

AI-driven wear prediction for linear guides with physical constraints toward resilient manufacturing systems

Yong Li^{1*}, Ping Lu¹, and Bin Wang¹

¹Chengdu Vocational & Technical College of Industry, Chengdu, China

Abstract

Monitoring the progressive degradation of linear guide systems under complex operating conditions is difficult using weak vibration signals, and existing data-driven methods generally suffer from insufficient integration of physical evolution laws and poor calibration of predictive uncertainty. This issue is particularly consequential in high-precision manufacturing, where unexpected wear-induced failures of linear guides can disrupt production continuity and undermine system resilience. To address these problems, this paper proposes an AI-driven wear prediction framework that integrates physical priors with adaptive physical constraints. First, coarse-grained physical prior features are extracted from wear profile data and concatenated with PCA-reduced vibration features to establish a dual-channel fusion mechanism. Second, a lightweight residual network enhances degradation feature extraction, while adaptive physical constraints dynamically balance prediction accuracy with physical regularity. Finally, within a pseudo-Bayesian physics-informed framework, a two-stage model selection and global variance scaling strategy is proposed to achieve well-calibrated uncertainty quantification under weak signal conditions. Experimental results on a linear guide degradation dataset show that the proposed method achieves an R^2 of 0.9553, MAE of 101.70 $\mu\text{m}\cdot\text{mm}$, and RMSE of 143.26 $\mu\text{m}\cdot\text{mm}$, with a 95% prediction interval coverage rate of 0.9460. These results demonstrate that the proposed framework provides reliable, uncertainty-aware wear-state prediction and RUL-oriented decision support for the predictive maintenance of linear guides. By enabling early degradation warning, informed maintenance planning, and rapid recovery preparation, the proposed method can contribute to reducing unplanned downtime and supporting the resilience of high-precision manufacturing systems.

Keywords: linear guide, wear prediction, physical prior, adaptive physical constraints, dual-channel feature fusion.

Received on 13 August 2026, accepted on 24 September 2026, published on 16 October 2026

Copyright © 2026 Yong Li *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ eetsis.14477

1. Introduction

1.1 Research Background

In modern high-precision manufacturing, the linear axis guideway system serves as the core load bearing and positioning component of computer numerical control (CNC) machine tools, semiconductor manufacturing equipment, and

industrial robots, playing a crucial role in achieving precise motion control [1-3]. However, with the continuous increase in processing speed, load cycles, and prolonged operation time of equipment, high-intensity mechanical stress and frictional environments accelerate the degradation of key components of linear guideways, directly affecting the processing accuracy and surface quality of workpieces [4-6].

*Corresponding author. Email: liy19780810@163.com

Additionally, excessive wear may lead to guideway jamming or processing assembly failure, resulting in significant equipment downtime losses and production safety risks [7-9].

Traditional industrial maintenance often relies on reactive maintenance strategies, which means repairs or replacements are only carried out when equipment malfunctions or completely fails [10, 11]. However, such methods are prone to sudden production halts due to unexpected equipment failures, resulting in unnecessary economic losses. Currently, the smart manufacturing industry is accelerating its transition towards Predictive Maintenance (PdM) and Prognostics and Health Management (PHM), aiming to assess the current degradation state of guide rails through real-time monitoring data and predict the Remaining Useful Life (RUL). Therefore, accurately estimating the current wear state of linear guide rails and providing reliable wear-area estimates for subsequent RUL assessment are crucial for ensuring the reliability of precision machining equipment throughout its entire lifecycle [12-14]. More importantly, reliable wear prediction can provide actionable time windows for maintenance scheduling, spare-part preparation, and degradation-aware production adjustment, thereby helping manufacturing systems avoid abrupt failures, shorten recovery time, and maintain stable operation under disturbances.

1.2. Related work

For resilient manufacturing systems, equipment health monitoring should not be limited to identifying component degradation but should also support rapid decision-making after abnormal wear is detected. Accurate and uncertainty-aware wear prediction of linear guides can help maintenance managers determine whether a machine can continue operating, whether production tasks should be rescheduled, and when recovery actions should be initiated. The research paths for guideway degradation, surface wear, and RUL prediction can be systematically summarized into three categories: physical model methods, data-driven machine learning approaches, and preliminary hybrid model methods [15,16].

1.2.1 Physics-based Models

Early research primarily focused on establishing and analyzing physical models based on contact mechanics and tribology. Cheng et al. [17] proposed a method for predicting the accuracy degradation of rolling guides, considering multiple random parameters and the effects of frictional thermal expansion. This approach constitutes a physical framework based on probability statistics. Its main contribution lies in quantifying how random load fluctuations and thermal deformations collaboratively accelerate accuracy degradation. Tao et al. [18] established a basic wear prediction model specifically for roller linear guides, providing a mathematical foundation for calculating operational wear. From a system-level and dynamic perspective, Zhang et al. [19] investigated the factors governing the running accuracy of linear ball guides, while Miao et al. [20] conducted a comprehensive

analysis of vibration characteristics and reliability on roller linear guideway workbenches, revealing how dynamic behavior correlates with structural reliability.

1.2.2 Data-driven machine learning approaches

With the advancement of sensor technologies, data-driven machine learning (ML) techniques have emerged as powerful tools for condition monitoring without relying on complex physical equations. To monitor the surface wear state of linear rolling guides in real-time, Nie et al. [21] proposed a surface wear state recognition method based on multi-scale fuzzy entropy and random forest algorithm, achieving robust feature extraction. In terms of structural assembly, Hui et al. [22] developed a data-driven modeling approach to evaluate the assembly quality of machine tool linear axes. Wu et al. [23] proposed a fault diagnosis framework for linear guide rails utilizing Synchrosqueezed S-transform (SSTG) combined with a CA-DenseNet convolutional architecture, significantly enhancing diagnostic accuracy. Li et al. [24] presented a structured framework that categorizes the use of ML across wear monitoring, friction prediction, and lubricant analysis, and identifies key challenges and future directions for data-driven tribological research.

1.2.3 Hybrid Models and Prognostic Frameworks

To bridge the gap between physical interpretability and data-driven adaptability, recent literature has shifted towards hybrid fusion models and comprehensive prognostic maintenance. At the infrastructure level, Sicard et al. [25] reviewed experimental setups for predictive maintenance of linear actuators and conducted surveys and classifications of test benches and sensor suites, which are crucial for the synchronous physical and data flows required for developing and validating hybrid models. Regarding tool and component wear, Saatci et al. [26] presented a systematic review on hybrid AI systems, evaluating how expert physical knowledge combines with AI models in modern manufacturing. Finally, Yang et al. [27] synthesized recent advancements in prediction methodologies for rail wear, emphasizing the paradigm shift from single-domain formulations towards multi-source data and physical model fusion. Gao et al. [28] developed a hybrid physics-data-driven model-based fusion framework for wear prediction, successfully merging physical boundaries with algorithmic flexibility.

1.3 Research Gaps and Contributions

Through an in-depth analysis of existing literature, this study identifies the following core gaps:

(1) Inadequate integration of physical laws and data characteristics: Existing research often overlooks the distribution characteristics (i.e., "profile") of wear along the lateral direction and along the path of the guide rail. Most existing hybrid methods rely on pre-generated physical simulation data, lacking the ability to effectively construct and embed physical-prior features directly from measured profile data governing the degradation of measured high-precision track profiles.

(2) Modeling bottleneck under weak feature conditions: In practical monitoring, the inertial vibration signals collected by the slider or workbench are often affected by machine tool vibrations, motor noise, etc., resulting in an extremely low signal-to-noise ratio. Existing data-driven models struggle to extract effective guideway degradation information from these "weak features" of machine tool operation.

(3) Calibration Defects in Uncertainty Quantification: Existing Bayesian neural networks are prone to exhibit an "over-confident" phenomenon under weak signal conditions. How to effectively calibrate epistemic uncertainty during the degradation process of guide rails remains a current research challenge.

Addressing the aforementioned issues, this study proposes a framework for predicting guideway wear. The main contributions are as follows:

(1) A dual-channel feature fusion mechanism based on physical profile evolution laws has been constructed: By dividing the degradation interval and analyzing the profile evolution laws, coarse-grained physical prior features are constructed from high-precision physical pro-files. Finally, by fusing multi-condition vibration features processed by PCA, a dual-channel input architecture of "macroscopic physical trend + microscopic data-driven" is formed, effectively solving the problem of information extraction under weak feature conditions.

(2) A lightweight residual network architecture with an adaptive physical constraint mechanism has been designed: To address the issue that pure data-driven models struggle to meet physical consistency under weak feature conditions, a lightweight residual network incorporating skip connections is constructed to enhance the ability to extract degradation information. Additionally, the loss function explicitly incorporates a monotonicity constraint on cumulative wear area and a training-data-derived quadratic degradation prior. Their contributions are adaptively balanced through learnable log-variance parameters, enabling the model to improve prediction accuracy while maintaining physically consistent degradation behavior.

(3) A pseudo-Bayesian uncertainty quantification framework has been developed. A calibration strategy integrating "two-stage hyperparameter search" with "global adaptive variance scaling" is proposed. This method significantly enhances the coverage rate of the 95% prediction interval, effectively mitigating the model's overconfidence behavior.

Overall, the proposed method links component-level wear prediction with system-level resilience enhancement by enabling early warning, uncertainty-aware maintenance planning, and faster recovery of high-precision manufacturing equipment.

