2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Bearing Remaining Useful Life Prediction via Multi-Scale Convolution and Bidirectional Gated Recurrent Unit Network

Jian Li[1,2], Rangyong Zhang[1,2(✉)], Hu Liang[1,2], and Yiming Zhang [3]

[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250013, China
[2] Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan 250013, China
[3] Wenzhou-Kean University, Wenzhou 325060, China
zhangry@sdas.org

Abstract. Accurate remaining useful life (RUL) prediction of rolling bearings plays a vital role in industrial predictive maintenance. Nevertheless, current approaches fail to effectively extract multi-scale degradation features in noisy environments, resulting in significant prediction inaccuracies. We propose a Multi-Scale Convolutional Bidirectional Gated Recurrent Unit (MSCNN-BiGRU) network for bearing remaining useful life prediction. First, raw vibration signals undergo deep feature extraction via a Stacked Denoising Autoencoder (SDAE), followed by dimensionality reduction using a Hierarchical Self-Organizing Map (HSOM) to generate a 1D degradation curve (DC). A Multi-Scale Convolution module is then constructed, incorporating 1D dilated convolution and a multi-scale strategy to extract degradation features from the DC, enabling the simultaneous capture of localized defects and global trend patterns. Finally, an attention layer is integrated at the feature input stage, combined with a GRU to construct a Bidirectional GRU (BiGRU) prediction model, which dynamically weights critical temporal dependencies for accurate RUL estimation. Experiments on the PHM2012 dataset that MAE is reduced by an average of 18.7% compared to sub-optimal models, and this work provides a generalizable framework for RUL prediction of rotating machinery, enhancing the reliability of industrial maintenance systems.

Keywords: Remaining Useful Life, Rolling bearings, Feature Extraction, Degradation Curve, Bidirectional Gated Recurrent Unit.

1 Introduction

In modern industrial production, the stability and reliability of machinery are critical to ensuring both production efficiency and operational safety. As a key rotating component in machinery, bearings play a pivotal role in determining the overall system's performance and operational effectiveness [1]. However, under harsh working conditions, bearings are prone to degradation caused by wear, fatigue, and aging, which can ultimately lead to failure. Consequently, implementing Prognostics and Health

Management (PHM) [2,3] strategies for bearings is essential to ensure the health and stability of equipment operations. Within the domain of PHM, accurately predicting the RUL of bearings has emerged as a critical yet challenging problem [4,5]. Precise RUL predictions not only enhance operational efficiency but also prevent unplanned failures and ensure the safety and continuity of industrial production processes [6].

The rapid advancements in deep learning technologies in recent years have highlighted their exceptional scalability, powerful representation learning capabilities, and ability to extract deep features from raw data. These advancements have driven the widespread application of deep learning techniques in bearing RUL prediction. Unlike traditional methods, deep learning approaches enable direct prediction of bearing RUL from raw vibration signals without requiring manual feature engineering. This eliminates issues such as subjectivity and inefficiency, as the models can autonomously learn and extract degradation features. Methods leveraging Convolutional Neural Networks (CNN) [7], Recurrent Neural Networks (RNN) [8] and Deep Belief Networks (DBN) [9] have been widely adopted for RUL prediction. For example, Tang et al. [10] developed a bearing RUL prediction framework based on Long Short-Term Memory (LSTM) networks and Transformer models. Their approach constructs a degradation feature set using time-frequency domain features and trains the LSTM-Transformer model to improve prediction accuracy. Similarly, Mou et al. [11] proposed a method combining Convolutional Deep Belief Networks (CDBN) and Bidirectional Long Short-Term Memory (BiLSTM) to handle large-scale, nonlinear, and high-dimensional degradation systems. This method uses CDBN to generate deep health indicators and BiLSTM to analyze time-series data and degradation trends, with RUL estimated through Monte Carlo simulation. Wang et al. [12] introduced a novel approach that converts raw vibration signals into time-frequency representations and employs a 3D deep CNN to extract degradation features comprehensively. Yang et al. [13] proposed an LSTM-based bearing RUL prediction model that incorporates a Dropout module to enhance training stability and prediction accuracy. Furthermore, Cao et al. [14] utilized raw vibration signals processed through edge spectral analysis, feeding them into a time-convolutional neural network to extract deep degradation features. Pan et al. [15] designed a performance degradation assessment method for gearbox bearings based on Deep Belief Networks and Self-Organizing Maps, where denoised vibration signals were analyzed and an improved particle filter was applied to predict gearbox RUL.

