t) = \mathcal{F}[\boldsymbol{\theta}(t)] + \mathbf{v}(t) \quad (15)$$

where $\mathbf{X}(t)$ denotes measurements of vibration signals, $\boldsymbol{\theta}(t)$ denotes the model internal parameters including the defect state vectors $[\theta_L(t), \theta_h(t)]$ and information on working conditions $[F_s(t), \omega_s(t)]$, and $\mathbf{v}(t)$ is supposed as a Gaussian noise.

It should be noticed that the above equation is only associated with the measurement noise term, and the estimation from the particle filtering step may suffer from significant variability. In order to retain the effectiveness of model calibration, the ensemble learning technique is employed in this procedure to minimize the model form uncertainty $\gamma(t)^2$. The model form uncertainty, also called the model epistemic uncertainty, accounts for the uncertainty in the model parameters caused by a lack of knowledge or information, which can be estimated as

$$\gamma(t)^2 \approx \frac{1}{T} \sum_{j=1}^T [\mathcal{F}_{\text{ref}}^j[\boldsymbol{\theta}(t)] - \bar{\mathcal{F}}[\boldsymbol{\theta}(t)]]^2 \quad (16)$$

where $\mathcal{F}_{\text{ref}}^j(\cdot)$ denotes T groups of measurement equations with different empirically initialized parameters and $\bar{\mathcal{F}}$ denotes the ensembled prediction of the groups. The ensemble process is performed on different candidate groups in order to be free of model from uncertainties $\gamma(t)^2$.

When new measurements $\mathbf{X}(t)$ are available, the poster distribution in the previous step is used as the prior information in the current step. Particles with small weights are eliminated and particles with large weights are duplicated in the resampling step based on the inverse CDF method. Procedures model dynamic calibration is summarized in Algorithm 1.

C. PI Bayes Deep Learning

Before feeding raw vibration signal $\mathbf{X}(t)$ and simulation $\hat{\mathbf{X}}(t)$ into the proposed network, essential data preprocessing techniques are conducted, including max-min normalization, time-scale alignment, time series-based features, and image-based features extraction and feature prognosis performance evaluation. The time series-based features and image-based features are denoted as $\{\mathbf{t}_\#^i\}_{i=1}^{N_\#}$ and $\{\mathbf{v}_\#^i\}_{i=1}^{N_\#}$, respectively, where $\# = S, T$ denotes the source or the target domain.

Algorithm 1 Particle Filters-Based Dynamic Calibration

Input: $\{\theta^i(t-1)\}_{i=1}^{N_p}, \mathbf{X}(t), \{\mathcal{F}_{\text{ref}}^j(\cdot)\}_{j=1}^T$
Output: $P(\hat{\boldsymbol{\theta}}(t)), P(\hat{\mathbf{X}}(t))$
 $\omega_{\text{sum}} = 0, \omega_{\text{cdf}} = 0, \text{cdf} = []$
for $i = 1, \dots, N_s$ **do**
 Step 1: Draw samples $\theta_t^i \sim P(\theta_t^i | \theta_{t-1}^i)$ using (14)
 Step 2: Calculate the ensemble $\hat{\mathbf{X}}^i(t) = \frac{1}{T} \sum_{j=1}^T \mathcal{F}_{\text{ref}}^j(\theta_t^i)$
 Step 3: Compute the likelihood $\omega_t^i = \frac{\text{Cov}(\hat{\mathbf{X}}^i(t), \mathbf{X}(t))}{\sigma_{\hat{\mathbf{X}}^i(t)} \sigma_{\mathbf{X}(t)}}$
 $\omega_{\text{sum}} = \omega_{\text{sum}} + \omega_t^i$
end
for $i = 1, \dots, N_{\text{particles}}$ **do**
 Compute the CDF value of each particle $\text{cdf}[i] = \frac{\omega_{\text{cdf}} + \omega_t^i}{\omega_{\text{sum}}}$
 $\omega_{\text{cdf}} = \omega_{\text{cdf}} + \omega_t^i$
 Step 4: $u = \text{rand}(0, 1), \text{Index} = \text{find}(\text{cdf} \geq u)$
 Assign the particles $\hat{\boldsymbol{\theta}}^i(t) = \theta_t^{\text{Index}}$
end
 $P(\hat{\boldsymbol{\theta}}(t)) = \frac{1}{N_{\text{particles}}} \sum_{i=1}^{N_{\text{particles}}} (\hat{\boldsymbol{\theta}}^i(t)) \sim \sigma(\sum_{i=1}^{N_{\text{particles}}} (\hat{\boldsymbol{\theta}}^i(t)))$
 $P(\hat{\mathbf{X}}(t)) = \frac{1}{T} \sum_{j=1}^T F_{\text{ref}}^j(\hat{\boldsymbol{\theta}}(t)) \sim \sigma(\sum_{j=1}^T F_{\text{ref}}^j(\hat{\boldsymbol{\theta}}(t)))$

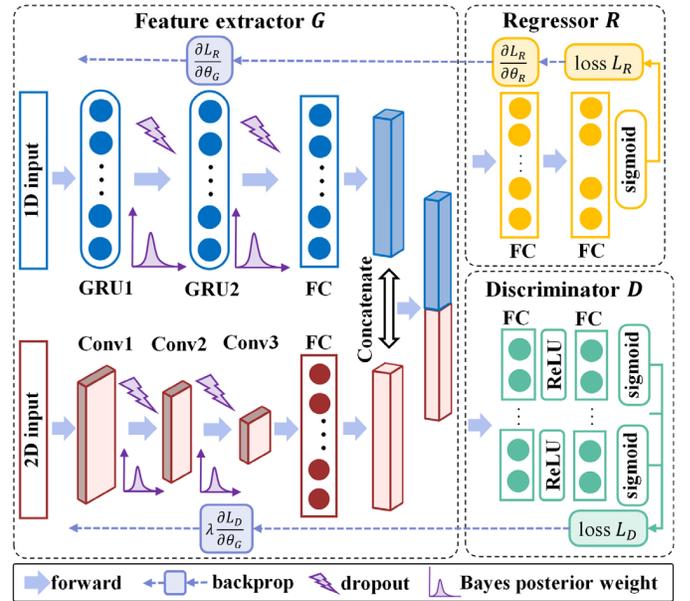


Fig. 3. Detailed structure of BDDN.

1) *Bayes Backpropagation-Based Deep Dual Network:* The detailed architecture is shown in Fig. 3. The proposed network mainly includes a Bayes dual feature extractor G , an RUL regressor R , and multiple subdiscriminators D , which are elaborated as follows.

For a deep dual network + ntations (TFRs), with GRU unit in the forward network, t is processed as

$$z_t = \text{sig}(\mathbf{W}_z \cdot [\mathbf{h}_{t-1}, \mathbf{t}_t] + b_z) \quad (17)$$

$$\mathbf{r}_t = \text{sig}(\mathbf{W}_r \cdot [\mathbf{h}_{t-1}, \mathbf{t}_t] + b_r) \quad (18)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \cdot [\mathbf{r}_t \odot \mathbf{h}_{t-1}, \mathbf{t}_t] + b_h) \quad (19)$$

$$\mathbf{h}_t = (1 - z_t) \odot \mathbf{h}_{t-1} + z_t \odot \tilde{\mathbf{h}}_t \quad (20)$$

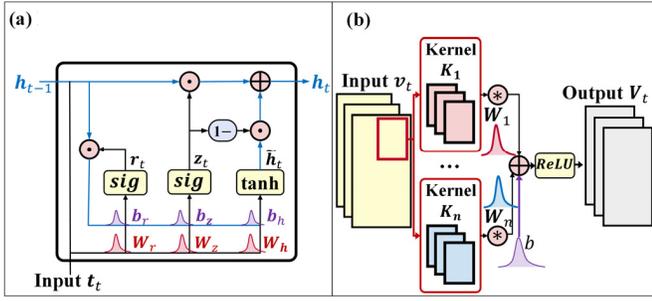


Fig. 4. Bayes propagation: (a) Bayes GRU unit and (b) Bayes CNN layer.

where \mathbf{W} and b with different subscripts z, r , and h are respective weight vectors and bias vectors, respectively; $\text{sig}(\cdot)$ and $\text{tanh}(\cdot)$ are sigmoid and hyperbolic tangent functions, respectively; \odot is a pointwise multiplication; z_t and r_t are outputs of update gate and reset gate, respectively; and \tilde{h}_t represents the candidate hidden state, which could be understood as the information added to the current memory. The output h_t is a linear interpolation between \tilde{h}_t and h_{t-1} and can be regarded as a combination of the prior memory and the current memory.