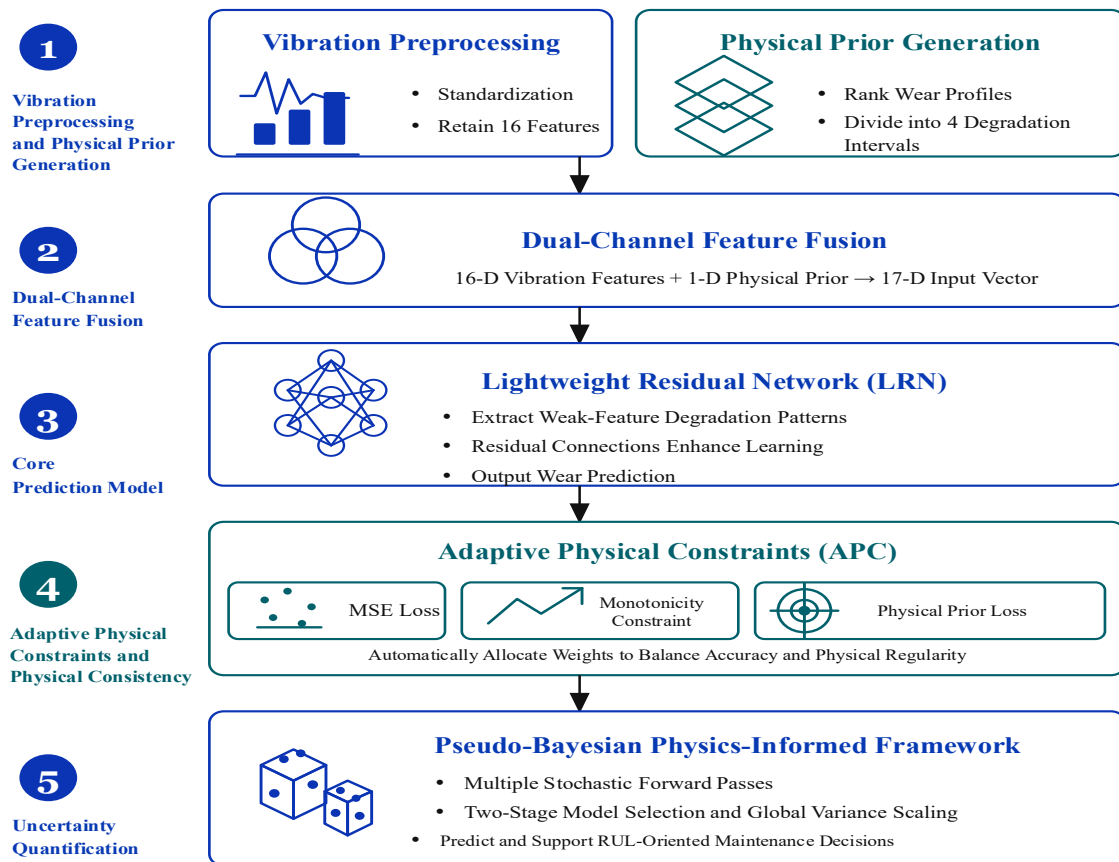


Figure 1. Schematic of the Wear Prediction Method for Linear Guides

2. Methodology

The wear prediction method for linear guides proposed in this study is illustrated in Figure 1. First, the vibration signals are standardized and dimensionality-reduced via Principal Component Analysis (PCA). Simultaneously, based on the evolution patterns of the wear profile, the degradation stages are ranked and partitioned to construct a one-dimensional physical prior. Subsequently, these two types of information are fused into a 17-dimensional input vector and fed into a lightweight residual network (LRN), where residual connections are utilized to enhance the extraction of weak degradation patterns against a noisy background. During the model training process, an Adaptive Physical Constraint (APC) is introduced. This jointly considers the data fitting loss, wear monotonicity constraint, and physical prior loss. Learnable weights are employed to dynamically balance prediction accuracy with physical consistency, while Post-processing of Cumulative Maximum (PCM) is adopted to further ensure that the prediction results are monotonically non-decreasing with the degradation stages. Finally, a pseudo-Bayesian physics-informed framework is established based on the optimal prediction model. By performing multiple random forward passes, two-stage model selection, and global variance scaling, the prediction uncertainty is calibrated to output the final predictions. This provides support for degradation-aware predictive maintenance decision-making of linear guides.

2.1 Guide rail degradation dataset and vibration feature preprocessing

This study utilizes the linear guideway degradation dataset published by the National Institute of Standards and Technology (NIST) [29]. This dataset comprises 14 degradation stages (Stage 1–14), with Stage 1 representing a healthy state without damage and Stage 14 representing the most severely damaged state. This multi-condition data was obtained through 50 bidirectional round-trip experiments conducted at three operating speeds for the trolley in each degradation stage.

Vibration feature extraction: First, acceleration and angular velocity segments of the trolley passing through the wear area are extracted based on the position signal. Then, a total of 51-dimensional features are extracted from each segment in both the time and frequency domains. Time-domain features include: root mean square (RMS), peak, peak-to-peak, kurtosis, skewness, form factor, and impulse factor of three-axis acceleration and three-axis angular velocity, as well as the RMS and peak values of the synthesized modulus of acceleration/angular velocity. For frequency-domain features, taking the acceleration X-axis as an example, the energy proportion of the power spectrum in the low, medium, and high frequency bands is calculated, as well as the spectral centroid and spectral spread.

To eliminate information redundancy and multicollinearity among the 51 extracted features, PCA [14] is applied to the high-dimensional feature set through covariance analysis and orthogonal linear transformation, retaining the leading principal components with a high cumulative explained variance to achieve dimensionality reduction while preserving the major information in the original features.

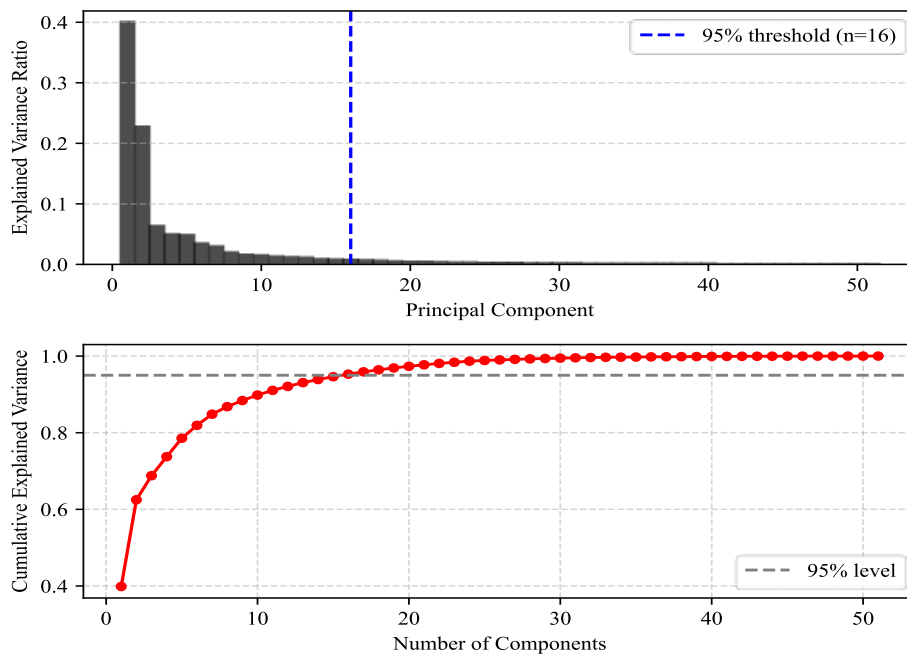


Figure 2. PCA analysis results of vibration features

Figure 2 displays the results of PCA analysis: the scree plot shows that the explained variance ratios of the first few principal components are 0.40, 0.23, 0.08, 0.06, 0.05, and so on. The cumulative variance curve indicates that the cumulative explained variance of the first 16 principal components reaches 95%. Therefore, retaining the 16 principal components can achieve effective dimensionality reduction while maintaining most of the information.

2.2 Profile analysis and physical prior generation

(1) Profile analysis

In addition to vibration signals, a micrometer was used to conduct point-to-point measurements of the guide rail wear profile at each degradation stage, recording wear depth data along the guide rail position (0–100 mm). These profile data directly embody the physical laws governing wear evolution. One of the core innovations of this study lies precisely in

constructing data-driven physical prior features directly from measured profile data, rather than relying on pre-generated physical simulation data.

Profile correction and degradation ranking: Based on the profile of Stage 1, the installation deviation is eliminated through a translation search algorithm, converting the original measurement values of each stage into relative wear depth. Using wear area as the degradation indicator, the experimental data across the 14 degradation stages are automatically ranked according to the principle of monotonic increase, resulting in 14 degradation stages. The ranked wear area increases monotonically from 0 $\mu\text{m}\cdot\text{mm}$ to 2215.62 $\mu\text{m}\cdot\text{mm}$, which is completely consistent with the physical order recorded in the experiment, reflecting the rationality of this indicator as a measure of degradation degree.

Degradation interval division and profile evolution: The four degradation intervals (Stages 1-4, 5-8, 9-10, and 11-14) were determined solely from the stage-wise wear-area growth rates computed on the training-set stages (Stages 1, 3, 4, 6, 7, 9, 10, 12, and 13).

