Proceedings of Engineering and Technology Innovation, vol. 32, 2026, pp. 01-13

Improved CNN-LSTM Bearing Remaining Useful Life Prediction Based on the Weibull Loss Function

Yanping Zhang[1,2,*], Kho Lee Chin[1], Xiaozheng Li[3], Mingqiang Zhang[4], Dongfeng Yuan[5], Annie Joseph[1]

[1]Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

[2]Shandong Provincial Key Laboratory of Industrial Big Data and Intelligent Manufacturing, Qilu Institute of Technology, Shandong, China

[3]School of Information and Electrical Engineering, Shandong Jianzhu University, Shandong, China

[4]School of Cyber Science and Engineering, Qufu Normal University, Shandong, China

[5]School of QILU Transportation, Shandong University, Shandong, China

Received 30 June 2025; received in revised form 16 October 2025; accepted 17 October 2025

DOI: https://doi.org/10.46604/peti.2025.15348

Abstract

The prediction of the remaining useful life (RUL) of rolling bearings is a critical task in predictive maintenance. This paper presents a deep learning model named knowledge-driven convolutional neural network–long short-term memory (KCNN-LSTM), enhanced by the Weibull-based loss function tailored with historical bearing failure data. By incorporating a probabilistic Weibull modeling mechanism, the model can accurately capture the uncertainty and accelerated degradation trend of bearing failure over time. The prognostics and health management (PHM) 2012 and XJTU-SY bearing datasets are utilized to evaluate the proposed KCNN-LSTM model. The results indicate that the proposed KCNN-LSTM achieves superior performance compared with the conventional CNN-LSTM model, leading to a 10.2% improvement in the score metric and a notable reduction in prediction error. The proposed model offers a practical and effective approach for enhancing predictive maintenance in high-reliability industrial systems.

Keywords: bearing life prediction, Weibull, CNN, deep learning, predictive maintenance

1. Introduction

Rolling bearings are essential components in rotating machinery. Their operational condition directly affects the performance, reliability, and safety of equipment in industrial systems. As industrial systems become increasingly complex, the equipment failure rate rises significantly, with bearing failures being the most common source. According to Zhuang et al. [1] and Nandi et al. [2], approximately 40% to 50% of equipment failures are caused by malfunctioning or deteriorating rolling bearings. Once a failure occurs, it not only interrupts production but can also lead to unexpected downtime, expensive repairs, and safety risks. Therefore, the accurate prediction of the remaining useful life (RUL) of bearings has become important in the prognostics and health management (PHM) field, possessing both theoretical significance and practical engineering value [3-4]. In recent years, data-driven approaches, particularly those implementing deep learning techniques, have significantly progressed in bearing life prediction. This is primarily due to their capability for automatic feature extraction and effective modeling of time-series signals.

Consequently, many studies have explored optimizing network architectures to enhance prediction accuracy and robustness. For example, Zhang et al. [5] proposed the IF-SCINet, which performs well in identifying the degradation initiation point. Li and Jian [6] introduced signal energy ratio (SER) based metrics into long-short-term memory (LSTM) network

* Corresponding author. E-mail address: ypzhang@qlit.edu.cn

models to enhance the degradation modeling capability. Furthermore, You et al. [7] and Guo et al. [8] enhanced model robustness across variable operating conditions by integrating improved temporal convolutional networks (TCNs) with multi-head attention mechanisms. Among hybrid architectures, the convolutional neural network–long short-term memory (CNN-LSTM) model has been widely adopted for degradation trend modeling and RUL prediction. This is due to its ability to integrate the local feature extraction strengths of the convolutional neural network (CNN) with the temporal sequence learning capability of LSTM [9-11].

To further improve the generalization in practical industrial settings, researchers have explored several strategies to address challenges such as complex working environments and data scarcity. These include multi-sensor fusion [12], transfer learning [13], and small-sample learning [14-18]. In terms of feature extraction, techniques like wavelet transforms [19-21], multi-scale attention mechanisms [22-24], and variational mode decomposition (VMD) with sparse representation [25-26] have been employed to improve non-smooth degraded signals. In addition, there have been efforts to integrate physics knowledge into network architectures to enhance predictive model interpretability and stability. For example, MPINet integrates physical priors into its multi-scale feature fusion process and has demonstrated superior performance in sample-limited scenarios [27].

Despite these advances, a significant gap remains in integrating the physical distribution information of degradation lifetime into data-driven models. Specifically, most existing approaches do not explicitly model uncertainty in the late degradation stage, nor do they capture the accelerated wear behavior characteristic of bearing end-of-life dynamics. As a result, prediction deviations often occur, particularly when degradation accelerates nonlinearly near failure. The main contributions of this paper are as follows:

(1) Propose a knowledge-driven convolutional neural network–long short-term memory (KCNN-LSTM) bearing a life prediction model that integrates the Weibull loss function. This design enhances feature extraction and uncertainty modeling in the degradation stage while maintaining the strengths of CNN-LSTM time-series modeling.
(2) Construct a new health index (HI) tailored to the nonlinear characteristics of bearing degradation, thereby providing better insight into the degradation process and improving RUL prediction accuracy.
(3) Conduct a series of experiments across different operating conditions using publicly available datasets. The proposed model outperforms existing methods in prediction accuracy, curve smoothness, and cross-condition generalization, demonstrating its potential for real-world deployment in predictive maintenance.

2. Methodology Related Technologies

This section introduces the key theoretical foundations and algorithms supporting the proposed bearing RUL prediction framework. It begins with the definition of bearing life prediction and then details the core technologies employed, including CNN, LSTM, and Weibull reliability modeling theory. Furthermore, it presents the evaluation indicators used to assess prediction accuracy and reliability. These related technologies collectively form the methodological basis for constructing an intelligent and interpretable bearing life prediction model.