With CNN unit in the forward network, v is processed as

$$V_t = \text{ReLU}\left(\sum_{n=1}^K \mathbf{W}_n K_n * v_t + b\right) \quad (21)$$

where \mathbf{W}_n is the weight vector of the n th convolutional kernel, b is the corresponding bias, $\text{ReLU}(\cdot)$ is a nonlinear activation function, and $*$ is the convolution operator. The output V_t can be regarded as features extracted with different scales of filters.

In order to address the problem of deep neural networks only providing point estimation without giving confidence intervals, the weight and bias uncertainty are embedded into the deep dual network. The principle of Bayes GRU and Bayes CNN is shown in Fig. 4; it can be seen that all weights and biases are represented by probability distributions over possible values rather than having a single fixed value, and the perturbation exhibiting in each weight is also learned in a way that explains variability in training data coherently.

The respective parameters ϑ_w and ϑ_b of distribution on weight $q(\mathbf{W} | \vartheta_w)$ and bias $q(b | \vartheta_b)$ can be learned by minimizing the Kullback–Leibler (KL) divergence with true Bayes posterior. The general form is given as follows:

$$\begin{aligned} \vartheta^* &= \arg \min_{\vartheta} \text{KL}[q(\mathbf{w} | \vartheta) \| P(\mathbf{w} | \mathcal{D})] \\ &= \arg \min_{\vartheta} \int q(\mathbf{w} | \vartheta) \log \frac{q(\mathbf{w} | \vartheta)}{P(\mathbf{w})P(\mathcal{D} | \mathbf{w})} d\mathbf{w} \\ &= \arg \min_{\vartheta} \text{KL}[q(\mathbf{w} | \vartheta) \| P(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w} | \vartheta)}[\log P(\mathcal{D} | \mathbf{w})] \end{aligned} \quad (22)$$

where $P(\mathbf{w} | \mathcal{D})$ is the posterior distribution of the weights given the training data. The approximate cost depends on the particular weights drawn from the variational posterior.

In order to handle the high-dimensional integration over \mathbf{w} in (22), an unbiased Monte Carlo gradients reparameterization

Algorithm 2 Unbiased Monte Carlo Gradients-Based Optimization

-
- Step 1.** Sample $\epsilon \sim \mathcal{N}(0, I)$
Step 2. Let $\mathbf{w} = \mu + \log(1 + \exp(\rho)) \odot \epsilon$
Step 3. Let $\vartheta = (\mu, \rho)$
Step 4. Let $f(\mathbf{w}, \vartheta) = \log q(\mathbf{w} | \vartheta) - \log P(\mathbf{w})P(\mathcal{D} | \mathbf{w})$ (23)
Step 5. Calculate the gradient with respect to the mean [32]

$$\Delta \mu = \frac{\partial f(\mathbf{w}, \vartheta)}{\partial \mathbf{w}} + \frac{\partial f(\mathbf{w}, \vartheta)}{\partial \mu}$$

Step 6. Calculate the gradient with respect to the standard deviation [32]

$$\Delta \rho = \frac{\partial f(\mathbf{w}, \vartheta)}{\partial \mathbf{w}} \frac{\epsilon}{1 + \exp(-\rho)} + \frac{\partial f(\mathbf{w}, \vartheta)}{\partial \rho}$$

Step 7. Update the variational parameters

$$\mu \leftarrow \mu - \alpha \Delta \mu, \quad \rho \leftarrow \rho - \alpha \Delta \rho$$
-

trick [32] is introduced, and the cost function is approximated

$$\mathcal{F}(\mathcal{D}, \vartheta) \approx \sum_{i=1}^n \log q(\mathbf{w}^{(i)} | \vartheta) - \log P(\mathbf{w}^{(i)}) - \log P(\mathcal{D} | \mathbf{w}^{(i)}) \quad (23)$$

where $\mathbf{w}^{(i)}$ denotes the i th Monte Carlo sample drawn from the variational posterior $q(\mathbf{w}^{(i)} | \vartheta)$. The variational posterior is supposed with a diagonal Gaussian distribution, and then, a sample of the weights \mathbf{w} can be obtained by sampling a unit Gaussian, shifting it by a mean μ , and scaling by a standard deviation σ .

The standard deviation is set as $\sigma = \log(1 + \exp(\rho))$ and the variational posterior parameters are $\vartheta = (\mu, \rho)$. Thus, the step of Bayes backpropagation can be conducted as follows.

2) *PI Transfer Learning*: The proposed prognosis network includes two optimization objects: 1) minimize the RUL prediction error \mathcal{L}_R on the source domain data and 2) minimize the domain adaptation loss \mathcal{L}_{D_i} with respect to the feature generator and maximize \mathcal{L}_{D_i} with respect to the multiple subdiscriminators.

Since target domain labels are assumed to be unavailable during the training process, only source domain labels are employed, and the regression loss could be formulated as

$$\mathcal{L}_R = \frac{1}{N_s} \sum_{i=1}^{N_s} (\hat{y}_S^i - y_S^i)^2 \quad (24)$$

where \hat{y}_S^i is the predicted RUL at the i th time step and y_S^i is the true underlying RUL labels.

Multiple subdiscriminators are designed to achieve the domain consistency for flexible prognosis knowledge transfer, and the domain adaptation can be jointly optimized as

$$\mathcal{L}_{D_k} = - \sum_{i=1}^{N_K} (\log(\hat{d}_k^i)) - \sum_{i=1}^{N_T} (\log(1 - \hat{d}_k^i)) \quad (25)$$

where \mathcal{L}_{D_k} is the binary cross entropy loss of the k th source discriminator, N_K is sample number in the k th degradation region, and \hat{d}_k^i is the domain prediction at the i th time step.

Based on the aforementioned loss \mathcal{L}_{D_k} and \mathcal{L}_R , a PI weighting optimization function is designed to balance the transferring influence of different source bearings: more attention is expected on those discriminators sharing a similar failure

mode, as well as less attention is laid on the discriminators for the outlier failures, which are formulated as

$$\begin{aligned} (\hat{\theta}_G, \hat{\theta}_R) = \operatorname{argmin} & \\ & \times \left(\mathcal{L}_R(\theta_G, \theta_R) - \min_{\theta_{D_1}, \dots, \theta_{D_K}} \right. \\ & \left. \times \alpha \sum_{k=1}^K \omega(\mathbf{X}_S^k, \mathbf{X}_T) \mathcal{L}_{D_k}(\theta_G, \theta_{D_k}) \right) \end{aligned} \quad (26)$$

$$(\hat{\theta}_{D_1}, \dots, \hat{\theta}_{D_K}) = \operatorname{argmin} \sum_{k=1}^K w_k \mathcal{L}_{D_k}(\theta_G, \theta_{D_k}) \quad (27)$$

where $\hat{\theta}_G$, $\hat{\theta}_R$, and $\hat{\theta}_{D_k}$ indicate the saddle point of the model parameters, α is the hyperparameter to trade off these objectives in the unified optimization problem, and $\omega(\mathbf{X}_S^k, \mathbf{X}_T)$ is the designed weights assigned for the i th source sub-discriminator, deciding which region of source entity knowledge should be activated and integrated for the current transfer task.