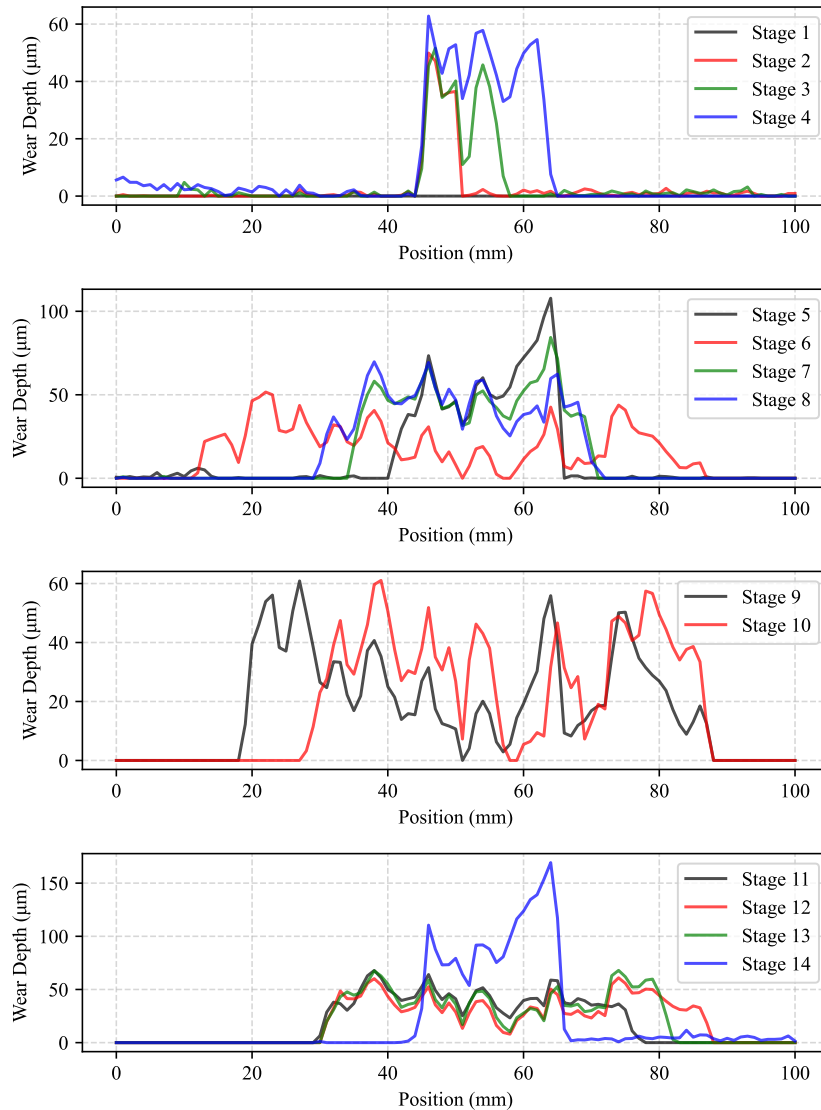


Figure 3. Evolution of interval profiles with different evolutionary characteristics

Figure 3 shows the continuous profile evolution within each interval:

Initial wear period (Stage 1–4): The wear area slowly increases, with an average of $419.21 \mu\text{m}\cdot\text{mm}$, and the profile shows local slight depressions;

Early expansion stage (Stage 5–8): The length of the wear area rapidly extends towards both ends, with the average area increasing to $1626.16 \mu\text{m}\cdot\text{mm}$ and the peak depth exceeding $100 \mu\text{m}$;

Mid-term stable phase (Stage 9–10): The expansion rate slows down, with an average area of $1834.59 \mu\text{m}\cdot\text{mm}$. The profile exhibits a "flat-bottomed" feature, and the depth slightly decreases;

Late accelerated stage (Stage 11–14): The wear area further expands and the depth increases again, with an average area reaching $2058.52 \mu\text{m}\cdot\text{mm}$, and the depth sharply increases to $169 \mu\text{m}$ in the final stage.

It can be seen that the increase in wear area is attributed to both the continuous expansion of the damaged area along the length of the guide rail and the contribution of a sudden increase in depth during the final stage. The four intervals were determined by identifying the change points in the stage-wise wear-area growth rate computed exclusively from the training-set stages, so that each interval corresponds to a distinct physical stage of wear evolution. This partitioning is therefore data-driven but derived only from training data, avoiding information leakage from the test stages. The Rank Wear Profiles are computed offline from the training-set wear profiles only, and the resulting physical prior remains fixed during inference without online updating.

(2) Physical prior features generation

Statistical calculations of physical prior features are only performed during the training set stages (Stage 1, 3, 4, 6, 7, 9, 10, 12, 13), and the test set stages (Stage 2, 5, 8, 11, 14) do not participate in any statistical calculations.

Firstly, calculate the average wear area within each degradation interval based on the wear area during the training set stage in that interval:

Initial wear period (training set includes Stage 1, 3, 4): average area is $468.73 \mu\text{m}\cdot\text{mm}$;

Early expansion stage (training set includes Stage 6, 7): average area is $1647.11 \mu\text{m}\cdot\text{mm}$;

Mid-term stable period (training set includes Stage 9 and 10): average area is $1834.59 \mu\text{m}\cdot\text{mm}$;

Late acceleration phase (training set includes Stage 12, 13): average area is $2055.90 \mu\text{m}\cdot\text{mm}$.

A mapping function from the degradation stage to the average wear area within a certain interval is established, assigning each sample in the dataset a corresponding average area within the interval. Subsequently, normalize these average areas by dividing them by the global maximum wear area in the training set. The maximum wear area among the training stages occurs at Stage 13, with a value of $2084.53 \mu\text{m}\cdot\text{mm}$.

The ultimately constructed one-dimensional coarse-grained interval-based physical prior feature provides coarse-grained guidance on the degradation stage for the model - it informs the model of the approximate degradation stage of the current sample, but does not provide precise continuous stage values, thereby forcing the model to combine vibration features to correct individual differences and achieve an organic combination of physical trends and data-driven methods.

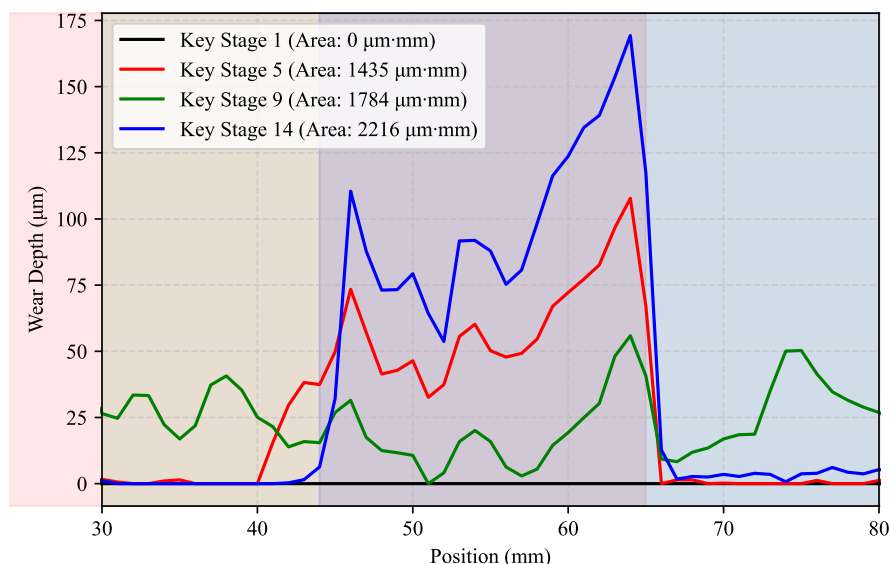


Figure 4. Comparison of wear profiles in four key degradation stages

Comparison of profiles at key degradation stages: Figure 4 focuses on comparing the profiles at four key stages: Stage 1, 5, 9, and 14. From this, it can be clearly observed that: The

peak wear depth of Stage 5 has reached $107.8 \mu\text{m}$, with damage concentrated in the 40–65 mm range; The depth of Stage 9 decreased to $60.9 \mu\text{m}$, but the length of the wear area

expanded to approximately 30–75 mm; The depth of Stage 14 increases sharply to 169.3 μm , with the wear zone covering almost 30–80 mm.

This comparison further reveals the compound physical law of "acceleration after depth saturation" and "continuous length expansion", providing a direct basis for the subsequent physical constraint.

2.3 Dual-channel feature fusion

Based on the two parallel feature-construction branches described above, the 16-dimensional vibration representation obtained through standardization and PCA is concatenated with the 1-dimensional physical prior derived from the wear-profile evolution. This fusion produces a 17-dimensional unified input vector, which is subsequently fed into the prediction model. By combining data-driven vibration information with coarse-grained physical guidance on the degradation stage, the proposed dual-channel representation enables the model to capture weak degradation characteristics while maintaining consistency with the overall wear-evolution trend. This fused representation therefore provides the feature basis for the subsequent lightweight residual network and physics-informed modeling.

2.4 Lightweight residual network

Residual architectures can facilitate gradient propagation and improve the learning of degradation-related representations [15]. Therefore, an LRN is employed to extract weak wear-related information from the fused features. The 17-dimensional fused input vector is first projected into a 64-dimensional latent representation through an input layer consisting of a fully connected layer, batch normalization (BN), and a ReLU activation function:

$$h_0 = \text{ReLU}(\text{BN}(W_{\text{in}}x + b_{\text{in}})) \quad (1)$$

where $x \in \mathbb{R}^{17}$ denotes the fused input feature vector and $h_0 \in \mathbb{R}^{64}$ denotes the latent feature representation. Here, W_{in} and b_{in} denote the weights and bias of the input projection layer, while $W_{l,1}$, $b_{l,1}$, $W_{l,2}$, and $b_{l,2}$ denote the parameters of the l -th residual block. The subscript l is used only for the input projection layer, and the subscript l is used for the residual blocks, so that the naming convention remains consistent across layers.