2.1. Bearing life prediction definition

Throughout the lifespan of bearing operation, it is mainly affected by load, speed, and temperature, which gradually lead to fatigue damage and, eventually, failure. Therefore, accurately predicting the RUL of bearings based on condition monitoring data has become a core issue in predictive maintenance and failure warning systems. The problem of bearing life prediction can be fundamentally understood as a time-series forecasting task. The goal is to utilize the multidimensional state data, including vibration, temperature, rotational speed, and others, collected throughout the degradation process to construct a mapping relationship from the raw sensor signals to the evolving health state and, subsequently, predict the bearing's RUL. Formally, this is described as follows:

Input: A time-ordered sequence of sensor measurements from the early stage of operation to the current time, often including time-domain and frequency-domain features of vibration signals and HI.
Output: The estimated remaining time that the bearing can operate before failure, commonly expressed as the RUL.

In order to achieve accurate RUL predictions, models can be constructed using either data-driven or hybrid approaches that incorporate physical insight with sensor data. Furthermore, quantifying the uncertainty in these predictions is important for making reliable maintenance decisions.

2.2. CNN

CNN is a feed-forward neural network designed for hierarchical feature extraction using convolutional operations. The typical architecture of a CNN, illustrated in Fig. 1, consists of five main components: input layer, convolutional layer, pooling layer, fully connected layers, and Softmax output layer.

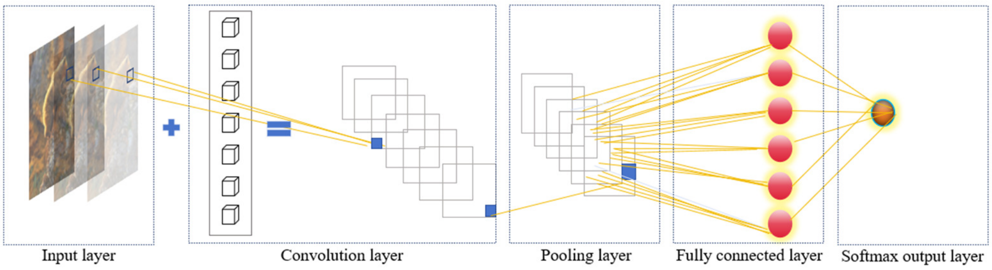


Fig. 1 CNN network architecture diagram

The convolutional layers utilize trainable filters (kernels) to extract local spatial features from the input data. Each filter is characterized by a set of weights and biases, which are optimized during training. Pooling layers follow the convolutional layers and serve to downsample the feature maps, effectively reducing spatial dimensionality while preserving essential information. This approach reduces the computational costs and improves generalization by limiting the risk of overfitting. In practice, CNNs are typically built with several convolutional and pooling layers arranged in alternating sequences, which enables the extraction of more complex and informative patterns from the input data.

2.3. LSTM

LSTM is an enhanced version of the recurrent neural network (RNN). It uses a gating mechanism to control the retention or discard of information over time, which improves the network's ability to handle sequential data. The cumulative structure in LSTM makes it more effective in calculating derivatives for back-propagation, which avoids the problem of gradient vanishing and allows the network to capture long-term dependencies between long sequences. As illustrated in Fig. 2, the LSTM network structure consists of four gates: input gate (i_t), output gate (o_t), forgetting gate (f_t), and storage unit (C_t).

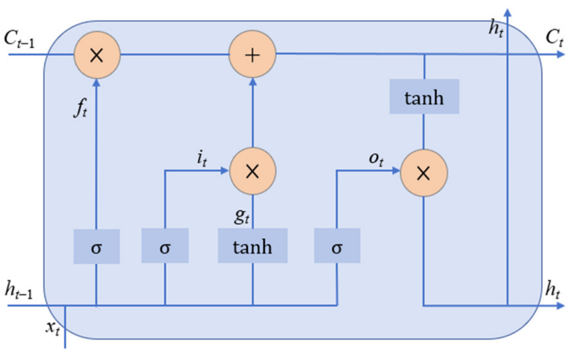


Fig. 2 Long-short-term memory unit

The formulas used to update the status of each gate and cell are defined below:

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{1}$$

$$f_t = \sigma(W_f + [h_{t-1}, x_t] + b_f) \tag{2}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tanh(W_c \times [h_{t-1}, x_t] + b_c) \tag{3}$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{4}$$

$$h_t = o_t \times \tanh(C_t) \tag{5}$$

where i_t determines the current information to be saved or updated; f_t decides which information to discard or retain, while also preventing gradient vanishing and explosion problems during iteration; o_t determines the portion of information to output from the memory cell; and the memory cell C_t contains the information stored at time t.

In LSTM, the cell state serves as long-term memory; the input and forget gates help the network automatically learn which information to retain or ignore, thereby making the network more suitable for processing long sequence data. The LSTM architecture is specifically designed to enable the network to adaptively process and retain critical information through the gate mechanism and optimize time series data processing.

2.4. Weibull reliability modeling theory

The Weibull distribution is a widely used probability distribution model in reliability engineering and life data modeling. Its flexibility in accommodating different failure modes makes it suitable for life modeling and RUL prediction of mechanical components such as bearings and cutting tools. The two-parameter Weibull Cumulative Distribution Function (CDF) represents the probability of failure at a given time or the proportion of equipment that experiences failure over time. In this study, the CDF is used to model the degradation behavior of bearings. The CDF function is defined as follows:

$$f(t) = 1 - e^{-(t/\eta)^{\beta}} \tag{6}$$

where β is the shape parameter, and η is the characteristic lifetime. When plotted on a log-log scale, β is the slope of the line. η represents the age at which 63.2% of the units fail. Furthermore, β corresponds to the common failure location of the component at the end of its lifetime.

Depending on β, the Weibull distribution reflects different failure mechanisms:
$\beta < 1$: early failure stage (failure rate decreases with time);
$\beta = 1$: conforms to an exponential distribution (failure rate is constant);
$\beta > 1$: wear failure stage (failure rate increases with time), which is suitable for modeling bearing and tool wear.