In this article, the proposed weighting mechanism follows the principle that the knowledge learned from similar physical failure properties could be more informative and indicative of the RUL prediction. The similarity between the bearing's dynamic model parameters $\hat{\theta}$ is designed to obtain the physics-guided weights, which can be formulated as

$$\omega(\mathbf{X}_S^k, \mathbf{X}_T) = \frac{\exp(1/\operatorname{dist}_{\text{MMD}}(\hat{\theta}_S^k, \hat{\theta}_T))}{\sum_{k=1}^K \exp(1/\operatorname{dist}_{\text{MMD}}(\hat{\theta}_S^k, \hat{\theta}_T))} \quad (28)$$

where $\hat{\theta}_S^k$ are the dynamic model calibrated parameters of the k th source bearing degradation region and $\hat{\theta}_T$ are the dynamic model parameters of the target bearing unit. $\operatorname{dist}_{\text{MMD}}(\cdot)$ is the maximum mean discrepancy (MMD) function, which is based on the notion of embedding probabilities in a reproducing kernel Hilbert space (RKHS). $\sup\{\cdot\}$ is the supremum of the input aggregate of RKHS, and \mathcal{H} is a nonlinear mapping function from the original space to RKHS.

The designed weighting mechanism provides a tighter upper bound of the target risk by assigning higher weights to the source domain, which shares a similar degradation level with the target domain, and the physical parameters of the bearing dynamic model provide a theoretical guarantee to regularize the domain-invariant prognosis feature learning process. The procedure of the proposed PI prognosis knowledge transfer network is shown in Algorithm 3.

IV. CASE STUDY

A. Dataset Description and Data Preprocessing

Two popular rolling bearing datasets from XJTU-SY [33] and PRONOSTIA [34] are utilized in this article to validate the proposed framework, and the bearing test beds are given in Fig. 5. Both datasets contain vibration signals from the horizontal and vertical axes, and there are great differences between two datasets, such as the tested bearing specifications, sampling duration, and sampling interval.

The first prediction time (FPT) of bearings from two datasets is identified by the kurtosis of vibration signal [5] first, and the measurement signal between FPT and

Algorithm 3 Physics-Informed Transfer Learning

Input: $\mathcal{D}_S, \hat{\mathcal{D}}_S, \mathcal{D}_T, \hat{\mathcal{D}}_T, P(\hat{\theta}_S), P(\hat{\theta}_T)$
Output: $\hat{\theta}_G, \hat{\theta}_R, \hat{\theta}_{D_k}$
Data Preprocessing: $\mathcal{D}_S = \{[t_S^i, P(\hat{t}_S^i)], [v_S^i, P(\hat{v}_S^i)]\}_{i=1}^{N_S}$,
 $\mathcal{D}_T = \{[t_T^i, P(\hat{t}_T^i)], [v_T^i, P(\hat{v}_T^i)]\}_{i=1}^{N_T}$
Initialize $G, R, \{D_k\}_{k=1}^K$, Batch size = N_b
for iteration = 1, 2, ..., max_iteration **do**
 draw samples $[\hat{t}_S^i, \hat{v}_S^i, \hat{t}_T^i, \hat{v}_T^i]$ from corresponding distributions
 for $k = 1, 2, \dots, K$ **do**
 Extract source fused feature $F_S = G([t_S, \hat{t}_S], [v_S, \hat{v}_S])$
 Extract target fused feature $F_T = G([t_T, \hat{t}_T], [v_T, \hat{v}_T])$
 Calculate the source predicted RULs $\hat{y}_S^k = R(F_S^k)$
 Feed F_S and F_T to multiple sub-discriminators D_k
 Calculate the prediction $\hat{d}_S^k = D_k(F_S^k)$, $\hat{d}_T = D_k(F_T)$
 Calculate the physical weight $\omega(\mathbf{X}_S^k, \mathbf{X}_T)$
 End
 Update $\theta_G, \theta_R, \{\theta_{D_k}\}_{k=1}^K$
end
Obtain the $\hat{\theta}_G, \hat{\theta}_R, \{\hat{\theta}_{D_k}\}_{k=1}^K$ until convergence of $G, R, \{D_k\}_{k=1}^K$

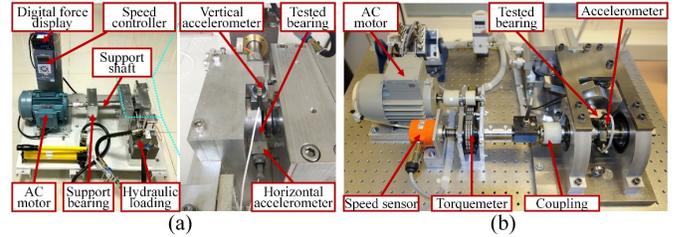


Fig. 5. Testbed of rolling bearing: (a) XJTU-SY and (b) PRONOSTIA.

end of life (EOF) is employed as the raw inputs. The detailed information of selected bearings is shown in Table I.

Essential data preprocessing should be conducted to eliminate the measurement noise and to improve the model training efficiency. First, data normalization is employed on all samples based on the min-max approach. It is noted that the testing dataset is not accessible during training, and therefore, the normalization boundary values for the testing dataset are predefined based on the minimum and maximum of the historical training dataset. Second, the time-scale alignment is implemented for the different sampling intervals and sampling duration across the above two datasets, the vibration signal of PRONOSTIA dataset is downsampling to 60 s, and the measurement length of XJTU-SY is reduced to 2560 samples to keep the time-scale consistency. Finally, both TFRs and time series features are extracted from the raw signal. The TFRs are extracted by continuous wavelet transform (CWT) and resized into a 2-D image with a size of 80×80 . The time series-based features include 15 time-domain features, 16 frequency-domain features, and one trigonometric feature, which are listed in detail in Table II.

In addition, feature ranking based on the monotonicity, the prognosability, and the trendability is conducted to further select effective inputs for RUL prediction, which are shown in Fig. 6. The top 16 features are chosen as the final time series inputs.

TABLE I
DETAILED INFORMATION OF THE DATASET

Bearing Dataset	Operating condition	Bearing number	FPT (min)	RUL (min)
XJTU-SY	C1 2100 rpm 12 kN	B1_1	78	45
		B1_2	36	125
		B1_3	58	100
		B1_4	29	23
	C2 2250 rpm 11kN	B2_1	452	39
		B2_2	46	115
		B2_3	302	231
		B2_4	121	218
	C3 2400 rpm 10 kN	B3_1	2346	192
		B3_2	340	30
		B3_3	1417	98
		B3_4	10	104
PRONOSTIA	C3 1800 rpm 4 kN	B4_1	219	248
		B4_2	137	8
		B4_3	287	109
		B4_4	180	56

TABLE II
DETAILED INFORMATION OF TIME SERIES FEATURES

F_1	Maximum:	F_2	Minimum
F_3	Mean value:	F_4	Peak to Peak
F_5	Mean absolute:	F_6	Variance
F_7	RMS:	F_8	Standard deviation
F_9	Skewness	F_{10}	Kurtosis
F_{11}	Shape factor	F_{12}	Mean absolute square:
F_{13}	Crest factor	F_{14}	Impulse factor
F_{15}	Margin factor	F_{16-3}	Energies of sub-bands
F_{32}	Standard deviation of inverse hyperbolic cosine		

TABLE III
PARAMETERS OF THE BEARING DYNAMIC MODEL

Item	Parameter	XJTU-SY	PRONO STIA
Tested bearing specification	Outer race diameter D_o (mm)	39.80	29.10
	Inner race diameter D_i (mm)	29.30	22.10
	Mean diameter D_m (mm)	34.55	25.60
	Ball diameter d (mm)	7.92	3.50
	Radial clearance C_r (mm)	30×10^{-3}	50×10^{-3}
	Eccentric distance e (mm)	50×10^{-3}	50×10^{-3}
	Number of balls n_b	8	13
	Static loading rate (kN)	6.65	2.47
	Dynamic loading rate (kN)	12.82	4.00
	Load deflection factor K (N/m ^{1.5})	9.545×10^9	
Model internal parameter	Mass m_s, m_p, m_r (kg)	1.2638, 12.638, 1	
	Stiffness k_s, k_p, k_r (N/m)	4.241×10^8 , 15.1056×10^8 , 8.8826×10^7	
	Damping c_s, c_p, c_r (Ns/m)	1376.8, 2210.7, 2424.8	

B. Implementation Details

The detailed procedures of the proposed hybrid prognosis framework are composed of physical simulation construction, particle filter-based calibration, and PI transfer learning network training. For the implementation of physical simulation construction, the parameters of rolling bearing dynamic models used in this article are listed in Table III.