The network subsequently employs two lightweight residual blocks. Each residual block contains two fully connected layers with 64 hidden units. The first linear transformation is followed by BN and ReLU activation, whereas the second linear layer is directly connected to the residual addition without an additional BN layer. For the l -th residual block, the residual mapping is expressed as:

$$\mathcal{F}_l(h_l) = W_{l,2}\text{ReLU}(\text{BN}(W_{l,1}h_l + b_{l,1})) + b_{l,2} \quad (2)$$

and the block output is given by

$$h_{l+1} = \text{ReLU}(\mathcal{F}_l(h_l) + h_l) \quad (3)$$

Here, $\mathcal{F}_l(\cdot)$ denotes the nonlinear residual transformation. The identity skip connection preserves the original feature representation and facilitates gradient propagation, allowing the network to learn subtle degradation-related deviations from weak vibration features without substantially increasing model complexity. After the two residual blocks, a single fully connected output layer maps the learned representation to the predicted wear area:

$$\hat{y} = W_{\text{out}}h_2 + b_{\text{out}} \quad (4)$$

The detailed architecture and configuration of the proposed LRN are presented in Table 1.

Table 1. Detailed architecture of the proposed LRN

Component	Detailed configuration
Input	17 features
Input projection	Linear (17, 64)
Normalisation	BatchNorm1d (64)
Activation	ReLU
Residual block	Linear (64,64) $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Linear (64,64) $\rightarrow$ skip addition $\rightarrow$ ReLU
Hidden dimension	64.0
Output layer	Linear (64,1)

2.5 APC and monotonicity post-processing

Inspired by the paradigm that is capable of learning and automatically balancing multiple loss terms [30], an APC method is proposed by integrating data fitting, a monotonicity constraint on the cumulative wear area, and a training-data-derived wear prior into a unified objective. The contributions of these loss terms are automatically balanced through learnable log-variance parameters, enabling the model to achieve accurate predictions while maintaining physically consistent degradation behavior. Moreover, a quadratic degradation prior with a fixed scale coefficient determined exclusively from the training data is introduced to explicitly guide the predicted wear trajectory and reduce physically implausible deviations. The expression of the total loss function is as follows:

$$L_{\text{total}} = e^{-s_{\text{mse}}}L_{\text{mse}} + s_{\text{mse}} + e^{-s_{\text{mono}}}L_{\text{mono}} + s_{\text{mono}} + e^{-s_{\text{phys}}}L_{\text{phys}} + s_{\text{phys}} \quad (5)$$

where L_{mse} , L_{mono} , and L_{phys} denote the data-fitting mean squared error loss, monotonicity constraint loss, and physical prior loss, respectively. The parameters s_{mse} , s_{mono} , and

s_phys are learnable log-variance parameters associated with the corresponding loss components. The terms $\exp(-s_mse)$, $\exp(-s_mono)$, and $\exp(-s_phys)$ serve as adaptive weights, whereas the additive terms s_mse , s_mono , and s_phys arise from the uncertainty-based weighting formulation and prevent degenerate solutions. During training, these parameters converge to stable finite values without degeneracy.

The parameters s_mse , s_mono , and s_phys are global, input-independent trainable scalars shared across all input samples, rather than sample-specific quantities. They are initialised as 0 and jointly optimised with the network weights through backpropagation using AdamW.

$$L_{mse} = 1/N \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (6)$$

$$L_{mono} = \frac{1}{N-1} \sum_{i=1}^{N-1} \max(0, -(\hat{y}_{(i+1)} - \hat{y}_{(i)})) \quad (7)$$

$$L_{phys} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\hat{y}_i}{y_{prior}(s_i)} - 1 \right)^2 \quad (8)$$

In Eq. (8), the physical prior is defined as:

$$y_{prior}(s) = \beta \times (s/S_{max})^2 \quad (9)$$

where $\beta = 2084.53$ is determined from the training data, corresponding to the maximum wear area observed among the training stages. It should be noted that the quadratic form of $y_{prior}(s)$ reflects the physical evolution pattern of wear, whereas the coefficient β is a data-derived degradation reference rather than a purely physical parameter. Thus, the proposed prior combines a physically grounded evolution trend with a data-calibrated magnitude. The variables s and S_{max} denote the degradation stage and the maximum degradation stage ($S_{max} = 14$), respectively. For the i -th sample, s is replaced by s_i .

After model training, Post-processing via PCM [31] is further applied to suppress minor non-physical fluctuations in the stage-wise predictions. Specifically, the mean predicted wear area is first calculated for each degradation stage, and the cumulative maximum is then taken along the degradation-stage direction. Thus, the corrected stage-wise prediction is expressed as $\bar{y}_s = \max_{1 \leq k \leq s} [\bar{y}_k]$, where $\bar{y}_k$ denotes the mean prediction at the k -th degradation stage and $\bar{y}_s$ denotes the PCM-corrected prediction at stage s . PCM only corrects minor non-physical fluctuations and does not alter the overall degradation trend learned by the model.

2.6 Pseudo-Bayesian physics-informed uncertainty quantification

Based on the optimal LRN+APC model, a pseudo-Bayesian physics-informed framework is constructed to quantify predictive uncertainty. Dropout layers are inserted after the input projection layer and after the first linear layer of each residual block. During stochastic inference, Dropout remains activated, and T stochastic forward passes are performed for

each input sample. The resulting prediction samples are used to estimate the MC predictive mean $\hat{\mu}_i$ and MC predictive standard deviation $\hat{\sigma}_{i,MC}$ [32] as follows:

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^T \hat{y}_i^{(t)} \quad (10)$$

$$\hat{\sigma}_{i,MC} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (\hat{y}_i^{(t)} - \hat{\mu}_i)^2} \quad (11)$$

where $\hat{y}_i^{(t)}$ denotes the prediction of the i -th sample obtained from the t -th stochastic forward pass.

To mitigate the overconfidence that may arise under weak-signal conditions, a two-stage model-selection strategy is further introduced. In Phase 1, the training Dropout rate is selected using the validation set. In Phase 2, based on the model selected in Phase 1, the inference Dropout rate and a global standard-deviation scaling factor α are jointly selected.

For each candidate configuration, Prediction Interval Coverage Probability (PICP), Mean Prediction Interval Width (MPIW), and Negative Log-Likelihood (NLL) [33-35] are calculated on the validation set.

A hierarchical selection criterion is adopted. Candidate configurations satisfying $|PICP - 0.95| \leq 0.03$ are given the highest priority. If no candidate satisfies this condition, configurations with $PICP \geq 0.90$ are considered. If neither condition can be satisfied, the configuration with the highest PICP is selected. Within the same priority level, the configuration with the smallest MPIW is preferred; if the MPIW values are identical, the one with the lower NLL is selected.

After the inference Dropout rate is determined, the calibrated predictive standard deviation $\hat{\sigma}_{i,cal}$ is obtained by scaling the MC predictive standard deviation with the global standard-deviation scaling factor α :

$$\hat{\sigma}_{i,cal} = \alpha \hat{\sigma}_{i,MC} \quad (12)$$

Through this two-stage selection and global uncertainty calibration procedure, the proposed framework simultaneously provides point estimates and calibrated prediction intervals while balancing interval coverage and compactness.

3. Experiments and results

3.1. Experimental setup

3.1.1 Data partitioning and validation protocol

To evaluate the generalization ability of the model throughout the entire life cycle of the guide rail, this study employs a multi-stage hold-out method to partition the dataset. Five stages, specifically Stage 2, 5, 8, 11, and 14, are selected as the test set, encompassing early, middle, and late stages of degradation; the remaining nine stages (Stage 1, 3, 4, 6, 7, 9, 10, 12, and 13) serve as the training set. This partitioning approach ensures that the test set covers different stages of the entire degradation process, thereby comprehensively

evaluating the model's predictive performance under varying degrees of degradation.

All models utilize the identical training and testing sets to ensure fairness in comparison. Furthermore, to eliminate dimensional differences among various features, the input features undergo standardization (with a mean of 0 and a standard deviation of 1) prior to training. The standardization parameters are solely fitted from the training set and subsequently applied to the testing set.

For hyperparameter calibration of the uncertainty-quantification framework, an internal validation subset is further separated from the training data. Specifically, Stages 1, 4, 7, 10, and 13 are used for validation during the two-stage Dropout and uncertainty-scaling selection procedure. The held-out test stages (Stages 2, 5, 8, 11, and 14) are not involved in hyperparameter selection or uncertainty calibration and are used only for final performance evaluation.

3.1.2 Evaluation metrics

Let y_i represent the actual wear area of the i -th sample, $\hat{y}_i$ the corresponding predicted value, $\bar{y}$ the mean of all actual values, and N the total number of samples. This study employs the following three types of indicators to comprehensively evaluate model performance:

(1) Point prediction indicators

Three indicators, namely root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), are used to measure prediction accuracy [36, 37]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (15)$$

RMSE: It assigns higher weights to larger errors, with the unit consistent with the wear area (i.e., $\mu\text{m} \cdot \text{mm}$).

MAE: It reflects the average magnitude of prediction errors, with the same unit as RMSE.

R^2 : It measures the extent to which a model explains the variance of the target variable. The closer the value is to 1, the better the fit.

(2) Error robustness

To measure the error robustness (E_{Rob}) of the model, the normalized complementary value of the average absolute error relative to the full range of wear area is calculated according to formula (16). The value range is $(-\infty, 1]$, where a value closer to 1 indicates a smaller error and better robustness.

$$E_{\text{Rob}} = 1 - (MAE / \max_{\text{wear}}) \quad (16)$$

where $\max_{\text{wear}} = 2215.62 \mu\text{m} \cdot \text{mm}$ denotes the maximum observed wear area over the complete 14-stage dataset.

(3) Uncertainty quantification indicators

PICP: It measures the proportion of true values falling within the 95% prediction interval. The ideal value should be close to the nominal coverage level (0.95).

$$PICP = \frac{1}{N} \sum_{i=1}^N I[y_i \in [\hat{y}_i - 1.96\sigma_i, \hat{y}_i + 1.96\sigma_i]] \quad (17)$$

where σ_i is equal to the $\hat{\sigma}_{(i, \text{cal})}$ given in Equation (12), and $I(\cdot)$ is an indicator function.