Ideally, β and η are estimated by maximum likelihood based on multiple failure samples t_i. For scenarios where only a small number of failures are recorded, the formula below can be used to estimate n, where r represents the number of observed failures without considering the right-censored samples. This procedure is part of the classical Weibull distribution modeling and uses the maximum likelihood method for parameter estimation.

$$\eta = \left[\sum_{i=1}^{N} \frac{t_i^{\beta}}{r} \right]^{1/\beta} \tag{7}$$

where t is the time or lifetime, r is the number of failed units, and N is the number of samples (failures + unfinished run-to-failure). Estimates for ball bearing roller bearings are 2.0 or 1.5.

For newer production facilities, estimates of β can be derived from production systems with similar layouts but longer operating histories. In this study, simple RUL experiments were conducted using two bearing datasets. However, both datasets had insufficient failure data, making it difficult to properly fit the Weibull distribution. Therefore, the Weibull equation was applied using commonly accepted bearing shape parameters, as shown in Table 1.

Table 1 Test sets for each operating condition and axis

Training set for different operating conditions	Test set for each condition	Same condition test set for each axis	
Axis 1-1	Axis 1-3	Axis 1-4	Axis 1-6
Axis 1-2		Axis 1-5	Axis 1-7
Axis 2-1	Axis 2-3	-	-
Axis 2-2		-	-
Axis 3-1	Axis 3-3	-	-
Axis 3-2		-	-

2.5. RUL forecast evaluation indicators

The RUL prediction of the rolling bearing problem is a deep learning-based regression task. Various evaluation metrics are typically employed to assess the effectiveness of the prediction using a deep learning model. This paper adopts the root-mean-square error (RMSE) and the scoring function (Score) provided by the PHM2012 dataset as evaluation metrics to assess the accuracy and reliability of the proposed deep learning model. RMSE is used to measure the difference between a model's predicted values and the actual values. It is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{8}$$

where $\hat{y}_i$ and y_i denote the predicted and actual values of healthiness, respectively, while n is the total number of test samples. Smaller values of RMSE indicate higher prediction accuracy.

2.6. Score

The score is an evaluation metric proposed in the NASA CMAPSS dataset. It considers the different impacts of predicting too early and too late on system maintenance, and is calculated as follows:

$$A_i = \begin{cases} e^{-\ln(0.5)\times(Er_i/5)}, & if\ Er_i \le 0 \\ e^{+\ln(0.5)\times(Er_i/20)}, & if\ Er_i > 0 \end{cases} \tag{9}$$

where Er_i is the percentage error in life prediction, which is calculated as

$$Er_i = \frac{act\text{RUL}_i - pred\text{RUL}_i}{act\text{RUL}_i} \tag{10}$$

Accuracy strongly depends on the confidence levels and distributions of the early and late predictions. In the current scoring function, a late prediction has a greater penalty compared to an early prediction, as it is more harmful. In actual situations, early prediction is also more meaningful.

Finally, the model score is given by the following formula, with higher scores indicating higher prediction accuracy:

$$\text{Score} = \frac{1}{n}\sum_{i=1}^{n} A_i \tag{11}$$

3. KCNN-LSTM Model Design

To enhance the prediction accuracy and interpretability of bearing RUL estimation, this section presents the design of a KCNN-LSTM model. The proposed model integrates data-driven deep learning with Weibull-based reliability theory to achieve physically interpretable predictions. It combines convolutional feature extraction, temporal sequence modeling, and Weibull-constrained learning to capture both local degradation patterns and long-term temporal dependencies in sensor data. The following subsections describe the model architecture, the Weibull loss function formulation, and the design of HI.

3.1. General structure of the model

To achieve accurate modeling and prediction of RUL for devices, this paper introduces a CNN-LSTM deep neural network model based on the Weibull loss function. The model's overall structure is illustrated in Fig. 3, which includes the input layer, the CNN module, the LSTM module, the Weibull parameter prediction layer, and the loss computation module.

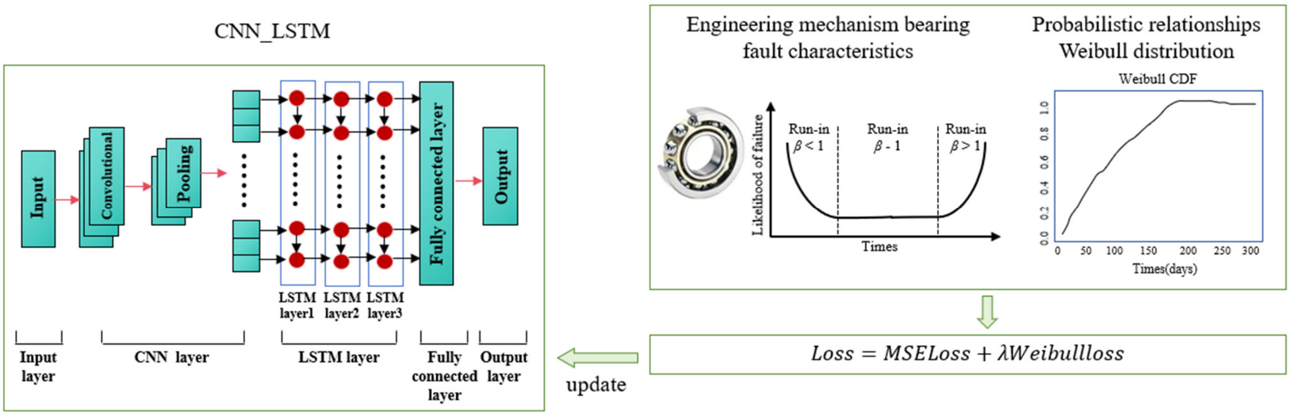


Fig. 3 Block diagram of the KCNN-LSTM model

(1) Input layer: input is multidimensional time series data, time window sequences collected by sensors such as vibration, temperature, current, etc.