Once the referenced dynamic model parameters are obtained, the simulation vibration signal can be obtained through a fixed-step fourth-order Runge–Kutta numerical method. The step Δt is as $1/25\ 600$ s and $t_2 - t_1 = 0.1$ s to be consistent with the sampling frequency and sampling duration

TABLE IV
HYPERPARAMETERS OF THE NETWORK

Parameter	Value	Parameter	Value
Optimizer	Adam	Learning rate	5.00×10^{-4}
Iteration number	1000	Batch size	256
Dropout rate	0.2	α	0.25

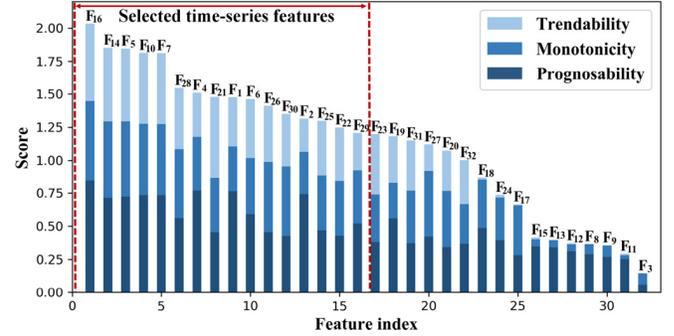


Fig. 6. Selected 16 time series features based on the evaluation index.

of the above datasets, respectively. The dynamic equations are iteratively solved with help of MATLAB software.

The model calibration is implemented to bridge the reality gap. Since latent failure properties are difficult to measure precisely from the real equipment, parameters of ensemble particle filters are initialized in a reasonable range empirically [5]. The respective initial damage width and the initial damage depth are set as 0–200 and 1.3–13 μm , respectively, and the degradation rates of crack and spall are set to 5–60 and 25–600 $\mu\text{m}/\text{min}$, respectively. The significance level is set as 2.5, and the number of particles is set as 100. The detailed hyperparameter setting of the network is given in Table IV.

C. Evaluation Metrics

Two evaluation metrics, root-mean-square error (RMSE) and mean absolute error (MAE), are utilized to measure the quantitative prediction performance of the proposed method, which are calculated as: $\text{RMSE} = (\sum_{i=1}^{N_T} (y_T^i - \hat{y}_T^i)^2 / N_T)^{1/2}$ and $\text{MAE} = \sum_{i=1}^{N_T} |y_T^i - \hat{y}_T^i| / N_T$, where y_T^i and \hat{y}_T^i are the actual RULs and predicted RULs of the target testing dataset, respectively. In order to measure the uncertainty quantification capacity of the proposed method, the coverage probability of prediction interval (PICP) is utilized in this article, which is calculated as: $(\sum_{i=1}^{N_T} c_i) / N_T$, where

$$c_i = \begin{cases} 1, & L_i \leq c_i \leq U_i \\ 0, & \text{else} \end{cases}$$

in which L_i and U_i are the minimum and maximum values of the i th predicted RULs, respectively.

D. Physical Simulation Verification

In order to verify the effectiveness of the proposed dynamic model, the simulated damage signal and actual measurements are investigated in this section. Three bearings with obvious fault type of XJTU dataset (B1_1, B2_2, and B3_1)

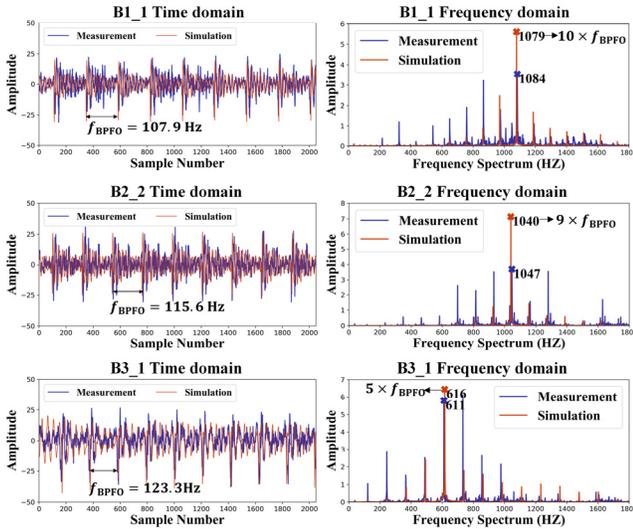


Fig. 7. Time-domain waveform and FFT spectrums comparisons between the physical simulation and actual measurement.

are selected for comparison. The time-domain waveforms of the last time step and the corresponding Fourier frequency transformation (FFT) spectrums are visualized in Fig. 7. From Fig. 7, it can be observed that the proposed dynamic model could generate high-fidelity vibration response under different rotation speeds and working load. From the FFT spectrum, it can be observed that there has a slight difference between the measured characteristic frequency and the theoretical value, which could be attributed to the frequency resolution and the tiny fluctuation of speed. On the other hand, the simulated characteristic frequencies are all in line with the theoretical fault frequency and its multiplier. In summary, the proposed physical model could effectively produce high-fidelity bearing simulation data.

The particle filter-based calibrated results are visualized through the root mean square (rms), which are shown in Fig. 8. It can be seen that the changing trend of each bearing is quite different, and the influence of operation conditions on the degradation trajectory is also very significant.

The rms values of simulation signals are quite approximate to those of the measured samples by dividing the whole degradation trajectory into two stages. Specifically, the early degradation stage could be attributed to the slowly growing deformation caused by the crack. When the crack damage accumulates until the spalling damage stage occurs, there would be two possible scenarios: 1) the spall damage is not severe. Thus, the “rolling in and rolling out phase” would dominate the contact behavior (as B2_2 and B3_1) and relatively small deformation will lead to a steady degradation process and 2) the spalling area is too large that the “fall into bottom phase” dominates the contact behavior (as B1_1) and the produced severe deformation would lead to a steep increase.

The calibration results of all bearings are shown in Fig. 9, and it can be observed that the calibrated model could generate simulations containing a steady evolution process and rapid degradation behavior, which retain the same changing trend

TABLE V
DESCRIPTIONS OF THE TRANSFER TASKS

Task	Training	Validation	Testing	Task	Training	Validation	Testing
A1	B1_1 ~ B1_4	B2_2	B2_1	B2	B1_1 ~ B1_4	B4_3	B4_4
A2	B1_1 ~ B1_4	B2_2	B2_3	B3	B2_1 ~ B2_4	B4_4	B4_2
A3	B2_1 ~ B2_4	B3_1	B3_3	B4	B2_1 ~ B2_4	B4_4	B4_3
A4	B2_1 ~ B2_4	B3_1	B3_4	B5	B4_1 ~ B4_4	B1_2	B1_1
A5	B3_1 ~ B3_4	B1_2	B1_1	B6	B4_1 ~ B4_4	B2_4	B2_2
A6	B3_1 ~ B3_4	B1_2	B1_4	B7	B4_1 ~ B4_4	B3_4	B3_3
B1	B1_1 ~ B1_4	B4_3	B4_1	B8	B4_1 ~ B4_4	B2_2	B2_4

as the actual measurement along the life cycle degradation process.

E. Comparative Analysis of RUL Prediction

In this section, five state-of-the-art transfer learning approaches published in the recent two years are employed to evaluate the performance of the proposed model comparatively: Ding’s method [14] and Cao’s method [15] established the backbone feature extractor with a stacked contractive auto-encoder (SCAE) and a bidirectional-GRU network, respectively, and the multikernel MMD is exploited to estimate the distribution discrepancy of the cross-domain features. Siahpour’s method [17] utilized the domain adversarial neural network (DANN) to extract the shared features across different domains, in which the consistency-based loss is combined into the training process to enhance the domain invariance. Miao’s method [18] designed a selective convolutional RNN (SCRNN) to learn temporal and spatial features for effective prognosis knowledge transfer. Mao’s method [19] designed a transfer domain validity index (T-DVI) to quantify the source domain contribution to achieve selective transfer learning.