Despite the demonstrated efficacy of deep learning in bearing RUL prediction through autonomous fault pattern analysis, critical challenges persist, including unbalanced feature importance allocation undermining prediction reliability, limitations in spatiotemporal feature extraction from vibration signals, and suboptimal model parameter configurations compromising robustness. To address these issues, this paper proposes a novel approach for bearing RUL prediction based on the bearing degradation curve and an improved MSCNN-BiGRU neural network. The main contributions are as follows:

(1) This method innovatively integrates the deep feature extraction capability of SDAE with the topology-preserving properties of HSOM. By implementing a hierarchical feature space compression mechanism via HSOM, it maintains the continuity of fault evolution trajectories during dimensionality reduction. The proposed approach

achieves nonlinear mapping between high-dimensional vibration signal features and low-dimensional health indicators in degradation curve construction, thereby providing a novel pathway to address accuracy deviations caused by distortion in degradation representation within high-dimensional vibration signals.

(2) This paper proposes a bearing RUL prediction model based on MSCNN-BiGRU. The framework integrates multi-scale convolution and dilated convolution to enhance feature learning, utilizes bidirectional gated recurrent unit networks to capture temporal dependencies while improving computational speed and reducing model complexity, and incorporates attention mechanisms to prioritize critical features, thereby boosting prediction accuracy and computational efficiency.

2 Methods

To address the prevalent issues in existing bearing degradation state modeling methods, such as excessive reliance on manual intervention and suboptimal feature extraction performance under noisy environmental signals, we introduce an SDAE-HSOM degradation curve construction method upstream of the BiGRU model. This approach employs SDAE to perform deep feature extraction on raw vibration signals, followed by HSOM to achieve layer-wise ordered dimensionality reduction of the SDAE-processed features, ultimately yielding a one-dimensional bearing degradation curve. Furthermore, we enhance traditional convolutional neural networks by integrating a 1D dilated convolution technique and a multi-scale concept, thereby improving the model's ability to capture multi-scale features and global information from sequential data. The complete bearing RUL prediction framework is illustrated in Fig. 1, where red dashed boxes highlight the proposed improvements in this study.

2.1 Construction of Bearing DC Based on SDAE-HSOM

The key to predicting the RUL of bearings lies in establishing a robust state degradation model. Vibration signals collected from operational equipment contain critical information reflecting bearing health conditions. By analyzing these signals, we can extract essential features that characterize the transition from healthy to faulty states. To accurately capture degradation trends, we propose a method for constructing bearing degradation curves based on a SDAE and HSOM. The SDAE [16], composed of multiple denoising autoencoders (DAEs) [16] trained layer-wise, enhances noise resistance, robustness, and generalization through noise injection mechanisms such as random masking or Gaussian noise.An overview of the three-layer DAE network structure designed in this study is shown in Fig. 2.

The DAE enhances feature learning through a noise injection mechanism. The input signal x is corrupted by random masking or Gaussian noise to generate the damaged signal $\hat{x}$, and the network reconstructs the original signal x'.The SDAE adopts a hierarchical feature extraction strategy, leveraging a three-layer stacked architecture to

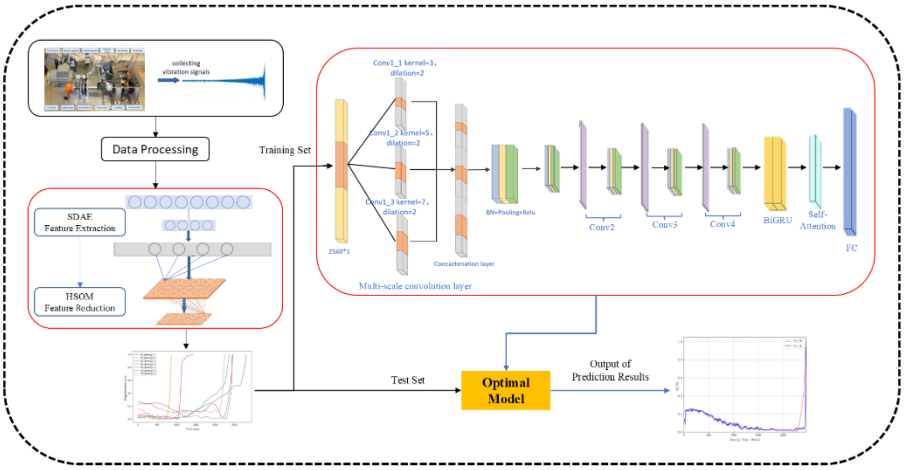

Fig. 1. The overall framework of MSCNN-BiGRU.