(2) CNN module: used to extract local temporal features from the input signal. Multi-scale feature extraction is performed on the time series using 32-dimensional convolution kernels to enhance the model's ability to perceive local change trends in the signal. The size of the convolution kernels remains fixed during training, and no dynamic adjustment is implemented. The model adopts a static multi-scale convolution design to capture diverse local pattern features under different receptive fields.

(3) LSTM module: employs a 2-dimensional input vector and constructs three hidden layers, each containing 64 neurons. Dropout regularization is applied to all layers during the training phase to prevent overfitting. The Rectified Linear Unit (ReLU) activation function is used for all layers except the last one, where the Sigmoid function is applied to produce the final output.

(4) Loss function module: adopts a constraint-based loss function derived from Weibull's CDF to match the predicted values with the actual failure probability, which guides backpropagation and model optimization. In this framework, a knowledge-guided feed-forward neural network based on the Weibull loss function embeds reliability engineering knowledge into the model training process, thereby improving the rationality and accuracy of bearing RUL prediction.

3.2. Weibull loss function

The Weibull CDF $f(t)$ is the summation of an approximate constrained loss function and the Weibull-based loss function. The detailed formulation is given by

$$L(t_i,\hat{t}_i)=\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(t_i-\hat{t}_i\right)^2}_{MSELoss}+\underbrace{\lambda\frac{1}{n}\sum_{i=1}^{n}\left(F(t_i)-F(\hat{t}_i)\right)^2}_{Weibullloss} \tag{12}$$

where n is the number of samples in the dataset. $F(t_i)$ is the true failure rate at time i, which is calculated by the CDF shown in Eq. (6). $F(\hat{t}_i)$ is the estimated failure rate at time i. The failure rate is calculated by the CDF. The estimated failure rate is calculated from the output of the neural network after each training step.

The above Weibull distribution-based loss function can be used for both sufficient and insufficient failure data. When insufficient failure data is present, an estimate β must be used in conjunction with Weibull's equation to generate an estimate η. The parameter η is dynamically computed according to Eq. (7), meaning that it is not a fixed constant but an adaptive value estimated by the network based on the input features. When the predicted η deviates from the actual degradation pattern, the loss function increases, and the backpropagation process transmits the error signal to the feature extraction layers (CNN-LSTM module), forcing the network to learn more physically meaningful feature representations so that η becomes more consistent with the true degradation behavior.

3.3. Design of HI

The degradation process of bearings is usually nonlinear, especially towards the end of their life cycle when the degradation rate accelerates. Exponential degradation functions are particularly well-suited for modeling this dynamic, as they can capture this nonlinear variation and provide a more accurate bearing health state assessment.

This paper focuses on the accelerated degradation of data generation, especially near the onset of failure modes; thus, a health metric based on cumulative degradation is adopted. This metric is modeled using an exponential decay function, which is expressed as follows:

$$\mathrm{HI}(t)=d-e^{t\tau+a} \tag{13}$$

$$\begin{cases}\mathrm{HI}(t_{\min})=1\\ \mathrm{HI}(t_{\max})=0\end{cases} \tag{14}$$

where a is the convergence rate hyperparameter, while d and τ are general parameters determined by solving Eq. (13) from Eq. (14). If HI = 1, the experimental system is functioning normally, and if HI = 0, the experimental system is in complete failure mode.

4. Experimental Design and Result Analysis

To validate the effectiveness and generalization performance of the proposed KCNN-LSTM model, a series of experiments was conducted using publicly available bearing degradation datasets. This section introduces the experimental datasets, model configurations, and comparison baselines, followed by a detailed analysis of parameter sensitivity, performance evaluation, and visual result interpretation. The experimental framework aims to comprehensively assess the model's capability in capturing nonlinear degradation trends, enhancing prediction accuracy, and maintaining robustness across different working conditions.

4.1. Experimental dataset

(1) Data presentation

The proposed model in this paper is validated using the PHM2012 dataset provided by the IEEE society [28]. This dataset contains 17 sets of full life cycle vibration signals of bearings collected using accelerometers, including 6 training sets and 6 test sets, as shown in Table 1.

(2) Data preprocessing

This experiment adopts the FEMTO-ST bearing dataset, provided for the IEEE PHM2012 Data Challenge (i.e., the PHM2012 bearing dataset) as the experimental verification data. In this dataset, horizontal vibration signals are utilized because bearing faults typically manifest more distinct fault-related features in the horizontal direction. Before normalization, the time-domain standard deviation feature is extracted from the raw vibration data. To represent the bearing's operating condition every 10 seconds, the system calculates features from each 0.1-second segment of the signal.

Subsequently, the time-domain vibration signal is converted into the frequency domain through a Fast Fourier Transform (FFT) to obtain spectral features. Each segment is windowed by a Hann window and transformed using a 2,048-point FFT. The mean and standard deviation are computed in both the time and frequency domains. In this study, the RUL labels are derived from the corresponding HI of each sample, with a normalized relationship defined as RUL(t) = HI(t), where a lower HI value indicates a shorter remaining lifetime.

(3) Parameters

In the experimental configuration, the proposed model was trained with a learning rate of 0.001. The network architecture comprised an input layer with two neurons, a CNN layer with 32 filters, an LSTM layer with 64 hidden units, and a single output neuron. The model was trained using a mini-batch size of 32 for 60 epochs to ensure training stability and convergence. Furthermore, this study analyzes the effects of different parameter (β, λ, η) configurations on the model's predictive performance, providing insights for subsequent model optimization. The selection of their values is summarized in Table 2.

Table 2 Parameters

Parameters	Value	Parameters	Value
Learning rate	0.001	β	2
Input layer	2	λ	2.28
CNN network layer	32	epoch	60
LSTM network layer	64	Batch size	32
Output layer	1	-	-

4.2. *Experimental scheme design and baselines*

To systematically evaluate the performance of different models in tool wear prediction, five groups of comparative experiments are designed in this paper:

(1) CNN-LSTM (MSE): a combination–LSTM hybrid model in which mean squared error (MSE) is used as the loss function.