Totally, 14 transfer RUL prediction tasks are constructed, including transfer tasks on an identical machine and transfer tasks across different machines, which are listed in Table V. The hyperparameters of comparative methods are initialized based on the settings given in reference [14], [15], [16], [17], [18], [19], and the confidence interval of comparative methods is obtained by the model ensemble learning.

1) *Comparative Results on the Identical Machine:* The RUL prediction results of tasks on the identical machine among the above methods are shown in Table VI. It is clear that the proposed approach gets superior performance in terms of predictive accuracy and uncertainty quantification.

To further verify the effectiveness of the proposed method, the comparative results are visualized in Fig. 10. Due to space limitation, two representative tasks A1 and A3 are chosen for illustration. From Fig. 10(a), it is clear that all methods could perform relatively well on task A1, which could be attributed to the similar failure data distribution between the source domain and the target domain. It can be observed that the proposed method achieves a tighter error bound and more reliable confidence interval compared with other methods. From Fig. 10(b), all comparative methods degenerate significantly with the larger domain gaps between C2 and C3. For the T-DVI which performs well on task A1 but struggles on task A3, the dramatic performance degeneration is because the

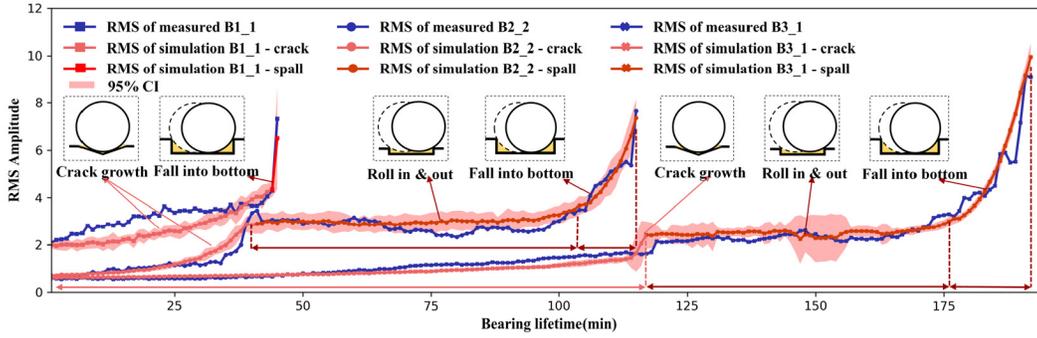


Fig. 8. Lifecycle bearing rms values of calibrated simulations and real measurements.

TABLE VI

PERFORMANCE COMPARISON ON THE IDENTICAL MACHINE

Task	Metric	SCAE	BDGRU	C-DANN	SCRNN	T-DVI	Proposed
A1	RMSE	0.124	0.111	0.207	0.106	0.088	0.058
	MAE	0.104	0.096	0.176	0.091	0.080	0.043
	PICP	0.128	0.256	0.384	0.692	0.487	0.615
A2	RMSE	0.164	0.185	0.145	0.155	0.148	0.123
	MAE	0.133	0.171	0.118	0.132	0.126	0.094
	PICP	0.463	0.277	0.696	0.775	0.129	0.917
A3	RMSE	0.236	0.152	0.212	0.202	0.461	0.120
	MAE	0.206	0.123	0.188	0.148	0.378	0.095
	PICP	0.087	0.500	0.263	0.421	0.026	0.438
A4	RMSE	0.148	0.112	0.097	0.104	0.425	0.084
	MAE	0.109	0.098	0.081	0.086	0.355	0.069
	PICP	0.316	0.346	0.663	0.662	0.001	0.377
A5	RMSE	0.217	0.210	0.180	0.200	0.052	0.075
	MAE	0.188	0.188	0.159	0.168	0.044	0.061
	PICP	0.156	0.244	0.377	0.222	0.640	0.756
A6	RMSE	0.216	0.205	0.232	0.167	0.107	0.078
	MAE	0.194	0.164	0.201	0.144	0.098	0.052
	PICP	0.043	0.652	0.478	0.522	0.783	0.565

parameter transfer strategy in T-DVI is sensitive to the data for pretraining, which would lead to unexpected negative transfer when the divergence between the target pretraining dataset and target testing dataset is too large. For the SCAE, C-DANN, and SCRNN which exploit raw time series and frequency spectrum data as inputs, the poor performance is caused by the unstable feature extraction in unsupervised domain adaptation. For the BDGRU, the feature selection based on Wasserstein distance improves the performance compared with the above methods, but there still is a nonnegligible predicted bias in the latter phase of life. For the proposed method, the predictive fluctuation of the whole life has been suppressed effectively to avoid negative transfer by distilling the prognosis knowledge selectively.

2) *Comparative Results Across Different Machines:* In this section, the effectiveness of the proposed method on the RUL transfer prediction across different machines has been verified. The comparative results are shown in Table VII. From the results, it can be found that the proposed method could retain significant model robustness among all cross-machine tasks, which provides lower values of RMSE and higher values of PICP compared with other DTL methods.

The RUL prediction results on tasks B1 and B5 are shown in Fig. 11. From Fig. 11(a), it can be observed that the proposed

TABLE VII

PERFORMANCE COMPARISON CROSS DIFFERENT MACHINES

Task	Metric	SCAE	BDGRU	C-DANN	SCRNN	T-DVI	Proposed
B1	RMSE	0.168	0.283	0.277	0.187	0.183	0.136
	MAE	0.145	0.243	0.239	0.165	0.141	0.119
	PICP	0.100	0.068	0.180	0.269	0.020	0.369
B2	RMSE	0.143	0.081	0.139	0.094	0.275	0.049
	MAE	0.125	0.064	0.121	0.078	0.233	0.040
	PICP	0.089	0.482	0.750	0.785	0.214	0.785
B3	RMSE	0.238	0.266	0.212	0.089	0.159	0.067
	MAE	0.204	0.237	0.179	0.083	0.137	0.047
	PICP	0.250	0.125	0.250	0.875	0.125	0.875
B4	RMSE	0.327	0.282	0.260	0.252	0.310	0.135
	MAE	0.275	0.245	0.225	0.218	0.271	0.118
	PICP	0.303	0.018	0.294	0.193	0.060	0.420
B5	RMSE	0.236	0.290	0.267	0.191	0.181	0.036
	MAE	0.211	0.257	0.234	0.136	0.156	0.029
	PICP	0.111	0.600	0.311	0.556	0.266	0.711
B6	RMSE	0.236	0.197	0.178	0.212	0.372	0.129
	MAE	0.192	0.182	0.147	0.194	0.319	0.100
	PICP	0.234	0.113	0.478	0.278	0.148	0.773
B7	RMSE	0.199	0.158	0.119	0.153	0.185	0.082
	MAE	0.155	0.141	0.098	0.122	0.163	0.067
	PICP	0.235	0.102	0.602	0.489	0.031	0.610
B8	RMSE	0.180	0.161	0.102	0.136	0.122	0.053
	MAE	0.150	0.145	0.080	0.111	0.115	0.043
	PICP	0.151	0.197	0.587	0.711	0.106	0.440

method shows much superior advantages on grabbing the degradation tendency, which could eliminate the information redundancy and ensure the prediction to be consistent with the true degradation state. From Fig. 11(b), it is observed that the performance across different machines is better than the predictive results on the identical machine, which means that the bearings across different machines also have similar degradation characteristics. This phenomenon is also proved in [21]. It is worth noting that all the compared methods suffer different degrees of fluctuations and prediction bias during the whole bearing life, which can be attributed to insufficient capacity of extracting discriminative features for the cross-machine RUL prediction, while the proposed method could effectively solve this problem, where the physics augmented inputs could suppress the prediction fluctuation of measurement noise and the physical informed weighting could exploit temporal information to provide a more reliable RUL prediction in terms of monotonicity and tendency.