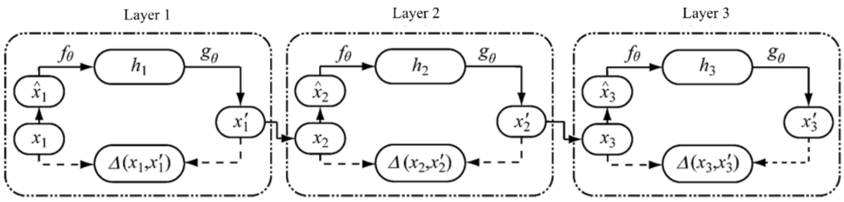

Fig.2. Three-Layer SDAE Structure. The SDAE comprises an input layer x_i, a hidden layer h_i, and an output layer x_i', where f_θ denotes the encoder, g_θ represents the decoder, and $\Delta(x_i, x_i')$ corresponds to the reconstruction error.

achieve progressive feature learning. Each network layer adjusts encoder parameters through reconstruction error minimization, progressively forming abstract features in high-dimensional space. This hierarchical training framework significantly enhances the expressive capability of deep features.

However, the SDAE-derived features often exhibit redundancy. To address this, the HSOM a nonlinear dimensionality reduction framework combining two Self-Organizing Map (SOM) layers—is employed. The HSOM compresses high-dimensional features into a 1D representation while preserving topological relationships and hierarchical data patterns, thereby mitigating overfitting risks. By integrating the SDAE and HSOM, the preprocessed vibration data is transformed into an interpretable 1D degradation curve that visually tracks the bearing's evolution from health to failure. This approach effectively compresses high-dimensional industrial data, enables intuitive degradation pattern visualization, and provides a reliable, interpretable foundation for RUL prediction through deep feature extraction and hierarchical topology mapping.

2.2 Hybrid MSCNN-BiGRU Architecture

In feature extraction for sequential data, the kernel size of convolutional layers critically impacts model performance. Conventional convolutional neural networks employ single-scale kernels, limiting their ability to capture multi-scale temporal features—smaller kernels focus on local details and high-frequency patterns, while larger kernels capture global trends and low-frequency modes. To address this issue, we propose a multi-Scale 1D Dilated Convolution Module integrated into traditional CNN architectures, enhancing the model's capacity to extract multi-scale features and global information.

This module combines 1D dilated convolution with multi-scale convolution techniques. Dilated convolution [18] expands the receptive field by inserting zeros at fixed intervals into the kernel, increasing coverage without adding parameters.The 1D dilated convolution is a specialized adaptation of the dilated convolution concept from image processing, extended to 1D sequential data. It extracts features from 1D signals with continuous or discrete numerical structures, such as time-series data. This method enables models to effectively capture long-range dependencies and global patterns in 1D signals.The module adapts to varying scales by adjusting dilation rates and employing multiple kernel sizes, enabling comprehensive feature extraction from local to global contexts and adaptive learning of hierarchical contextual information. Specifically, in the first convolutional layer, kernels of diverse sizes and dilation rates process the input to generate three distinct feature maps. These multi-scale features are then concatenated into a unified representation, serving as the input to subsequent layers. This approach overcomes the limitations of traditional single-scale convolution, significantly improving the modeling of long-range dependencies and global structures in 1D signals.

The BiGRU model captures temporal dependencies in time-series data by integrating a self-attention mechanism, which dynamically optimizes feature weights. The self-attention layer computes correlations between time-step features and assigns adaptive weights, allowing the model to prioritize critical degradation patterns. This design enhances the accuracy of remaining useful life prediction and strengthens long-term memory capabilities. By balancing local and global temporal contexts, the model achieves superior performance and generalization in handling complex temporal dependencies and nonlinear degradation dynamics.