(2) CNN-LSTM (RMSLE): a variant that adopts root-mean-squared logarithmic error (RMSLE) as the loss function to reduce sensitivity to small-value errors.

(3) LSTM and Bi-LSTM: one-way LSTM and two-way Bidirectional Long Short-Term Memory (Bi-LSTM) structures are adopted, respectively, with MSE employed as the loss function.

(4) FNN (MSE + Weibull): a two-layer feedforward neural network (FNN) taken from von Hahn and Mechefske [29], which utilizes the same composite loss (MSE + Weibull constraint) and serves as a knowledge-informed shallow baseline for comparison with the KCNN-LSTM, thereby highlighting the improvement of deep temporal modeling under the same Weibull-constrained framework.

(5) KCNN-LSTM (proposed): the kurtosis convolution kernel is introduced into the CNN-LSTM structure, with a Weibull-weighted term added to the loss function to enhance the modeling of critical wear transitions.

The above models enable a comparative evaluation of deep learning and machine learning methods across different wear stages, from the perspectives of structural design and loss function.

4.3. Results analysis

(1) Parameter sensitivity analysis

To evaluate the influence of Weibull parameters on model performance, a sensitivity analysis was conducted on the shape parameter β and the weighting coefficient λ in the loss function Eq. (12). The parameter β determines the form of the degradation trend, where $\beta < 1$ indicates early-failure behavior, $\beta = 1$ corresponds to constant failure rate, and $\beta > 1$represents accelerated degradation, which aligns with the wear-out process of bearing deterioration. The coefficient λ balances the contribution of the data-driven MSE term and the physically constrained Weibull term-small λ weakens the physical constraint, while excessively large λ may cause numerical instability.

Table 3 summarizes the results obtained by varying β from 1 to 3 while keeping λ fixed. The model exhibits the best performance when $\beta = 2.0$, achieving the lowest RMSE (0.071) and the highest Score (0.821). When $\beta < 1.6$ or $\beta > 2.5$, the performance decreases, and for $\beta = 3$, the Score drops sharply, indicating that an excessively steep degradation assumption reduces generalization. This confirms that $\beta \approx 2$ provides a reasonable balance between physical interpretability and numerical stability. Table 4 presents the results for different λ values with β fixed at 2.0. The model performance improves as λ increases from 0.1 to 2.0, where the RMSE decreases from 0.082 to 0.078, and the Score rises from 0.808 to 0.818. When $\lambda \approx 2.0$ to 2.3, the model achieves its best performance (RMSE = 0.071, Score = 0.821). Further increasing λ to 2.5 or 3.0 causes a slight deterioration, reflecting over-emphasis on the Weibull term and minor oscillations during training.

Overall, the sensitivity analysis demonstrates that $\beta = 2.0$ and $\lambda = 2.28$ yield the optimal trade-off between accuracy and stability. These values are consistent with the physical degradation characteristics of bearings and are adopted as the default parameters in subsequent experiments.

Table 3 β value selection

β	1	1.6	2.0	2.5	3
RMSE	0.079	0.078	0.071	0.0804	0.0878
Score	0.807	0.812	0.821	0.803	0.518

Table 4 λ value selection

λ	0.1	0.5	1.0	2.0	2.28	2.5	3
RMSE	0.082	0.086	0.081	0.078	0.071	0.0776	0.0784
Score	0.808	0.085	0.794	0.818	0.821	0.813	0.804

(2) Performance evaluation

To validate the generalization capability of the proposed model, experimental evaluations were conducted on the PHM2012 bearing dataset and the XJTU-SY bearing dataset (http://biaowang.tech/xjtu-sy-bearing-datasets). To quantitatively assess the prediction performance of different models, six comparative experiments were designed, with RMSE and Score selected as the main evaluation metrics. The experimental results of each model on both datasets are summarized in Table 5, from which the overall performance in terms of prediction accuracy and stability can be intuitively compared.

According to Table 5, the results of each model are analyzed as follows:

<u>Performance of CNN-LSTM under different loss functions</u>: The CNN-LSTM model trained with RMSLE loss outperforms the one with MSE loss, achieving a higher Score (maximum 0.756) and a lower RMSE (minimum 0.083), indicating that RMSE loss provides better sensitivity to degradation trends. The proposed KCNN-LSTM model, integrating a Weibull-based loss function, achieves either the lowest or second-lowest RMSE and the highest Score across all test channels. For instance, in the bearing 1-3 channel of the PHM2012 dataset, the RMSE is 0.071, and the Score reaches 0.821, demonstrating its effectiveness in multi-scale structural feature extraction and progressive wear modeling.

Performance comparison of different models under the same loss function: The Bi-LSTM model generally surpasses the conventional LSTM owing to its bidirectional temporal modeling capability, yielding improved RMSE and Score results. Conversely, the standard LSTM shows limited capacity in capturing complex degradation dynamics. Although the Weibull loss + machine learning approach introduces prior knowledge of degradation, it lacks deep structural representation, resulting in relatively high RMSE values (minimum 0.133) and weaker predictive robustness compared with the KCNN-LSTM.

Generalization validation on the XJTU-SY dataset: To further evaluate cross-dataset generalization, the models were tested on the XJTU-SY bearing dataset. The proposed KCNN-LSTM model maintains superior performance, achieving the lowest RMSE of 0.054 and the highest Score of 0.779 among all compared methods. This consistent advantage across two distinct datasets confirms the model's strong generalization ability and robust adaptation to varying degradation patterns and operating conditions.

Overall, incorporating the Weibull-based loss function within a structurally enhanced deep model significantly improves the accuracy, stability, and robustness of RUL prediction. These improvements are particularly evident in multi-stage wear processes, where the ability to capture complex temporal-spatial degradation dynamics is essential for accurate remaining life estimation.