Furthermore, the high-level features learned by the proposed method and other comparative models are visualized by the

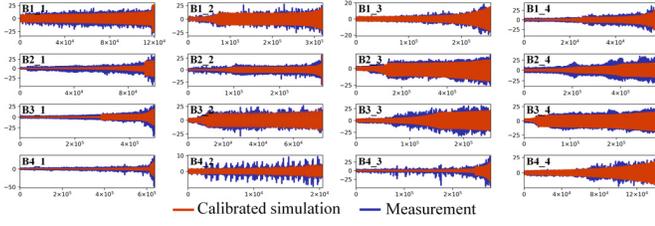


Fig. 9. Calibration simulation and measurement results of all bearings.

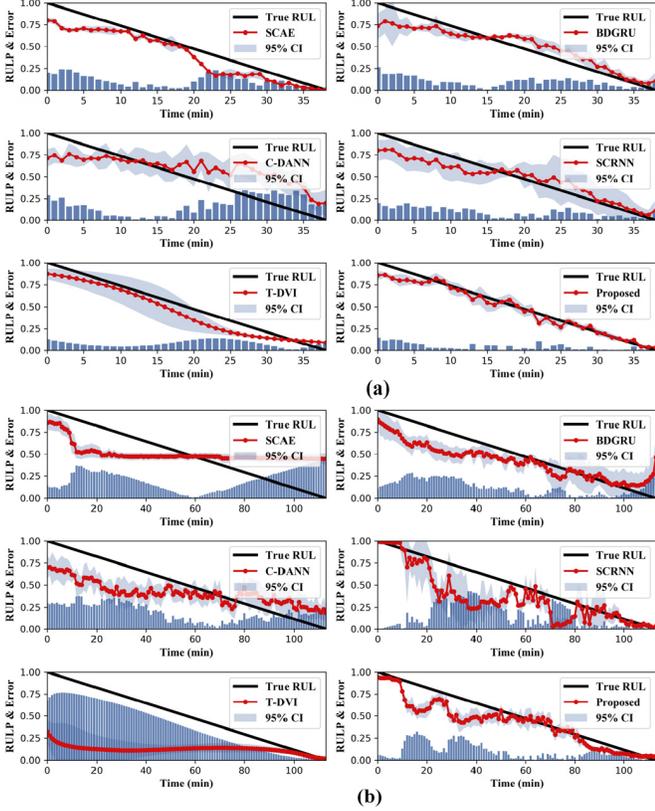


Fig. 10. Comparative results of (a) task A1 and (b) task A3.

t-SNE technique [35]. Taking task B5 as an example, the visualization results of the hidden features are shown in Fig. 12. From the results, it can be seen that the degradation features extracted by all the methods are drawn very closer, which indicates that the distribution discrepancy between the source and target domain has been effectively reduced. Although all the methods show satisfactory performance on domain adaptation, the compared methods still suffer from the problem of degradation trajectories extraction and degradation level alignment. Specifically, the BDGRU and C-DANN models fail to capture the degradation traces. The SCAE, SCRNN, and T-DVI models could capture the degradation traces, but the cross-domain features in the similar degradation level are not aligned. In the proposed method, the degradation trajectories of the cross-domain features are well-aligned and degrade toward the same direction, which validates the effectiveness of the designed PI knowledge transfer learning module on the cross-domain RUL prediction.

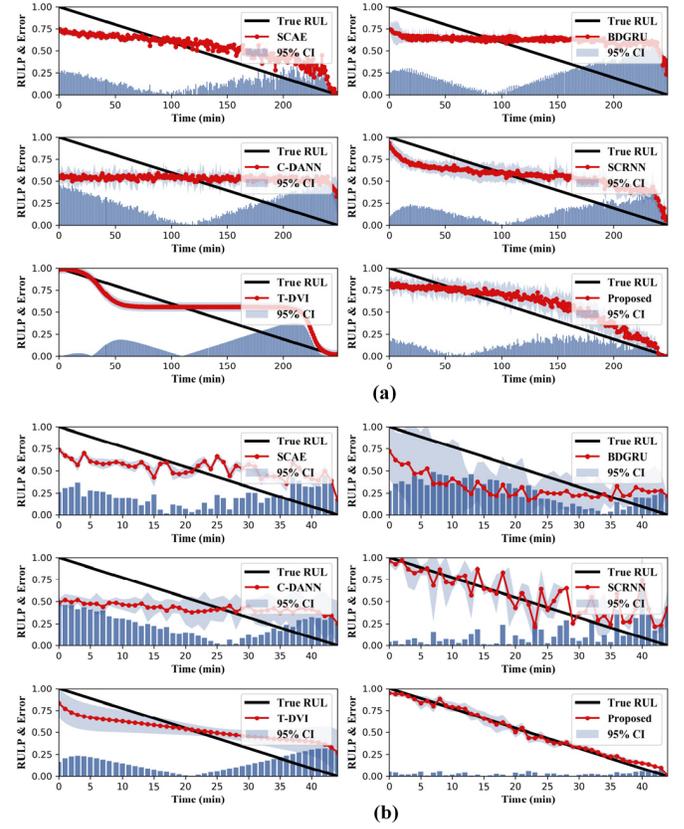


Fig. 11. Comparative results of (a) task B1 and (b) task B5.

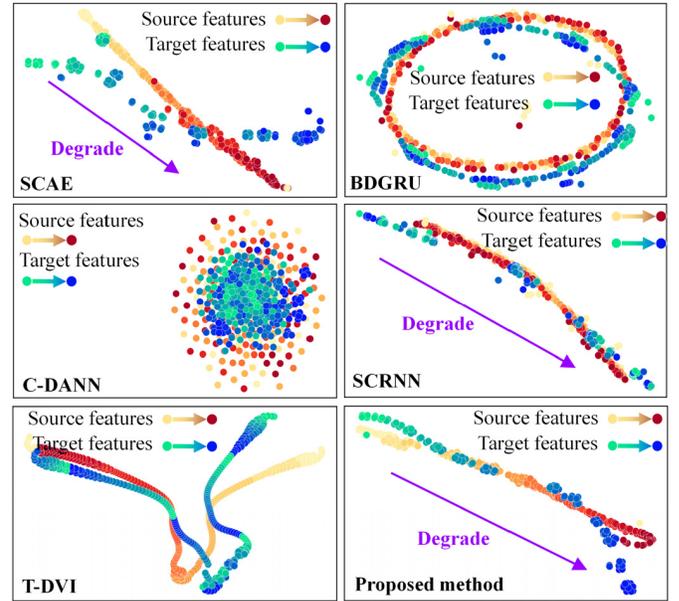


Fig. 12. Feature visualization by different methods on task B5.

F. Ablation Study

In this section, the ablation study is conducted to evaluate the impact of physics-derived information. Totally, six variants of the proposed model are investigated, and the details are listed in Table VIII. Variants V1–V3 are designed to evaluate the contribution of each designed module, including

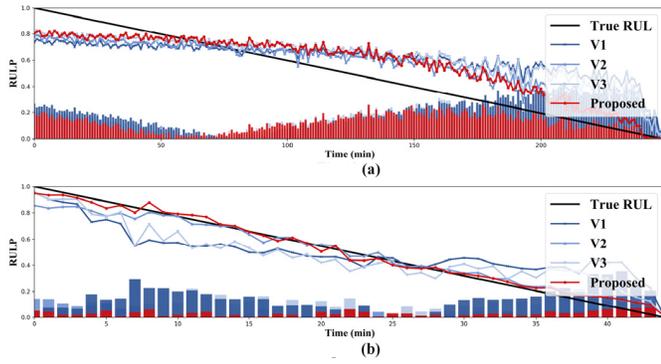


Fig. 13. Predictive results on V1–V3: (a) task B1 and (b) task B5.