As a type of attention mechanism, Self-Attention aggregates information by computing correlations between elements, capturing global relationships, and enhancing the model's adaptability and performance, especially in long sequence tasks. The calculation formula for the self-attention mechanism is as follows:

$$Attention(Q,K,V) = SoftMax\left(\frac{QK^T}{\sqrt{d}}\right) \tag{1}$$

In the formula, Q, K and V represent matrices composed of vectors derived from the input data through different linear transformations;Softmax(·) is the activation function used for normalization; $\sqrt{d}$ is the scaling factor, which prevents the dot product from becoming too large as the dimensionality increases.

3 Experimental Results and Analysis

3.1 Datasets

The PHM2012 dataset [19] is a data resource from the Prognostics and Health Management Challenge organized by the IEEE Reliability Society in 2012, specifically targeting the remaining useful life prediction of rolling bearings.Provided by the FEMTO-ST Institute in France, this dataset captures degradation processes of 17 bearings under three representative operating conditions, For detailed information, refer to Table 1.Regarding data division, From each operating condition, the first two bearing samples are used for training, and the rest for testing. Data mainly includes horizontal and vertical accelerometer readings, with some datasets incorporating temperature sensors. Samples are recorded every 10 seconds at 25.6 kHz for 0.1 s, yielding 2560 points per instance. Each sample is structured as a six-column matrix, containing timestamp components and bi-directional vibration signals.

Table 1. Dateset of PHM2012

Condition	Training Set	Training Set
Rotation: 1800rpm Radial Force: 4000N	Bearing1_1 Bearing1_2	Bearing1_3 Bearing1_4 Bearing1_5 Bearing1_5 Bearing1_7
Rotation: 1650rpm Radial Force: 4200N	Bearing2_1 Bearing2_2	Bearing2_3 Bearing2_4 Bearing2_5 Bearing2_6 Bearing2_7
Rotation: 1500rpm Rotation: 1500rpm	Bearing3_1 Bearing3_2	Bearing3_3

3.2 Training Details and Evaluation Metrics

Training Details. The experiments are implemented in the Pytorch framework. We employ an Nvidia RTX 4060 GPU for accelerated computing, with CUDA technology utilized to optimize training efficiency.The first-layer multi-scale 1D dilated convolutional kernels were configured to 3, 5, and 7 for multi-frequency degradation feature extraction, while layers 2–4 employed a uniform kernel size of 3. Training hyperparameters included a learning rate of 0.0005, 200 epochs, an overfitting control coefficient of 0.18, and a dilation rate of 2 to optimize temporal context coverage.

Evaluation Metrics. To validate the effectiveness of the RUL prediction method, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used to evaluate the accuracy of the model. The smaller the computed loss, the closer the predicted values are to the ground truth, reflecting greater prediction precision.The calculation methods are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}(x_i))^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{n}^{1}|\hat{y}(x_i) - y_i| \tag{3}$$

where n is the number of samples, y_i represents the true value of the i-th sample, and $\hat{y}(x_i)$ denotes the predicted value of the i-th sample.

3.3 Results on PHM2012

This study employs MAE and RMSE as evaluation metrics to comprehensively assess the discrepancies between model predictions and ground-truth values across multiple dimensions. To ensure the reliability and persuasiveness of conclusions, MAE and RMSE were calculated for all test bearings under three representative operating conditions.As detailed in Table 2, which quantifies prediction errors for each test case, both MAE and RMSE metrics exhibit consistently low values across the majority of datasets. These results demonstrate the proposed model's superior predictive performance in most samples, characterized by minimal average absolute deviations from true values. Furthermore, the model demonstrates robust stability and consistency across heterogeneous samples, underscoring its adaptability to diverse operational scenarios.

Table 2. The performance of bearing RUL prediction.

Bearing Name	MAE (%)	RMSE (%)
Bearing1_3	0.119	1.400
Bearing1_4	0.225	2.685
Bearing1_5	0.644	6.335
Bearing1_6	0.622	8.064
Bearing1_7	0.089	3.321
Bearing2_3	0.785	5.842
Bearing2_4	0.868	4.528
Bearing2_5	0.945	10.884
Bearing2_6	0.920	6.411
Bearing2_7	1.524	9.127
Bearing3_3	3.200	11.142

We adopted an improved MSCNN-BiGRU network model to predict the RUL of five test bearings (1_3, 1_4, 1_5, 1_6, and 1_7) under working condition 1. The predictive performance of the model is visually demonstrated, as shown in Fig 3, where the predicted results exhibit a good fit with the actual values. The prediction results are expressed as the bearing degradation ratio, with the full lifespan serving as the baseline (assigned a value of 1). The formula for calculating the remaining life proportion is as follows:

$$RUL(\%) = 1 - DL(\%) \tag{2}$$

Where DL represents the degradation ratio (the proportion of elapsed time), RUL represents the remaining life ratio.