MPIW: The average width of the prediction interval, which should be as small as possible while ensuring coverage.

$$MPIW = \frac{1}{N} \sum_{i=1}^N (Upper_i - Lower_i) \quad (18)$$

where $Upper_i = \hat{y}_i + 1.96\sigma_i$, $Lower_i = \hat{y}_i - 1.96\sigma_i$. Under the premise of satisfying coverage, MPIW should be as small as possible.

NLL: NLL based on the Gaussian distribution assumption, where a lower value indicates better probability calibration.

$$NLL = -\frac{1}{N} \sum_{i=1}^N \log[\mathcal{N}(y_i | \hat{y}_i, \sigma_i^2)] \quad (19)$$

3.1.3 Implementation and hyperparameter settings

To ensure the fairness and reproducibility of the experiment, all methods involving neural networks adopt the following unified configuration:

Optimizer: AdamW, learning rate: 0.001, weight decay: 1×10^{-4} . Learning rate scheduling: Cosine Annealing. Training epochs: 800 epochs. Batch size: 128. Random seed: Both PyTorch and NumPy are fixed to 42.

Traditional machine learning methods (such as XGBoost, LightGBM, Gradient Boosting, etc.) all utilize default hyperparameters, with the number of trees uniformly set to 100 and the learning rate set to 0.1. All experiments were conducted in the same computing environment equipped with an Intel Core i7 CPU and an NVIDIA GeForce RTX GPU.

For the pseudo-Bayesian uncertainty-quantification experiments, $T=50$ stochastic forward passes are performed for each sample during MC inference. In Phase 1 of the two-stage selection procedure, six equally spaced candidate values within $[0.1, 0.7]$ are considered for the training Dropout rate. In Phase 2, six equally spaced candidate values within $[0.1, 0.7]$ are considered for the inference Dropout rate, while fifteen equally spaced candidate values within $[0.1, 3]$ are considered for the global standard-deviation scaling factor α . Hyperparameter selection is performed according to the hierarchical PICP-MPIW-NLL criterion described in Section 2.6.

The computational overhead of the pseudo-Bayesian inference was also considered. The LRN+APC model is lightweight, with only two residual blocks and 64 hidden units, and the input dimension is only 17. With $T = 50$ stochastic forward passes, the total inference cost remains modest. Since PHM decisions are typically made on a periodic basis rather than in hard real-time, this latency is acceptable for condition-based maintenance and RUL-oriented planning.

3.2. Ablation experiment

3.2.1 BL (Baseline) model

To evaluate the independent contributions of each module, we first constructed a naive baseline model (BL). This model employs a simple fully connected neural network with two hidden layers (dimensions 64 and 32, respectively), uses mean squared error (MSE) as the loss function, and applies PCM post-processing, but incorporates no physical constraints.

3.2.2 BL + LRN and BL + LRN - PCM

The baseline network is replaced with the lightweight residual network (LRN) described in Section 2.4 to investigate whether residual connections can help extract richer degradation information from weakly correlated vibration features. All other settings remain identical to those of the baseline. Furthermore, to isolate the effect of PCM, we remove the PCM unit from model (BL+LRN-PCM) and

compare its prediction accuracy with the version that retains PCM.

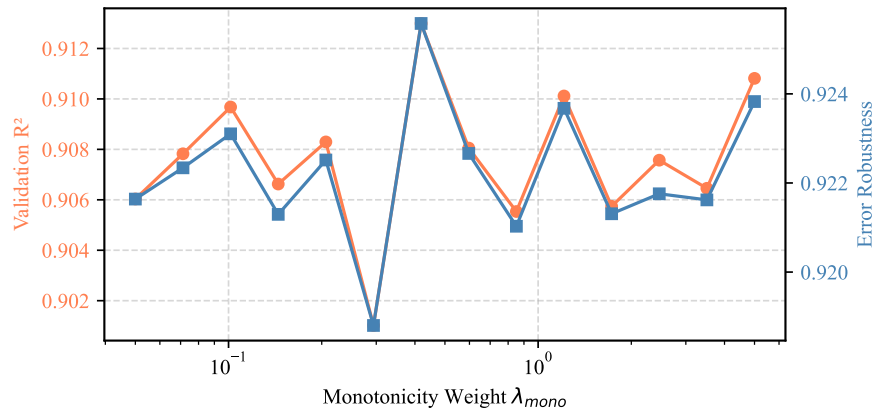
3.2.3 Sensitivity analysis of physical information weight

To investigate the impact of physical constraints on model performance, monotonicity constraint and quadratic degradation prior are explicitly incorporated into the loss function of the baseline model. By fixing the weight of one item and scanning the weight of the other, the influence patterns of both on the prediction results are examined.

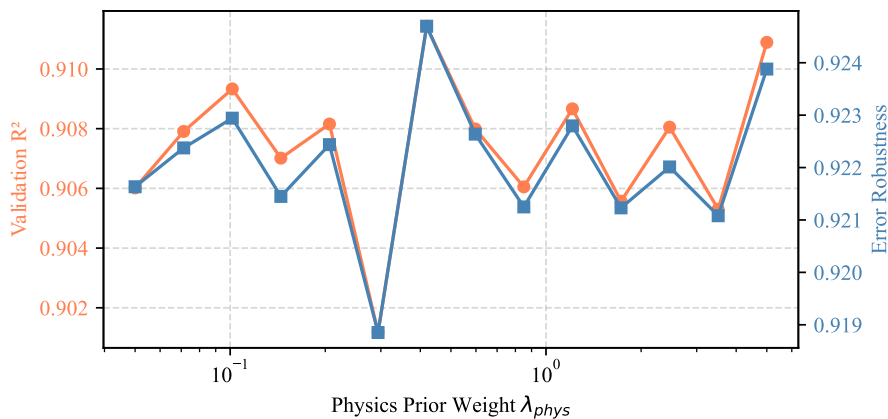
Monotonicity weight λ_{mono} sensitivity: Fix the quadratic degradation prior weight $\lambda_{phys}=1.0$ and take 14 values at logarithmic equal intervals within the interval $[0.05, 5.0]$.

Sensitivity of the λ_{phys} : Fix the $\lambda_{mono}=1.0$ and similarly scan 14 logarithmically equally spaced points within the range of $[0.05, 5.0]$.

Through the aforementioned weight scanning, the results of the sensitivity analysis are shown in Figure 5.



(a) Monotonicity weight sensitivity



(b) Quadratic degradation prior weight sensitivity

Figure 5. Sensitivity analysis of physical information weights

3.2.4 BL + APC and BL + APC - PCM

The sensitivity analysis clearly indicates that determining fixed optimal weights requires a tedious grid search. To address this limitation, we introduce the Adaptive Physical Constraints (APC) described in Section 2.5 into the baseline model. APC treats each loss term as an independent Gaussian likelihood and automatically infers reasonable weights via learnable log-variance parameters. We compare two variants: BL+APC (with PCM) and BL+APC-PCM (without PCM) to assess the individual contributions of APC and PCM to prediction accuracy.

3.2.5 BL + LRN + APC

Based on the independent effectiveness of residual networks (Section 3.2.2) and adaptive physical constraints (Section 3.2.4), we construct a model that combines both LRN and APC while retaining PCM (denoted as BL+LRN+APC). This is our proposed model.

3.2.6 BL + LRN + APC + Disent

To further examine whether feature disentanglement can improve prediction under weak wear-related vibration characteristics, a disentanglement variant is constructed on the basis of BL + LRN + APC. This configuration, denoted as BL + LRN + APC + Disent, introduces separate degradation-related and operating-condition-related feature branches while keeping the remaining experimental settings unchanged. Specifically, the network consists of a shared input projection followed by two parallel branches. The shared projection is a fully connected layer with 64 hidden

units, batch normalization, and ReLU activation. Each branch contains two lightweight residual blocks with 64 hidden units. The degradation-related branch outputs the predicted wear area, while the operating-condition-related branch outputs the predicted degradation stage as an auxiliary task. The total loss extends the APC loss by adding an auxiliary stage prediction loss and the disentanglement regularization, and all loss terms are adaptively balanced by learnable log-variance parameters. The purpose of this experiment is to evaluate whether explicit feature separation provides additional benefits beyond the proposed LRN-APC configuration.

The results of the ablation experiment are summarized in Table 2.

3.3 Comparison with representative learning models

To comprehensively evaluate the performance of our method, this section compares it with other representative models.

Gradient Boosted Decision Trees (GBDT) methods: Gradient Boosting, XGBoost, LightGBM.

Deep network model: TabNet (deep network with sparse attention).

Variants of attention mechanisms: Self-Attention, Squeeze-and-Excitation (SE) attention.

Method in this paper (Ours): Based on the optimal model obtained from ablation experiments. The comparison summary of modern methods is summarized in Table 3.

Table 2. Summary of ablation experiments

Experimental methods	R ²	MAE ($\mu\text{m}\cdot\text{mm}$)	RMSE ($\mu\text{m}\cdot\text{mm}$)	E_Rob
BL	0.7992	269.48	301.73	0.8784
BL + LRN	0.9223	159.57	187.66	0.9280
BL + LRN-PCM	0.8217	197.15	284.31	0.9110
BL + APC	0.9304	164.20	177.64	0.9259
BL + APC-PCM	0.8979	183.17	221.35	0.9173
BL + LRN +APC (Ours)	0.9553	101.70	143.26	0.9541
BL + LRN +APC +Disent	0.9325	166.75	174.90	0.9247

Table 3. Comparison summary of representative learning models

Methods	R ²	MAE ($\mu\text{m}\cdot\text{mm}$)	RMSE ($\mu\text{m}\cdot\text{mm}$)	E_Rob
Gradient Boosting	0.9305	169.09	177.56	0.9237
XGBoost	0.9371	160.77	168.85	0.9274
LightGBM	0.936	160.39	170.29	0.9276
TabNet	0.9225	160.71	187.47	0.9275
SE Attention	0.9444	144.81	158.71	0.9346

Self-Attention	0.9215	176.52	188.64	0.9203
Ours	0.9553	101.70	143.26	0.9541

3.4 Uncertainty quantification experiments and results

This section experimentally evaluates the uncertainty-quantification framework described in Section 2.6. The analysis focuses on the calibration performance of the proposed method, the visualization of predictive uncertainty, and comparison with representative uncertainty-quantification approaches.