Computational efficiency: The proposed KCNN-LSTM model was implemented in PyTorch and trained on an NVIDIA RTX 4090 GPU. The total training process for one complete epoch with approximately 1,200 samples required about 10 seconds. This indicates that the model achieves both satisfactory prediction accuracy and acceptable computational efficiency for practical industrial deployment, supporting its potential use in real-time bearing health monitoring.

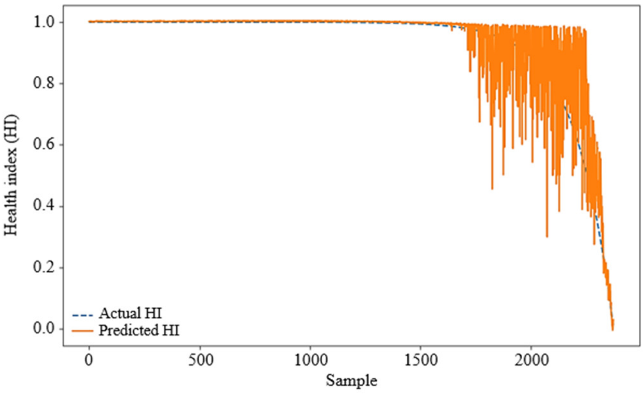
Table 5 RMSE and Score test results of each model

Model	Index	PHM2012					XJTU-SY (bearing 1-3)
		bearing 1-3	bearing 1-4	bearing 1-7	bearing 2-3	bearing 3-3	
CNN-LSTM (MSE)	Score	0.719	0.678	0.684	0.683	0.724	0.734
	RMSE	0.083	0.287	0.126	0.171	0.107	0.083
CNN-LSTM (RMSLE)	Score	0.691	0.749	0.756	0.702	0.701	0.721
	RMSE	0.083	0.289	0.118	0.100	0.115	0.082
LSTM	Score	0.644	0.366	0.578	0.554	0.572	0.643
	RMSE	0.071	0.330	0.087	0.181	0.201	0.182
Bi-LSTM	Score	0.753	0.662	0.738	0.612	0.673	0.764
	RMSE	0.064	0.315	0.079	0.157	0.126	0.074
FNN (MSE + Weibull)	RMSE	0.133	-	-	0.168	0.200	-
KCNN-LSTM (proposed)	Score	0.821	0.745	0.768	0.725	0.771	0.779
	RMSE	0.071	0.264	0.123	0.094	0.086	0.054

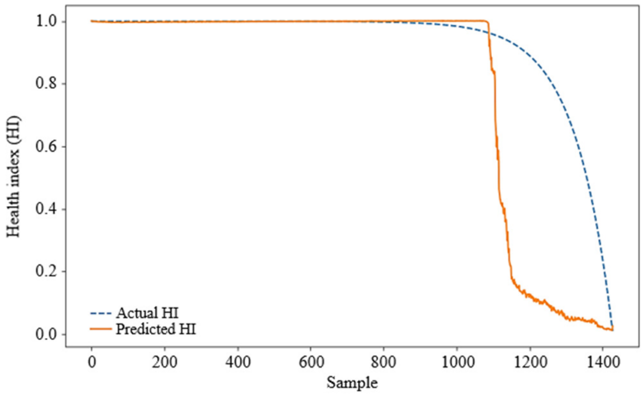
(3) Visual analysis of forecast results

To further verify the adaptability and robustness of the proposed KCNN-LSTM model, a visual analysis on six groups of bearing data (bearing 1-3, bearing 1-4, bearing 1-7, bearing 2-3, bearing 3-3, and XJTU-SY bearing 1-3) was conducted, as shown in Fig. 4. In these figures, the blue curve is the life degradation trend predicted by the model, and the orange curve is the normalized actual life. The results show that the model can accurately capture the overall degradation trend and identify the inflexion point of accelerated degradation in the later stages.

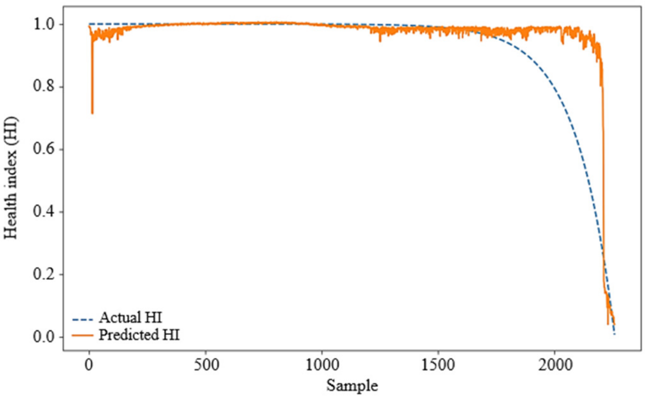
The model can also maintain stable prediction performance under different operating conditions. The prediction curve is smooth and resilient to noise, reflecting strong generalization ability and potential for practical deployment. The proposed model provides reliable technical support for bearing life prediction under complex real-world conditions.