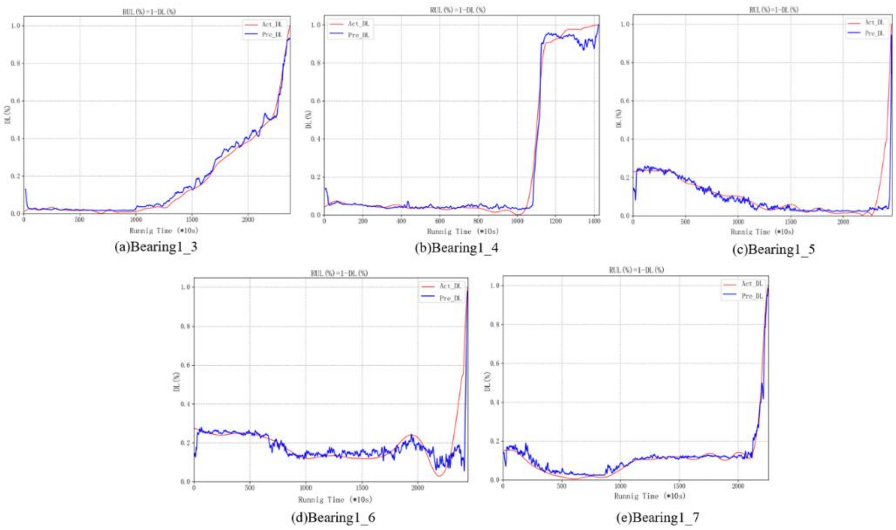

Fig.3. The RUL prediction results of the MSCNN-BiGRU model on the test set under working condition 1.

3.4 Ablation experiment analysis

To validate the effectiveness of the key components in the proposed MSCNN-BiGRU model, this section compares the performance differences of various configurations by progressively removing or replacing core components, using the PHM2012 dataset. The comparative models involved in the ablation experiment are as follows:

1. BiGRU: This model integrates the bidirectional gating mechanism on top of the GRU to capture both forward and backward temporal dependencies.
2. CNN-BiGRU: This model combines convolutional neural networks with bidirectional gated recurrent units, utilizing convolutional layers to extract spatial local features and leveraging the bidirectional gating mechanism to capture temporal dependencies.

3. MCNN-BiGRU: Building upon traditional convolutional neural networks, this model incorporates one-dimensional dilated convolution techniques and multiscale ideas. Through a multi-branch structure, it synchronously extracts both local details and global long-range features from the sequence data, while using bidirectional gated recurrent units to capture temporal dependencies. This enhances the model's ability to represent complex sequence patterns.
4. MSCNN-BiGRU: Based on MCNN-BiGRU, this model adds a degradation curve construction module to accurately capture deep structural features reflecting the changes in bearing conditions.

As shown in Table 3, the proposed MSCNN-BiGRU model achieves the lowest values in both MAE and RMSE metrics among the compared methods, validating its superior predictive performance. Compared to models that only capture temporal dependencies through the bidirectional gating mechanism, MCNN-BiGRU introduces CNN convolutional layers and adds a spatial local feature extraction module. This effectively reduces the prediction bias caused by neglecting the spatial correlation of sensor signals, as seen in BiGRU. MCNN-BiGRU captures both local details and global long-range features of the vibration signals synchronously through one-dimensional dilated convolutions and a multiscale branching structure. MSCNN-BiGRU further enhances this by adding a degradation curve construction module, explicitly modeling the deep structural features that reflect bearing condition changes. This design specifically strengthens the model's sensitivity to key stages of RUL, significantly improving prediction accuracy in the late stages of degradation. In addition, as shown in Table 5, compared to the second-best model, our model achieves a 1.7% reduction in parameter count while improving the training speed by 13.9%.