3.4.1 Calibration and test results

The uncertainty-quantification model is calibrated on the validation set according to the two-stage selection procedure described in Section 2.6 and the hyperparameter settings specified in Section 3.1.3. After the calibration procedure is completed, the selected configuration is directly applied to the held-out test set without further parameter adjustment.

On the test set, the proposed method achieves a PICP of 0.9460, an MPIW of 803.26 $\mu\text{m}\cdot\text{mm}$, and an NLL of 6.8974. The PICP is close to the nominal coverage level of 0.95, indicating that the calibrated prediction intervals achieve the intended coverage level.

3.4.2 Visualization of uncertainty quantification prediction

Figure 6 presents the stage-wise predictive mean and calibrated 95% prediction band obtained using the complete framework. The true mean wear area, predicted mean wear area, and calibrated prediction band are shown for comparison. Figure 7 further presents the distribution of the calibrated predictive standard deviation across the evaluated samples. These visualizations provide complementary information on the stage-wise prediction behavior and the distribution of predictive uncertainty.

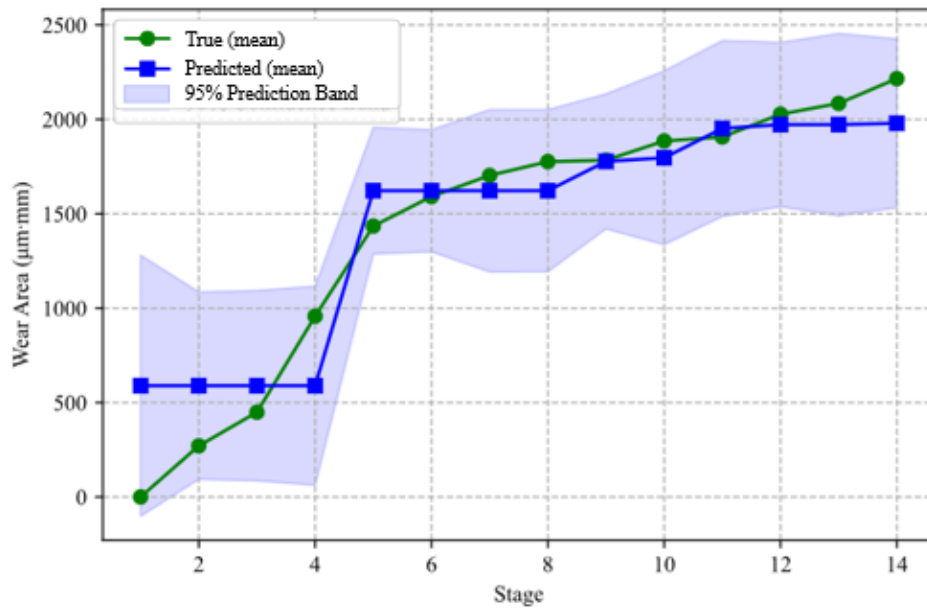


Figure 6. Stage-wise uncertainty prediction results and 95% prediction band across all degradation stages

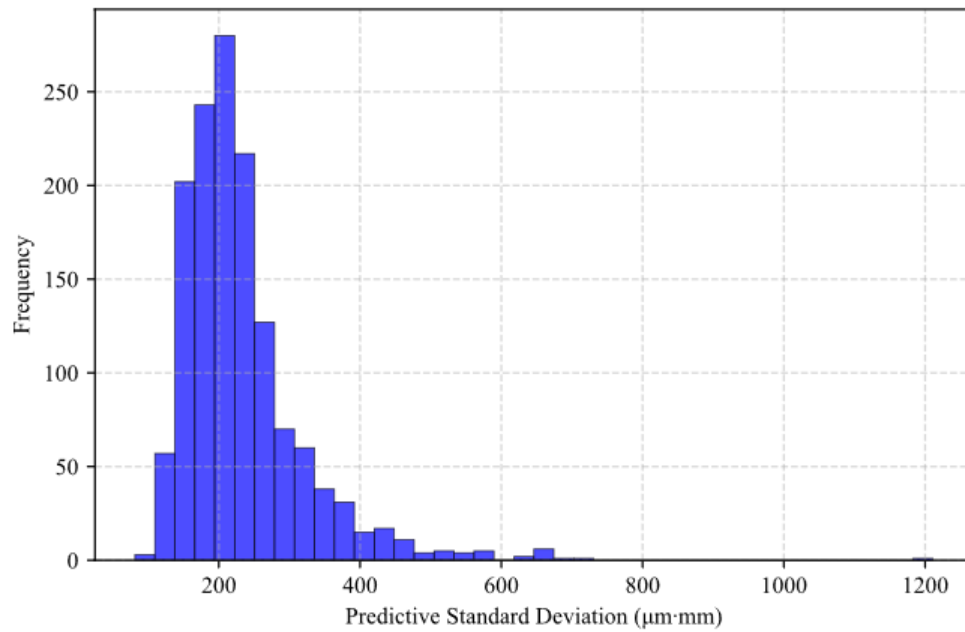


Figure 7. Histogram of the Predicted Standard Deviation After Calibration for All Samples

3.4.3 Comparison with representative uncertainty-quantification methods

To comprehensively evaluate the performance of the proposed uncertainty quantification method, five representative uncertainty estimation methods are selected for comparison in this section.

(1) Quantile Gradient Boosting:

Using quantile loss as the objective function, two independent models are trained to predict $\alpha=0.025$ (low quantile) and $\alpha=0.975$ (high quantile), respectively. To reduce training overhead, the midpoint of the lower and upper quantile predictions is used as an approximate central prediction. Post-calibration employs the residual quantile method: the residuals are calculated by subtracting this central prediction from the true values on the validation set, and the 2.5% and 97.5% quantiles of the residuals are taken as offsets. The original quantile interval of the test set is then adjusted using these offsets to obtain the calibrated prediction interval.

(2) Deep Ensemble:

Epistemic uncertainty is quantified by training five independently initialized deterministic neural networks with the same structure. All ensemble members are trained with the same APC loss. However, because the physical knowledge is introduced only through the loss function and is not embedded in the network architecture or governing equations, these models are referred to as deterministic physics-constrained neural networks rather than PINNs. During inference, the mean of the predictions from the five models is used as the point estimate, and the variance between models is taken as the epistemic uncertainty. Meanwhile, the average variance of the prediction residuals of each model on the validation set is used to estimate the aleatoric uncertainty.

After combining epistemic and aleatoric uncertainties, the optimal global scaling factor is searched on the validation set to scale the total standard deviation, and this factor is applied to the test set to generate the final prediction interval. This is a bulleted list.

(3) SE-Attention + MC Dropout

The SE-Attention is constructed by introducing the SE attention mechanism into the basic residual network. Each residual block is followed by an SELayer, and Dropout layers are added after the input layer and the first linear layer of each residual block. MC Dropout is performed during both training and inference. A joint search is conducted on the inference dropout rate and global scaling factor on the validation set, selecting the combination that makes the PICP as close to 0.95 as possible. The prediction interval for the final test set is calculated from the mean after monotonicity processing and the scaled standard deviation.

(4) Self-Attention + MC Dropout

In the residual network, a multi-head Self-Attention layer is embedded to capture long-range dependencies between features. The network structure is as follows: after linear projection and batch normalization, the input is connected to two residual blocks through a multi-head self-attention layer. Dropout is set in the input projection layer, the first linear layer of each residual block, and within the self-attention layer. The inference and calibration strategies are completely consistent with (3).

(5) Variational Inference PINN (VI-PINN)

All linear layers in the deterministic network are replaced with Bayesian linear layers, approximating the posterior distribution of parameters through variational inference. Each Bayesian linear layer learns mean parameters and logarithmic variance parameters for weights and biases, and samples from

the parameter distribution during forward propagation. The prior adopts a mixture Gaussian distribution, and the KL divergence term constrains the distance between the posterior and the prior. During inference, variational randomness is maintained, and forward propagation is performed 50 times for each test sample, followed by direct calculation of the predicted mean and standard deviation.

The uncertainty quantification performance of all comparison methods on the test set is summarized in Table 4. In this study, the term PINN is reserved for models in which physical laws are embedded in the network architecture or governing equations. Models that incorporate physical knowledge only through the loss function are referred to as physics-constrained models.

Table 4. Comparison of Performance of Different Uncertainty Quantification Methods on the Test Set

Methods	PICP (target 0.95)	MPIW ($\mu\text{m}\cdot\text{mm}$)	NLL
Quantile Gradient Boosting	0.8100	1513.3200	—
Deep Ensemble	0.6070	428.7800	7.1186
SE-Attention + MC Dropout	0.9990	1186.1800	6.8454
Self-Attention + MC Dropout	1.0000	1415.6300	6.9580
VI-PINN	0.7040	678.4800	7.5895
Ours	0.9460	803.26	6.8974

4. Discussion

The following discussion is structured around three dimensions: the integration mechanism of ablation experiments and physical constraints, a comparison with modern data-driven methods, and pseudo-Bayesian uncertainty quantification.