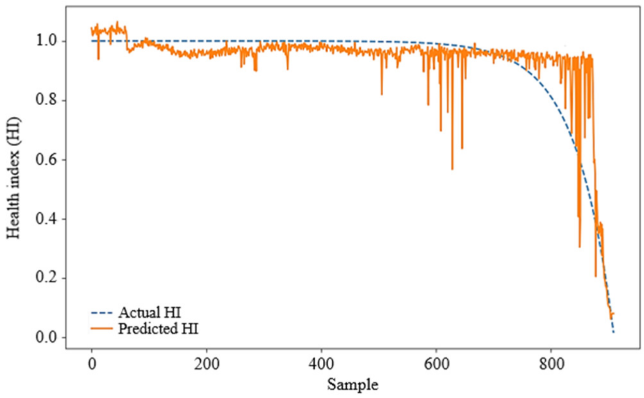
(a) Bearing 1-3



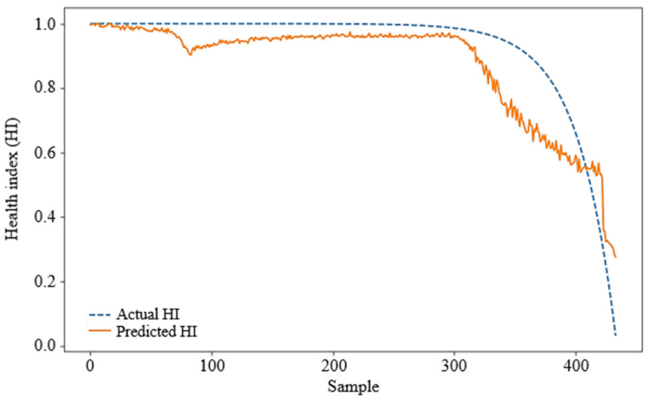
(b) Bearing 1-4



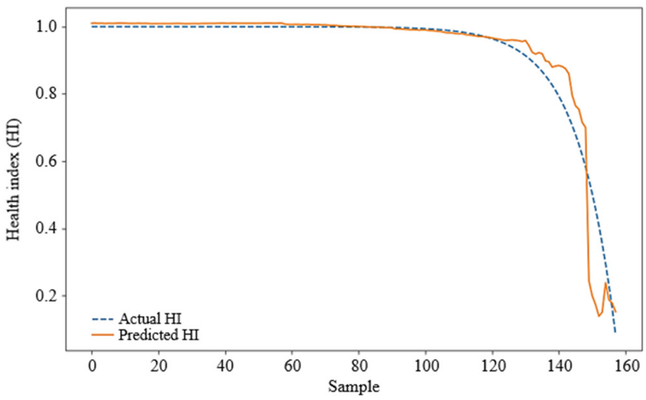
(c) Bearing 1-7



(d) Bearing 2-3



(e) Bearing 3-3



(f) XJTU-SY(bearing1-3)

Fig. 4 KCNN-LSTM prediction results for six bearing datasets

5. Conclusions

This study proposes a KCNN-LSTM prediction model that integrates an enhanced convolutional structure with a Weibull-based loss function to address the challenges of bearing degradation prediction. A series of comparative experiments under different deep learning models and loss function configurations was comprehensively evaluated. The model's performance, robustness, and generalization capability were verified using multi-channel bearing datasets (PHM2012 and XJTU-SY). The key findings are summarized as follows:

(1) Leading prediction accuracy: Among the models evaluated, the KCNN-LSTM model achieves the best RMSE and Score values across multiple channels. For instance, it attains an RMSE as low as 0.071 and a Score up to 0.821 on the PHM2012 dataset, while maintaining the lowest RMSE (0.054) and highest Score (0.779) on the XJTU-SY dataset. These results confirm its superior prediction accuracy and robustness compared with CNN-LSTM, Bi-LSTM, and traditional LSTM models.

(2) Weibull loss enhances degradation modeling ability: Compared with the conventional MSE-based losses, incorporating the Weibull-based composite loss better conforms to the physical characteristics of the degradation process. Consequently, this enhances the model's sensitivity to the accelerated wear phase and improves prediction accuracy in the end-of-life region.

(3) Visualization of degradation prediction performance: In the visual analysis across two bearing datasets under six different operating conditions, the KCNN-LSTM model produces smooth and stable prediction curves that closely match the true degradation trajectories. The model accurately identifies degradation inflection points and demonstrates strong noise suppression ability, especially during rapid wear stages.

(4) Strong generalization ability: The proposed model achieves consistently excellent prediction results across multiple datasets, channels, and working conditions, demonstrating high adaptability and universality. This confirms its potential for engineering applications in real-world bearing health monitoring and RUL prediction.

The KCNN-LSTM model proposed in this study is an accurate, robust, and generalizable approach. It provides a reliable solution for equipment health management and residual life prediction in actual industrial scenes. For future work, research on multi-modal signals, dynamic degradation process modeling mechanisms, and uncertainty quantification strategies can be further considered and explored to enhance the practicability and interpretability of the model.

Acknowledgments

This work was supported in part by the Key Research and Development (Major Scientific and Technological Innovation) Project of Shandong Province under Grant No. 2019TSLH0202 and No. 2019JZZY01011, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant No. 2021B1515120066.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] J. Zhuang, M. Jia, Y. Cao, and X. Zhao, "Semi-Supervised Double Attention Guided Assessment Approach for Remaining Useful Life of Rotating Machinery," Reliability Engineering & System Safety, vol. 226, article no. 108685, 2022.

[2] S. Nandi, H. A. Toliyat, and X. Li, "Condition Monitoring and Fault Diagnosis of Electrical Motors—A Review," IEEE Transactions on Energy Conversion, vol. 20, no. 4, pp. 719-729, 2005.

[3] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and A. Abraham, "Multi-Fault Diagnosis of Industrial Rotating Machines Using Data-Driven Approach: A Review of Two Decades of Research," Engineering Applications of Artificial Intelligence, vol. 123, part A, article no. 106139, 2023.

[4] N. Costa and L. Sánchez, "Variational Encoding Approach for Interpretable Assessment of Remaining Useful Life Estimation," Reliability Engineering & System Safety, vol. 222, article no. 108353, 2022.

[5] J. Zhang, C. Zhang, S. Xu, G. Liu, H. Fei, and L. Wu, "Remaining Life Prediction of Bearings Based on Improved IF-SCINe," IEEE Access, vol. 12, pp. 19598-19611, 2024.

[6] L. Li and Q. Jian, "Remaining Useful Life Prediction of Wind Turbine Main-Bearing Based on LSTM Optimized Network," IEEE Sensors Journal, vol. 24, no. 13, pp. 21143-21156, 2024.

[7] K. You, P. Wang, and Y. Gu, "Toward Efficient and Interpretative Rolling Bearing Fault Diagnosis via Quadratic Neural Network with Bi-LSTM," IEEE Internet of Things Journal, vol. 11, no. 13, pp. 23002-23019, 2024.