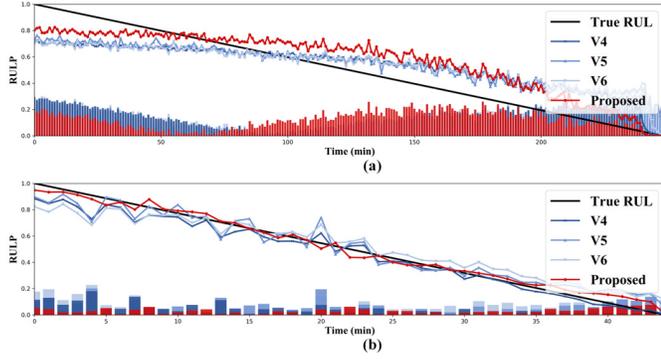


Fig. 14. Predictive results on V4–V6: (a) task B1 and (b) task B5.

TABLE VIII
DETAILS OF MODEL VARIANTS

No	Model component	
	physical simulation augmentation	physics-informed adversarial loss
V1	×	×
V2	✓	×
V3	×	✓
V4	Noisy simulation $SNR_{db} = 10$	✓
V5	Biased simulation $\alpha = \pm 0.1$	✓
V6	Noisy and biased simulation $SNR_{db} = 10, \alpha = \pm 0.1$	✓

physical simulation augmentation and PI adversarial loss. Variants V4–V6 are designed to evaluate the sensitivity of the proposed model to calibration performance, i.e., the impact of low-quality estimates of simulation signals on the RUL prediction performance. Concretely, variant V4 is designed to model the noisy calibration, where the white noise with signal-to-noise ratios (SNR_{db}) of 10 db is imposed on the simulation signal. Variant V5 is designed to model the biased calibration, where an increasing (decreasing) shift proportional to the nominal level is imposed on the calibrated parameters as $\theta_\alpha = \theta^{(t)} + \alpha(\theta^{(0)} - \theta^{(t)})$. Variant V6 considers the potential situation where $\hat{\theta}$ is noisy and also where $\hat{\theta}$ is affected by the model bias.

The comparative results of the variant models on tasks B1–B8 are listed in Table IX. From the results, it can be observed that variant V1 gets the worst performance, and the proposed model shows the best performance; variants V2 and V3 both improve the prognosis performance compared with V1,

TABLE IX
PERFORMANCE COMPARISON OF MODEL VARIANTS

Task	Metric	V1	V2	V3	V4	V5	V6	Proposed
B1	RMSE	0.228	0.159	0.203	0.144	0.161	0.167	0.136
	MAE	0.196	0.140	0.173	0.116	0.140	0.148	0.119
	PICP	0.140	0.265	0.124	0.256	0.261	0.205	0.369
B2	RMSE	0.089	0.058	0.084	0.055	0.058	0.062	0.049
	MAE	0.073	0.045	0.069	0.047	0.048	0.053	0.040
	PICP	0.428	0.768	0.214	0.464	0.750	0.803	0.785
B3	RMSE	0.221	0.094	0.175	0.119	0.139	0.140	0.067
	MAE	0.182	0.079	0.148	0.092	0.130	0.127	0.047
	PICP	0.125	0.500	0.250	0.875	0.125	0.125	0.875
B4	RMSE	0.271	0.141	0.233	0.174	0.145	0.182	0.135
	MAE	0.230	0.109	0.193	0.142	0.117	0.147	0.118
	PICP	0.064	0.568	0.156	0.440	0.376	0.807	0.420
B5	RMSE	0.167	0.057	0.129	0.060	0.078	0.091	0.036
	MAE	0.142	0.044	0.114	0.044	0.063	0.075	0.029
	PICP	0.177	0.578	0.222	0.600	0.422	0.356	0.711
B6	RMSE	0.184	0.147	0.176	0.150	0.154	0.155	0.129
	MAE	0.143	0.127	0.135	0.113	0.116	0.130	0.100
	PICP	0.113	0.052	0.122	0.461	0.313	0.122	0.773
B7	RMSE	0.140	0.106	0.139	0.097	0.089	0.098	0.082
	MAE	0.122	0.079	0.104	0.079	0.077	0.073	0.067
	PICP	0.289	0.357	0.551	0.184	0.357	0.693	0.610
B8	RMSE	0.165	0.106	0.093	0.108	0.087	0.127	0.053
	MAE	0.131	0.089	0.073	0.088	0.073	0.103	0.043
	PICP	0.050	0.252	0.384	0.413	0.312	0.339	0.440

which validates the effectiveness of designed components on improving the performance on cross-domain RUL prediction. In addition, compared with the proposed model, there has a slight performance decrease of variants V4–V6, indicating the negative effect of low-quality model calibration.

In order to further reveal the performance improvement among different variants, the predictive results on tasks B1 and B5 are visualized in Figs. 13 and 14. As shown in Fig. 13(a), V2 effectively suppresses data fluctuations compared with V1, proving that the physical simulation augmentation could facilitate the model to extract more discriminative features, which are beneficial to suppress the negative transfer caused by the noisy sensory data. On the other hand, as shown in Fig. 13(a) and (b), V3 could decrease the prediction bias in the latter phase of whole life, proving that the PI weighting loss could further improve the prognosis transfer across different degrees of domain gaps. Moreover, the proposed model with both modules achieves the best results in terms of prediction accuracy and degradation tendency, demonstrating that the designed modules play an important role in providing reliable and stable performance. From Fig. 14, the proposed model would suffer prediction fluctuation when the physical simulation is noisy or biased. However, even with a noisy and biased physical simulation augmentation, the proposed model is still to achieve a better prognosis performance than the pure DDM (V1), which shows the robustness of the hybrid prognosis framework to the uncertainty of simulation noise and parameter bias.

V. CONCLUSION

In this article, a calibration-based hybrid prognosis framework is proposed for rolling bearing RUL prediction across different machines. The critical issue of this approach is exploiting the physical model to guide meaningful DDM

construction. In order to solve the reality gaps between the physical simulation and actual measurement, the underlying parameters of the physical model are inferred through the particle filter-based calibration. Subsequently, the physical simulation and inference parameters are combined with the condition monitoring signal as inputs to the proposed Bayes deep neural network to develop a robust prognosis model.

The performance of the proposed model is evaluated on two run-to-failure bearing datasets, including transfer tasks across different operating conditions and different machines. From the experimental results, the conclusion is drawn as follows.

- 1) The particle filter-based calibration could effectively narrow the reality gap between the physical simulation and measurements, ensuring that the physical model generates high-fidelity data in terms of fault characteristic frequency and lifecycle failure tendency.
- 2) The comparative experimental results prove the superiority of the proposed method. The proposed method could grab the degradation tendency with less prediction error under different domain discrepancies, and the Bayes backpropagation could provide reliable confidence intervals to cover more real labels.
- 3) The discussions of model components demonstrate that the augmented input space could suppress the prediction fluctuation caused by sensor noise, and physical inference could improve model transferability by selecting vital prognosis knowledge effectively.

The potential future research is developing a hybrid prognosis framework for other mechanical components. Furthermore, an additional avenue is to explore hybrid frameworks for scenarios where only part of physics-based knowledge is available.