4.1 Ablation experiment analysis and physical constraint adaptive fusion mechanism

The progressive ablation experiment (Table 2) quantifies the independent contributions and synergistic effects of each core module on the prediction accuracy of the wear area. BL relies solely on a fully connected network, and a data-driven mean squared error loss. Under the condition where the original vibration features have a very weak linear correlation with the target, the R^2 is only 0.7992, and the MAE reaches 269.48 $\mu\text{m}\cdot\text{mm}$. This indicates that relying solely on statistical features is highly susceptible to interference from operational condition noise and multicollinearity, causing the model to fall into local optima.

After introducing the LRN with skip connections, the performance of BL+LRN achieves a qualitative leap, with R^2 rising to 0.9223 and MAE plummeting by over 40%. The residual structure effectively mitigates the vanishing gradient problem, enabling the network to extract weak degradation trend signals from a strong noise background. Comparing the results of BL+LRN with BL+LRN-PCM shows that PCM has a significant positive effect on the prediction monotonicity correction of the residual network, eliminating minor non-physical fluctuations caused by vibration feature noise. Even with the shallow network unchanged and only APC (BL + APC) applied, R^2 increases to 0.9304. It is worth noting that further comparing BL+APC with BL+APC-PCM reveals that R^2 decreases from 0.9304 to 0.8979, corresponding to an

absolute decrease of 0.0325, or approximately 3.49% relative to the BL+APC result; MAE increases from 164.20 to 183.17. This indicates that PCM post-processing has a positive effect, but its influence significantly weakens when APC is introduced - APC itself has already increased R^2 from the baseline of 0.7992 to 0.9304 (an absolute increase of 0.1312, or approximately 16.4% relative to the baseline), far exceeding the decrease brought about by removing PCM. This further confirms the dominant role of adaptive physical constraints in weak feature scenarios, with PCM serving only as an auxiliary correction method.

The weight sensitivity analysis (Figure 5) reveals the inherent flaws of the fixed weight mechanism. The APC mechanism automatically infers the optimal weights for multi-objective losses through learnable logarithmic variance parameters, completely eliminating the need for manual grid search. The R^2 of BL + APC even surpasses that of the optimal fixed weight model obtained through fine scanning, highlighting the scale matching advantage of uncertainty weighting in multi-task optimization.

When LRN and APC are synergistically integrated (BL + LRN + APC), the optimal performance of the entire group is achieved: $R^2 = 0.9553$, MAE = 101.70 $\mu\text{m}\cdot\text{mm}$, and error robustness E_{Rob} is 0.9541. This demonstrates a significant complementary synergistic effect between deep feature extraction capability and adaptive physical trend guidance - LRN focuses on mining degradation patterns hidden in noise, while APC ensures that these patterns are extrapolated on a manifold that conforms to the physical evolution laws of the guideway.

In contrast, the performance declines after introducing the feature decoupling module (BL + LRN + APC + Disent). This can be attributed to two factors: Firstly, the decoupling regularization term forces the degradation-related features to be orthogonal to the condition features, which compresses the rank of effective representations, resulting in the degradation-related branch losing some vibration information that is

weakly correlated with the wear area but indispensable. Secondly, there is nonlinear interference between the auxiliary task of the condition branch and the main regression task, and the additional MSE loss diverts the network's fitting attention away from the wear area. This indicates that in the scenario of this study, a unified input of dual channels is sufficient for the model to autonomously coordinate condition fluctuations. When training data are scarce for a new system, the data-derived degradation reference β can be estimated from (i) a similar guide system with available degradation data, (ii) a physics-based wear model calibrated with limited measurements, or (iii) expert knowledge of the allowable wear limit. The quadratic evolution pattern is expected to remain valid, while the magnitude coefficient may require transfer learning or Bayesian calibration with a small number of observations.

4.2 Comparative analysis with modern data-driven methods

The horizontal comparison in Table 3 further confirms the unique value of physics-informed guidance from an architectural perspective. Gradient boosting tree models, represented by XGBoost and LightGBM, exhibit robust nonlinear fitting capabilities (R^2 approximately 0.93–0.94). However, their nature based on stepwise spatial partitioning lacks continuous derivative information, making it difficult to spontaneously perceive the complex degradation patterns of "acceleration after deep saturation" and "continuous length expansion" when faced with multi-stage testing throughout the entire lifecycle. The deep tabular model TabNet performs mediocrely ($R^2 = 0.9225$), and its sparse attention mechanism lacks stability in feature selection on low-dimensional strong noise feature streams. The introduction of Self-Attention significantly deteriorates performance ($R^2 = 0.9215$), indicating that the long-range dependency assumption is prone to overfitting on tabular features after PCA dimensionality reduction and recombination. Conversely, the SE Attention, which focuses on feature channel weighting, achieves superior results ($R^2 = 0.9444$), revealing that dynamically emphasizing weights on highly correlated principal components has higher information focusing efficiency than spatial global transformations.

The positioning of the proposed framework can be further clarified through comparison with representative previous studies. Tao et al. [18] established a physics-based wear prediction model for roller linear guides based on contact mechanics and wear theory, demonstrating the effectiveness of explicit mechanical modeling for guideway wear estimation. Staroszyk et al. [16] further considered real operating conditions by combining damage accumulation with dynamic model-based condition monitoring for linear-guideway service-life prediction. These studies provide strong physical interpretability, whereas their implementation relies substantially on explicit mechanical or dynamic modeling. In contrast, Nie et al. [21] proposed a vibration-based surface wear-state identification method for linear rolling guides using multi-scale fuzzy entropy and an

optimized random forest, achieving a recognition accuracy of 97.9%. Their work demonstrates the effectiveness of data-driven vibration analysis for discrete wear-state identification, while the present study focuses on continuous wear-area regression under weak-feature conditions.

From the perspective of physics-data fusion, Gao et al. [28] developed a hybrid physics-data-driven framework for machining tool wear prediction by integrating physical information with sensor-derived features and combining prediction results through Bayesian fusion. More recently, Zhang et al. [31] investigated a physics-informed Transformer framework for rolling-bearing RUL prediction, representing the growing trend of embedding physical degradation knowledge into deep-learning prognostic models.

Compared with these approaches, the present framework constructs a physical prior directly from measured guideway wear profiles and integrates it with weak vibration features, while adaptive physical constraints jointly regulate data fitting, cumulative wear-area monotonicity, and the degradation prior. In addition, the proposed pseudo-Bayesian calibration provides uncertainty-aware wear predictions, achieving an R^2 of 0.9553 and a PICP of 0.9460 for the 95% prediction interval. Since the compared studies involve different components, degradation targets, datasets, and evaluation metrics, their numerical results should not be interpreted as a direct performance ranking. Instead, this comparison highlights the contribution of the present study in combining continuous wear estimation, measured-profile-derived physical guidance, adaptive physical consistency, and calibrated predictive uncertainty.

It should be noted that Transformer-based time-series models such as Informer, Autoformer, and PatchTST are designed for long-sequence forecasting with high-dimensional temporal inputs. In contrast, the present task is a low-dimensional tabular regression problem with only 17 input features. Directly applying such models would require reshaping tabular features into artificial sequences, which may introduce inductive bias and overfitting. Therefore, this study compares with representative tabular and attention-based models, including TabNet, Self-Attention, and SE Attention, which are more suitable for the present setting.

4.3 Pseudo-Bayesian uncertainty quantification and engineering applicability

The proposed pseudo-Bayesian physics-informed framework achieved a PICP of 0.9460 on the test set, which is close to the nominal coverage level of 0.95. The MPIW is tightly controlled at $803.26 \mu\text{m} \cdot \text{mm}$ (approximately 36.2% of the full range). This high-quality probabilistic calibration stems from a two-stage hyperparameter search and global variance calibration strategy, which accurately balances coverage and interval width.

Comparative analysis of uncertainty quantification methods (Table 4): Quantile Gradient Boosting employs a tree model to directly fit conditional quantiles, eliminating the need for distribution assumptions. However, its PICP is only

0.810, indicating severe undercoverage, and the interval width reaches $1513.32 \mu\text{m}\cdot\text{mm}$, suggesting that tree models struggle to capture heteroscedasticity in scenarios with weak features and high noise. Residual calibration fails to fully compensate for inherent flaws. Deep Ensemble characterizes epistemic uncertainty through the variance of five independently initialized deterministic PINNs. However, its PICP is only 0.607, indicating substantial undercoverage. One possible explanation is that the ensemble members may exhibit similar dependence on the coarse-grained physical prior, resulting in limited inter-model diversity and consequently an underestimation of epistemic uncertainty represented by the inter-model variance. Both SE-Attention + MC Dropout and Self-Attention + MC Dropout achieve PICPs close to 1.000 (overcoverage), with excessively wide intervals (MPIW reaching 1186.18 and $1415.63 \mu\text{m}\cdot\text{mm}$, respectively), and NLLs of 6.8454 and 6.9580, respectively. This situation arises from the attention mechanism excessively amplifying minor fluctuations under weak feature conditions, leading to excessively high MC Dropout sampling variance. While ensuring safety, it sacrifices interval compactness. VI-PINN employs variational inference to approximate parameter posteriors, but its PICP is only 0.704, and it has the highest NLL (7.5895), indicating a mismatch between the Gaussian mixture prior and the true posterior, and excessive regularization of the KL divergence term under limited data, leading to underfitting. More specifically, the KL divergence term dominates the training objective and over-regularizes the posterior toward the Gaussian mixture prior, which suppresses the model's ability to fit the wear-area observations and results in underfitting. In addition, the Gaussian mixture prior may not match the true posterior of the wear prediction task, further degrading the probability calibration and increasing the NLL. In contrast, the proposed pseudo-Bayesian physics-informed framework achieves a PICP of 0.9460 through two-stage tuning and calibration (coverage priority + rational interval distance priority), which is closest to the nominal level of 0.95, with MPIW= $803.26 \mu\text{m}\cdot\text{mm}$ and NLL=6.8974. This verifies that in weak feature degradation prediction, physically guided variance calibration is more effective than complex attention or variational structures. Therefore, the method proposed in this paper strikes an optimal balance between coverage and compactness, making it more suitable for condition-based maintenance decisions and RUL-oriented maintenance planning.