[8] Y. Guo, J. Zhou, Z. Dong, H. She, and W. Xu, "Rolling Bearing RUL Prediction Based on Fusion of Multi-Head Attention and Improved TCN-BiLSTM," IEEE Access, vol. 12, pp. 95641-95658, 2024.

[9] D. K. B. Kulevome, H. Wang, and X. Wang, "Rolling Bearing Fault Diagnostics Based on Improved Data Augmentation and ConvNet," Journal of Systems Engineering and Electronics, vol. 34, no. 4, pp. 1074-1084, 2023.

[10] Z. Chen, B. Xu, and Z. Zhang, "Rolling Bearing Fault Diagnosis Based on Recurrence Plot," IEEE Access, vol. 12, pp. 149710-149721, 2024.
[11] Y. Chang and G. Bao, "Enhancing Rolling Bearing Fault Diagnosis in Motors Using the OCSSA-VMD-CNN-BiLSTM Model: A Novel Approach for Fast and Accurate Identification," IEEE Access, vol. 12, pp. 78463-78479, 2024.
[12] J. Tong, C. Liu, J. Bao, H. Pan, and J. Zheng, "A Novel Ensemble Learning-Based Multisensor Information Fusion Method for Rolling Bearing Fault Diagnosis," IEEE Transactions on Instrumentation and Measurement, vol. 72, article no. 9501712, 2022.
[13] J. Mi, M. Chu, Y. Hou, J. Jin, W. Huang, T. Xiang, et al., "A Fault Diagnosis Method for Rolling Bearing Based on Deep Adversarial Transfer Learning with Transferability Measurement," IEEE Sensors Journal, vol. 24, no. 1, pp. 984-994, 2024.
[14] J. Zheng, S. Cao, K. Feng, and Q. Liu, "Zero-Phase Filter-Based Adaptive Fourier Decomposition and Its Application to Fault Diagnosis of Rolling Bearing," IEEE Transactions on Instrumentation and Measurement, vol. 73, article no. 3512111, 2023.
[15] M. Sun, X. Xiao, T. Chen, Q. He, Z. Long, and L. Li, "A Novel Domain Incremental Learning Method for Bearing Fault Diagnosis Based on F&K," IEEE Transactions on Industrial Informatics, vol. 21, no. 1, pp. 980-989, 2025.
[16] H. Liu, R. Yuan, Y. Lv, X. Yang, H. Li, and G. Song, "Degradation Tracking of Rolling Bearings Based on Local Polynomial Phase Space Warping," IEEE Transactions on Reliability, vol. 73, no. 2, pp. 1380-1392, 2024.
[17] C. Peng, Y. Sheng, W. Gui, Z. Tang, and C. Li, "A Rolling Bearing Fault Diagnosis Method Based on Multimodal Knowledge Graph," IEEE Transactions on Industrial Informatics, vol. 20, no. 11, pp. 13047-13057, 2024.
[18] C. Wang, X. Wang, and C. Zeng, "Saw Blade Wear Identification Based on Data Enhancement and Feature Fusion," IEEE Access, vol. 11, pp. 123677-123687, 2023.
[19] C. Li, Y. Tang, N. Lei, and X. Wang, "An Intelligent Fault Diagnosis Method Based on Optimized Parallel Convolutional Neural Network," IEEE Sensors Journal, vol. 25, no. 4, pp. 6160-6175, 2025.
[20] Y. Chen and L. Xiao, "A Multisource–Multitarget Domain Adaptation Method for Rolling Bearing Fault Diagnosis," IEEE Sensors Journal, vol. 24, no. 3, pp. 3406-3419, 2024.
[21] X. Zhao, L. Wang, M. Yang, Y. Chen, and J. Xiang, "A Novel Small-Sample Fault Diagnosis Method for Rolling Bearings via Continuous Wavelet Transform and Siamese Neural Network," IEEE Sensors Journal, vol. 24, no. 15, pp. 24988-24996, 2024.
[22] Z. Cao, J. Dai, W. Xu, and C. Chang, "Sparse Bayesian Learning Approach for Compound Bearing Fault Diagnosis," IEEE Transactions on Industrial Informatics, vol. 20, no. 2, pp. 1562-1574, 2024.
[23] Z. Meng, T. Shao, J. Liu, J. Li, L. Cao, and F. Fan, "Scale-Demodulation Adaptive Chirp Mode Decomposition with Application to Rolling Bearing Fault Diagnosis," IEEE Sensors Journal, vol. 24, no. 20, pp. 32554-32565, 2024.
[24] J. Feng, Y. Xing, Y. Yao, and B. Wang, "Artificial Feature Bias Rectified by Self-Supervised Learning for Rolling Bearings Fault Diagnosis Under Limited Labeled Vibration Signals," IEEE Transactions on Instrumentation and Measurement, vol. 74, article no. 3502812, 2024.
[25] B. Chang, X. Zhao, D. Guo, S. Zhao, and J. Fei, "Rolling Bearing Fault Diagnosis Based on Optimized VMD and SSAE," IEEE Access, vol. 12, pp. 130746-130762, 2024.
[26] A. Yan, Y. Zhao, Z. Lu, Y. Pang, S. Jin, Z. Liu, et al., "An Online Fault Detection and Remaining Life Prediction Method Based on SVDD for Rolling Bearings," IEEE Transactions on Instrumentation and Measurement, vol. 73, article no. 3516112, 2024.
[27] C. Gao, Z. Wang, Y. Guo, H. Wang, and H. Yi, "MPINe: Multiscale Physics-Informed Network for Bearing Fault Diagnosis with Small Samples," IEEE Transactions on Industrial Informatics, vol. 20, no. 12, pp. 14371-14380, 2024.
[28] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Morello, N. Zerhouni, et al., "PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests," IEEE International Conference on Prognostics and Health Management, pp. 1-8, 2012.
[29] T. von Hahn and C. K. Mechefske, "Knowledge Informed Machine Learning Using a Weibull-Based Loss Function," https://doi.org/10.48550/arXiv.2201.01769, 2022.