Table 3. Ablation Experiment Performance Evaluation of Different Models

Network model	Bearing1_3		Bearing1_4		Bearing1_5	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
BiGRU	0.592	2.537	0.711	4.376	2.535	9.693
CNN-BiGRU	0.516	2.164	0.655	3.911	1.814	8.712
MCNN-BiGRU	0.375	1.731	0.576	3.463	0.987	7.648
MSCNN-BiGRU	**0.088**	**1.400**	**0.174**	**2.685**	**0.644**	**6.335**

Table 4. Comparison of training speed and number of training parameters

Network model	Training speed	Volume of training parameters
MCNN-BiGRU	86min	14,159,895
Ours	**74min**	**13,993,487**

3.5 Comparative analysis with other algorithms

To thoroughly demonstrate the superiority of the MSCNN-BiGRU model proposed in this study, we selected a range of representative and widely adopted classical models and SOTA models for comparison and analysis. The models included in the comparison are BiLSTM, GAU, Transformer, BiGRU-Att, BiLSTM-Att, and iTransformer. Under consistent parameter configurations, the MAE metric is used to evaluate the prediction performance of all models. Taking five bearing samples (1_3 to 1_7) under Condition 1 as an example, the RUL prediction results of different models are shown in Table 5. A comprehensive comparison and analysis of the final prediction performance of different bearing sample models are presented. Compared to other models, our method achieves the lowest MAE across all datasets.On the PHM 2012 dataset, compared to the second-best method, our approach reduces the MAE by 27.3%, 29.5%, 11.9%, 13.0%, and 13.8%, respectively, validating the effectiveness of our method. The baseline model BiLSTM lacks the ability to extract spatial features, which results in larger prediction errors. BiGRU-Att, BiLSTM-Att, and GRU, which incorporate self-attention mechanisms to improve the capture of critical information, still fail to adequately capture the importance of deep features, thus limiting their prediction performance. The SOTA Transformer model benefits from its advanced internal attention module but has limited capacity in extracting deep features, leaving room for improvement in prediction accuracy. In contrast, MSCNN-BiGRU covers the key aspects of SOTA models and effectively utilizes deep degraded features through the in-depth exploration and integration of historical and current information. Therefore, it is more suitable for RUL prediction of rotating machinery.

Table 5. MAE of different forecasring models

Network model	Bearing 1_3	Bearing 1_4	Bearing 1_5	Bearing 1_6	Bearing 1_7
BiLSTM	0.612	0.635	1.724	1.236	0.672
GAU	0.311	0.435	0.927	1.001	0.504
Transformer	0.355	0.476	0.876	1.020	0.375
BiGRU-Att	0.942	1.154	1.845	1.788	0.937
BiLSTM-Att	0.207	0.326	0.885	0.815	0.188
iTransformer	0.121	0.247	0.731	0.715	0.218
Ours	**0.088**	**0.174**	**0.644**	**0.622**	**0.188**

4 Conclusion

In this paper, we propose an MSCNN-BiGRU network for RUL prediction of rolling bearings. First, we develop a degradation curve construction method based on SDAE and HSOM. The SDAE extracts deep features from raw vibration signals, while the HSOM performs hierarchical and ordered dimensionality reduction on SDAE-processed features to precisely capture deep structural characteristics that reflect bearing state transitions, thereby constructing degradation curves. To enhance prediction accuracy, we employ multi-scale convolution with varied kernel sizes to capture multi-level features and dilated convolution to model long-range contextual relationships. Additionally, a self-attention mechanism is integrated to highlight prediction-critical features, ensuring both efficient model training and improved prediction performance. Experimental validation on the PHM2012 open dataset demonstrates that our method achieves lower prediction errors and higher accuracy in bearing RUL estimation compared to baseline approaches. Future work will focus on enhancing cross-domain generalization capabilities through meta-learning, and developing an FPGA-based edge-cloud collaborative real-time RUL monitoring system to enable timely and precise maintenance decision-making for industrial equipment.

Acknowledgments. This work was supported by the Research and Development of Soil Multi-Parameter Composite Sensors and Intelligent Monitoring Systems (2024CXGC010905), in part by the Research, Development, and Application of Intelligent Central Air Conditioning and Integrated IoT Configuration Systems for High-End Residential Buildings (2024TSGC0603), in part by the Construction and Application Demonstration of R&D Public Service Platforms for Intelligent Innovation in New-Type R&D Institutions (YDZX2023050), in part by the Intelligent Manufacturing Empowerment Platform Based on Industrial Internet and Its Applications (YDZX2024121), in part by the Research on Key Technologies for Building Trusted Data Spaces and High-Quality Data Elements, and Industry Applications (2024ZDZX08), in part by the Research and Development of Key Sensing Technologies for Growth Factors and Vital Signs of Greenhouse Crops (2023TSGC0111), in part by the Research on Key Technologies for Intelligent Management and Control of Agricultural Machinery and Information Platforms, and Their Applications (2023TSGC0587), by the Research and Application of Key Technologies for Intelligent Management and Control of Agricultural Machinery and Information Platforms in Tai'an City (2023TATSGC042), and by the Innovation Capability Enhancement Project for Technology-based Small and Medium Enterprises of Shandong Province, titled Industrial Equipment Health Monitoring and Diagnosis System (2023TSGC0201).