5. Conclusion

To address the issue of wear degradation prediction for high-precision linear guides under weak feature conditions, this paper proposes a lightweight residual network framework that integrates physical prior information with data-driven learning. Ablation and comparative experiments demonstrate that, in industrial degradation prediction scenarios characterized by weak features and strong noise, the residual network is capable of extracting subtle degradation trend signals from a background of strong noise. Furthermore,

compared to the cumulative maximum post-processing method, the adaptive physical constraint mechanism remains the dominant factor in the model's predictive performance. From the perspective of manufacturing system resilience, the proposed method converts weak vibration and wear-profile information into actionable degradation knowledge. By enabling early identification of wear evolution, it supports maintenance scheduling, spare-part preparation, and risk-aware recovery planning, thereby helping to potentially reduce unplanned downtime and improve the rapid recovery capability of manufacturing systems. Despite the promising results, several limitations should be acknowledged. The proposed method relies on manual stage-level annotations, and the monotonic degradation assumption may not hold under abnormal wear, local spalling, lubrication failure, or maintenance interventions. In addition, its generalisability to other guide systems or wear scenarios has not been fully verified, since the current validation is limited to the NIST linear guide dataset and the data-derived coefficient beta may require re-calibration.

The current framework uses vibration signals and static wear profiles, while additional signals such as temperature and friction torque are often available in industrial environments. Future work will explore unsupervised or semi-supervised multi-source fusion to improve robustness with limited labeled degradation data and further embed the framework into shop-floor PHM and resilient manufacturing decision systems to link wear prediction with maintenance execution, production recovery, and system-level resilience optimization.

Acknowledgements

This work is supported by the Sichuan Engineering Re-search Center for Smart Operation and Maintenance of Rail Transit Lines (No. 2025GD-C08).

References

- [1] Li L, Yang H, Zhang Y, et al. Dynamic positioning error analysis and modeling of CNC machine tool guideway system. *Journal of Mechanical Science and Technology*. 2021; 35(5): 1955.
- [2] Kumar R, Sharma S, Kumar R, et al. Review of lubrication and cooling in computer numerical control (CNC) machine tools: A content and visualization analysis, research hotspots and gaps. *Sustainability*. 2023, 15(6): 4970.
- [3] Huang T, Kang Y, Du S, et al. "A survey of modeling and control in ball screw feed-drive system." *The International Journal of Advanced Manufacturing Technology*, 121.5 (2022): 2923-2946.
- [4] Dagne H, Lou Z, Zhang J, et al. "Roll angle error measurement of linear guideways using mini autocollimators." 2024 10th International Conference on Mechatronics and Robotics Engineering (ICMRE). IEEE, 2024.
- [5] Sutar, Manjushree D., and Bhagyesh B. Deshmukh. "Linear motion guideways—a recent technology for higher accuracy and precision motion of machine tool." *Int. J. Innovations Eng. Technol* 3.1 (2013).

- [6] Wang, Xiao Yi , C. G. Zhou , and Y. Ou . "Experimental analysis of the wear coefficient for the rolling linear guide." *Advances in Mechanical Engineering*, 11.1(2019).
- [7] Yang Z , Yinhu XI , Junhong M, et al. "Stripe theory based numerical method for solving asymmetrical hysteresis of friction force in linear rolling guideways." *Science China(Technological Sciences)*, 59(2016): 1320-1327.
- [8] Wang Wei; Zhang Yimin; Li Changyou, et al. "Effects of wear on dynamic characteristics and stability of linear guides." *Meccanica* 52.11 (2017): 2899-2913.
- [9] Bao, Hong, Mengjie Hao, Yuchen Du, et al. "Optimal Design of Multilevel Composite Lubrication Structures on Sliding Guide Rail Surfaces." *Coatings* 14.10 (2024): 1286.
- [10] Gackowiec, Paulina. "General overview of maintenance strategies—concepts and approaches." *Multidisciplinary Aspects of Production Engineering 2* (2019).
- [11] Erbiyik, Hikmet. "Definition of maintenance and maintenance types with due care on preventive maintenance." *Maintenance Management-Current Challenges, New Developments, and Future Directions*. IntechOpen, 2022.
- [12] Lee, Jay , and J. Singh. "Intelligent maintenance systems and predictive manufacturing." *Journal of manufacturing science and engineering* 142.11 (2020): 110805.
- [13] Zio, Enrico. "Data-driven prognostics and health management (PHM) for predictive maintenance of industrial components and systems." *Risk-informed methods and applications in nuclear and energy engineering*. Academic Press, 2024. 113-137.
- [14] Calabrese F , Regattieri A , Bortolini M, et al. "Predictive maintenance: A novel framework for a data-driven, semi-supervised, and partially online prognostic health management application in industries." *Applied Sciences* 11.8 (2021): 3380.
- [15] Cao Y , Li J , Sun Y, et al. "A review on rail remaining useful life prediction." *Transportation Safety and Environment* 7.4 (2025): tda062.
- [16] Staroszyk, Danny, Jens Müller, and Steffen Ihlenfeldt. "Increasing the service life prediction accuracy of linear guideways considering real operating conditions." *Mechanism and Machine Theory* 201 (2024): 105734.
- [17] Cheng Q , Zhang H Y , Zhang T, et al. "Prediction method of precision deterioration of rolling guide under multi-random parameters based on frictional thermal expansion effect." *Tribology International* 189 (2023): 108883.
- [18] Tao W , Zhong Y , Feng H, et al. "Model for wear prediction of roller linear guides." *Wear* 305.1-2 (2013): 260-266.
- [19] Zhang, Penghai, Tao Wang, and Jun Zha. "A study on accuracy of linear ball guide." *Proceedings of the Institution of Mechanical Engineers, part C: Journal of Mechanical Engineering Science* 236.7 (2022): 3293-3312.
- [20] Miao, Huihui, Chenyu Wang, Changyou Li, et al. "Vibration characteristics and reliability analysis of roller linear guideway workbench." *Nonlinear Dynamics* 111.23 (2023): 21461-21485.
- [21] Nie C , Zhou C , Wang T ,et al. Linear Rolling Guide Surface Wear-State Identification Based on Multi-Scale Fuzzy Entropy and Random Forest. *Lubricants*. 2025; 13 (8): 323.
- [22] Hui Y , Mei X , Jiang Z T T. "Assembly quality evaluation for linear axis of machine tool using data-driven modeling approach." *Journal of Intelligent Manufacturing* 33.3 (2022): 753-769.
- [23] Wu Y , Song J , Wu X, et al. "Fault diagnosis of linear guide rails based on SSTG combined with CA-DenseNet." *Journal of Dynamics, Monitoring and Diagnostics* (2024): 1-10.
- [24] Li Y , Liu B , Chang Z, et al. "Machine learning in tribology: A review on framework, case studies, and future perspectives." *Journal of Manufacturing Processes* 159 (2026): 243-270.
- [25] Sicard B , Butler Q , Ziada Y, et al. "Experimental setups for linear feed drive predictive maintenance: A review." 2023 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2023.
- [26] Saatci B T , Ulas M , Gurgenc T. "Hybrid AI systems for tool wear monitoring in manufacturing: A systematic review." *Applied Sciences* 16.1 (2025): 208.
- [27] Yang Y , Wang J , An W, et al. "Advances in prediction methods of wear of rails." *Wear* 564 (2025): 205676.
- [28] Gao T , Zhu H , Wu J, et al. "Hybrid physics data-driven model-based fusion framework for machining tool wear prediction." *The International Journal of Advanced Manufacturing Technology* 132.3 (2024): 1481-1496.
- [29] Gregory W. Vogl (2018), Linear Axis Testbed - Rail Degradation Experiment 01, National Institute of Standards and Technology, <https://doi.org/10.18434/T4/1502585> (Version: 1.1, Accessed 2026-04-12)
- [30] Kendall Alex, Yarin Gal, and Roberto Cipolla. "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [31] Zhang Zehua, Ning Shen, and Xingfen Wang. "Physics-Informed Transformer for Rolling Bearing RUL Prediction: A Dual-Stream Trend-Residual Decomposition Architecture." *IEEE Access* (2026).
- [32] Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Proceedings of the 33rd International Conference on Machine Learning*, 2016, 48: 1050–1059.
- [33] Xue Long, Kai Zhou, and Xiaoge Zhang. "Continuous optimization for construction of neural network-based prediction intervals." *Knowledge-Based Systems* 293 (2024): 111669.
- [34] Eressa Muluken Regas, Hakim Badis, and Dorian Grosso. "Distribution based upper lower bound estimation in deep neural nets." 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2022.
- [35] Chen L , Khalid O W , Tiang J J , et al. "Uncertainty aware and explainable construction cost prediction using a hybrid probabilistic learning model." *Scientific Reports* 16.1(2026).
- [36] Yau Her-Terng, Ping-Huan Kuo, and Song-Wei Hong. "Milling wear prediction using an artificial neural network model." *Engineering Applications of Artificial Intelligence* 135 (2024): 108686.
- [37] Si Sumei, Deqiang Mu, and Hailiang Tang. "A hybrid CBiGRUPE model for accurate grinding wheel wear prediction." *Sensors* 25.9 (2025): 2935.