REFERENCES

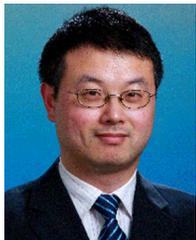
- [1] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, nos. 1–2, pp. 314–334, Jan. 2014.
- [2] J. Zhuang, M. Jia, Y. Ding, and P. Ding, "Temporal convolution-based transferable cross-domain adaptation approach for remaining useful life estimation under variable failure behaviors," *Rel. Eng. Syst. Saf.*, vol. 216, Dec. 2021, Art. no. 107946.
- [3] I. El-Thalji and E. Jantunen, "Dynamic modelling of wear evolution in rolling bearings," *Tribol. Int.*, vol. 84, pp. 90–99, Apr. 2015.
- [4] M. Behzad, H. A. Arghan, A. R. Bastami, and M. J. Zuo, "Prognostics of rolling element bearings with the combination of Paris law and reliability method," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Jul. 2017, pp. 1–6.
- [5] L. Cui, X. Wang, H. Wang, and H. Jiang, "Remaining useful life prediction of rolling element bearings based on simulated performance degradation dictionary," *Mechanism Mach. Theory*, vol. 153, Nov. 2020, Art. no. 103967.
- [6] H. Zhang, P. Borghesani, R. B. Randall, and Z. Peng, "A benchmark of measurement approaches to track the natural evolution of spall severity in rolling element bearings," *Mech. Syst. Signal Process.*, vol. 166, Mar. 2022, Art. no. 108466.
- [7] N. Z. Gebrael, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Trans.*, vol. 37, no. 6, pp. 543–557, Jun. 2005.
- [8] N. Gebrael and J. Pan, "Prognostic degradation models for computing and updating residual life distributions in a time-varying environment," *IEEE Trans. Rel.*, vol. 57, no. 4, pp. 539–550, Dec. 2008.
- [9] Z. Pang, X. Si, C. Hu, D. Du, and H. Pei, "A Bayesian inference for remaining useful life estimation by fusing accelerated degradation data and condition monitoring data," *Rel. Eng. Syst. Saf.*, vol. 208, Apr. 2021, Art. no. 107341.
- [10] Z. Ye, N. Chen, and K.-L. Tsui, "A Bayesian approach to condition monitoring with imperfect inspections," *Qual. Rel. Eng. Int.*, vol. 31, no. 3, pp. 513–522, 2015.
- [11] W. Mao, J. He, and M. Zuo, "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1594–1608, Apr. 2020.
- [12] C.-G. Huang, H.-Z. Huang, Y.-F. Li, and W. Peng, "A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing," *J. Manuf. Syst.*, vol. 61, pp. 757–772, Oct. 2021.
- [13] W. Peng, Z. S. Ye, and N. Chen, "Bayesian deep-learning-based health prognostics toward prognostics uncertainty," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2283–2293, Mar. 2019.
- [14] Y. Ding, P. Ding, and M. Jia, "A novel remaining useful life prediction method of rolling bearings based on deep transfer auto-encoder," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
- [15] Y. Cao, M. Jia, P. Ding, and Y. Ding, "Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network," *Measurement*, vol. 178, Jun. 2021, Art. no. 109287.
- [16] F. Li, T. Tang, B. Tang, and Q. He, "Deep convolution domain-adversarial transfer learning for fault diagnosis of rolling bearings," *Measurement*, vol. 169, Feb. 2021, Art. no. 108339.
- [17] S. Siahpour, X. Li, and J. Lee, "A novel transfer learning approach in remaining useful life prediction for incomplete dataset," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
- [18] M. Miao and J. Yu, "A deep domain adaptive network for remaining useful life prediction of machines under different working conditions and fault modes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.
- [19] W. Mao, J. Liu, J. Chen, and X. Liang, "An interpretable deep transfer learning-based remaining useful life prediction approach for bearings with selective degradation knowledge fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.
- [20] M. Yan, X. Wang, B. Wang, M. Chang, and I. Muhammad, "Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model," *ISA Trans.*, vol. 98, pp. 471–482, Mar. 2020.
- [21] J. Wang, Y. Liang, Y. Zheng, R. X. Gao, and F. Zhang, "An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples," *Renew. Energ.*, vol. 145, pp. 642–650, Jan. 2020.
- [22] Y. Qin, X. Wu, and J. Luo, "Data-model combined driven digital twin of life-cycle rolling bearing," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1530–1540, Mar. 2022.
- [23] Y. A. Yucesan and F. A. C. Viana, "A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation," *Mech. Syst. Signal Process.*, vol. 171, May 2022, Art. no. 108875.
- [24] W. Deng, K. T. P. Nguyen, C. Gogu, J. Morio, and K. Medjaher, "Physics-informed lightweight temporal convolution networks for fault prognostics associated to bearing stiffness degradation," in *Proc. PHM Soc. Eur. Conf.*, 2022, vol. 7, no. 1, pp. 118–125.
- [25] S. Fukata, E. H. Gad, T. Kondou, T. Ayabe, and H. Tamura, "On the radial vibration of ball bearings: Computer simulation," *Bull. JSME*, vol. 28, no. 239, pp. 899–904, 1985.
- [26] H. Cao, L. Niu, S. Xi, and X. Chen, "Mechanical model development of rolling bearing-rotor systems: A review," *Mech. Syst. Signal Process.*, vol. 102, pp. 37–58, Mar. 2018.
- [27] C. Mishra, A. K. Samantaray, and G. Chakraborty, "Ball bearing defect models: A study of simulated and experimental fault signatures," *J. Sound Vib.*, vol. 400, pp. 86–112, Jul. 2017.
- [28] D. An, J.-H. Choi, and N. H. Kim, "Prognostics 101: A tutorial for particle filter-based prognostics algorithm using MATLAB," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 161–169, Jul. 2013.
- [29] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Rel. Eng. Syst. Saf.*, vol. 217, Jan. 2022, Art. no. 107961.
- [30] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [31] Y. Dong, Y. Li, H. Zheng, R. Wang, and M. Xu, "A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample problem," *ISA Trans.*, vol. 121, pp. 327–348, Feb. 2022.

- [32] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1613–1622.
- [33] B. Wang, Y. G. Lei, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2018.
- [34] P. Nectoux et al., "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (PHM)*, Jun. 2012, pp. 1–8.
- [35] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.



Yafei Deng received the B.E. degree in mechanical engineering from the Harbin Institute of Technology, Harbin, China, in 2017. He is currently pursuing the Ph.D. degree in industrial engineering with Shanghai Jiao Tong University, Shanghai, China.

His research interests include hybrid physical model and data-driven model for prognostics and health management, condition monitoring, and intelligent fault diagnosis.



Shichang Du (Member, IEEE) received the B.S. and M.S.E. degrees in mechanical engineering from the Hefei University of Technology, Hefei, China, in 2000 and 2003, respectively, and the Ph.D. degree in industrial engineering and management from Shanghai Jiao Tong University, Shanghai, China, in 2008.

He was at the University of Michigan, Ann Arbor, MI, USA, from 2006 to 2007, as a Visiting Scholar. He is currently a Professor with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University. His current research interests include quality engineering, including quality control with analysis of variation flow, monitoring, and intelligent diagnosis and prognosis of manufacturing systems.

Dr. Du is an Area Editor of *Computers and Industrial Engineering*, an Associate Editor of *Flexible Services and Manufacturing Journal*, and an Editorial Board Member of *IEEM*.



Dong Wang (Member, IEEE) received the Ph.D. degree in systems engineering and engineering management from the City University of Hong Kong, Hong Kong, in 2015.

He was a Senior Research Assistant, a Post-Doctoral Fellow, and a Research Fellow at the City University of Hong Kong. He is currently an Associate Professor with the State Key Laboratory of Mechanical System and Vibration, Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China. His research interests include sparse and complex measures, signal processing, prognostics and health management, condition monitoring and fault diagnosis, statistical learning and machine learning, statistical process control, and nondestructive testing.

Dr. Wang is an Associate Editor of the *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, *IEEE SENSORS JOURNAL*, *Journal of Low Frequency Noise Vibration and Active Control*, and *Journal of Dynamics Monitoring and Diagnostics* and an Editorial Board Member of *Mechanical Systems and Signal Processing*.



Yiping Shao received the B.E. degree in industrial engineering from the Beijing Institute of Technology, Beijing, China, in 2014, and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2020.

He is currently an Assistant Professor with the College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, China. His current research interests include measurement and metrology, intelligent manufacturing, quality engineering, monitoring, and diagnosis.



Delin Huang received the B.E. degree in industrial engineering from Northeastern University, Qinhuangdao, China, in 2013, and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019.

He is currently an Assistant Professor with the College of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai. His current research interests include machine vision, intelligent manufacturing, and monitoring and control of manufacturing process.