References

1. Chen, J.X., Mao, W.T., Liu, J., et al.: Remaining useful life prediction of bearing based on deep temporal feature transfer. Control and Decision **36**(7), 1699–1706 (2021)
2. Shen, B., Chen, B., Zhao, C., et al.: A review of deep learning in machinery fault prediction and health management. Mach. Tool Hydraul. **49**(19), 162–171 (2021)
3. Zeng, S.K., Pecht, M.G., Wu, J.: Status and perspectives of prognostics and health management technologies. Acta Aeronautica Et Astronautica Sinica **26**(5), 626–632 (2005)

4. Ma, M., Mao, Z.: Deep-convolution-based LSTM network for remaining useful life prediction. IEEE Trans. Indust. Inform. **17**, 1658–1667 (2020)
5. Wan, S., Li, X., Zhang, Y., et al.: Bearing remaining useful life prediction with convolutional long short-term memory fusion networks. Reliab. Eng. Syst. Saf. **224**, 108528 (2022)
6. Yu, W., Pi, D., Xie, L., Luo, Y.: Multiscale attentional residual neural network framework for remaining useful life prediction of bearings. Measurement **177**, 109310 (2021)
7. Wen, L., Li, X.Y., Gao, L., et al.: A new convolutional neural network-based data-driven fault diagnosis method. IEEE Trans. Ind. Electron. **65**(7), 5990–5998 (2018)
8. Chen, D., Choo, M.K., Ng, S.C.: Modeling and Forecasting of nanoFeCu Treated Sewage Quality Using Recurrent Neural Network (RNN). Computation **11**(2), 39–48 (2023)
9. Mao, Z., Wu, L., Ma, Y., et al.: CAN Bus Intrusion Detection Model Based on DBN and GRU with Attention Mechanism. Journal of Wuhan University (Natural Science Edition) **69**(5), 598–608 (2023)
10. Tang, X., Xi, H., Chen, Q., et al.: Rolling bearing remaining useful life prediction based on LSTM-transformer algorithm. In: Proceedings of IncoME-VI and TEPEN 2021: Performance Engineering and Maintenance Engineering. pp. 207–215. Springer, Cham (2022)
11. Mou, H., Zheng, J., Hu, C., et al.: Remaining useful life prediction of multi-degrading equipment based on CDBN and BiLSTM. J. Aeronaut. **43**(07), 308–319 (2022)
12. Wang, X., Wang, T., Ming, A., et al.: Deep spatiotemporal convolutional-neural network-based remaining useful life estimation of bearings. Chin. J. Mech. Eng. **34**(1) (2021)
13. Yang, J., Peng, Y., Xie, J., et al.: Remaining useful life prediction method for bearings based on LSTM with uncertainty quantification. Sensors **22**, 4549 (2022)
14. Cao, Y., Ding, Y., Jia, M., et al.: A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. Reliab. Eng. Syst. Saf. **215** (2021)
15. Pan, Y., Hong, R., Chen, J., et al.: A hybrid DBN-SOM-PF-based prognostic approach of remaining useful life for wind turbine gearbox. Renew. Energ. **152**, 138–154 (2020)
16. Vincent, P., Larochelle, H., Lajoie, I., et al.: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research* **11**(12), 3371–3408 (2010)
17. Vincent, P., Larochelle, H., Bengio, Y., et al.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, pp. 1096–1103. ACM, New York (2008)
18. Yu, F., Koltun, V., Funkhouser, T.: Dilated residual networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. pp. 472–480. IEEE (2017)
19. Nectoux, P., Gouriveau, R., Medjaher, K., et al.: PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In: 2012 IEEE International Conference on Prognostics and Health Management. pp. 1–8. IEEE (2012)