

RUL prediction method based on cross-view hybrid network model

Yandi Ai¹, Dong Fang^{2,*}, Zhiping Tian¹, and Kaiyang Yan³

¹School of Frontier Interdisciplinary, Hunan University of Technology and Business Changsha, China

²School of Computer Science, Hunan University of Technology and Business, Changsha, China

³School of Intelligent Engineering and Smart Manufacturing, Hunan University of Technology and Business, Changsha, China

Keywords: RUL prediction, Cross-view learning, Dual-channel feature extraction.

Abstract. Remaining Useful Life (RUL) prediction has become a core technology in the field of prognostics and health management (PHM). However, due to the non-stationarity, weak signal characteristics and concurrent multiple faults of original signals, the estimation of RUL in a single view tends to ignore the structural relationship of samples in different spaces. To this end, this paper designs a RUL prediction framework based on a cross-view hybrid network model (CVHNet). Firstly, a dual-channel feature extraction hybrid network (DCF-HybridNet) is constructed. The original features are decomposed into time-frequency features and implicit spatial features through short-time Fourier transform (STFT) and Gramian angular difference field (GADF) - CNN, and then fused into a comprehensive feature representation with cross information. Secondly, a RUL regression algorithm integrating Transformer encoder and nonlinear fitter is developed to automatically learn the correlation between features in different views and predict RUL. The XJTU-SY rolling bearing dataset was used as an example for experimental verification. The results show that the Average CRA of the CVHNet method proposed in this paper is 0.9195 under three different working conditions. Compared with the methods in the same field, it shows superior performance and can provide strong support for the predictive maintenance of equipment.

1 Introduction

With the development of the Internet of Things and Industry 4.0, RUL prediction has become increasingly important in the manufacturing and maintenance fields. RUL prediction models can use data such as vibration, pressure, current and temperature collected from sensors to estimate and warn of potential failure time nodes of equipment in

*Corresponding author: fdvynq@163.com

advance. In recent years, under the combined effect of massive industrial monitoring data and the exponential leap in computing power, data-driven prediction methods have become a research hotspot in the field of RUL prediction^[1-2]. Relying on rich data resources, it automatically identifies complex patterns, extracts high-value features, and guides prediction decisions, effectively reducing the dependence on manually preset model frameworks or rules.

Feature quality is regarded as a key factor in driving the improvement of RUL prediction accuracy and model effectiveness^[3]. How to extract the most discriminative and representative fault feature information from the huge amount of raw data constitutes one of the core challenges of RUL prediction technology. To this end, Mi et al.^[4] combined the hierarchical clustering algorithm with the feature set optimization technology of shared information theory to effectively screen out the most representative feature set. Zhu et al.^[5] proposed a time-frequency domain feature extraction method based on wavelet transform, which can efficiently extract a representation that combines local features of time and frequency. He et al.^[6] proposed a cascade framework combining deep separable convolution (DSC) and dilated causal convolution (DCC) to achieve spatiotemporal degradation feature extraction. Ma et al.^[7] proposed a channel attention (ECA) mechanism to extract local and global features. Wen et al.^[8] proposed a deep residual shrinkage network (DRSN) for adaptive feature extraction, which can automatically learn features of the original signal. Although existing research has done relevant work in feature extraction and fusion, it has failed to achieve an effective balance in integrating time-frequency domain features with the intrinsic spatial structure characteristics of the signal.

As the core component of RUL prediction, the prediction model carries the key mission of extracting knowledge from complex data, revealing the laws of system degradation and predicting future states^[9]. For example, Magadán et al.^[10] used BiLSTM to predict the remaining service life of bearings. Liu et al.^[11] combined BiLSTM with self-attention mechanism to construct a BiLSTM-SA model to achieve RUL prediction. Que et al.^[12] proposed a RUL prediction model based on dynamic time warping (DTW) attention mechanism and GRU to improve the adaptability of the prediction model. Huang et al.^[13] proposed a RUL prediction method that integrates deep convolutional neural network and multi-layer perceptron (DCNN-MLP) to effectively realize RUL prediction. Zhu et al.^[14] proposed a recursive graph convolutional network with uncertainty estimation (RGCNU) for RUL prediction, which improved the robustness of the model to noisy data. Although existing studies have made many explorations on RUL prediction models, existing models pay less attention to the processing of negative values in signals and the computational efficiency of prediction algorithms, which may lead to the amplification of the deviation of model prediction results and pose a potential risk to prediction-based decision-making.

To overcome the above research deficiencies, the main contributions of this paper are as follows:

(1) A RUL prediction framework based on a cross-view hybrid network model (CVHNet) is proposed. The framework integrates a multi-dimensional feature extraction module and a deep learning network for RUL estimation, aiming to achieve efficient prediction under complex conditions with non-stationary and weak signal features.

(2) A dual-channel feature extraction hybrid network (DCF-HybridNet) was constructed. The network captures the time-frequency domain features in the monitoring data through STFT, and combines GADF and CNN to mine the spatial structure information in the vibration signal, so as to achieve multi-view and in-depth analysis of the signal's spatiotemporal features.

(3) A RUL regression algorithm integrating Transformer encoder and nonlinear fitter is designed. The weight-sharing multi-head attention mechanism is used to automatically learn the importance and complementarity between features of different views, and the

nonlinear fitter is integrated to handle negative input and gradient vanishing problems, so as to improve the model's understanding and prediction accuracy of dynamic working condition data.

2 RUL prediction method based on CVHNet

The CVHNet model mainly consists of a data preprocessing layer, a dual-channel feature extraction module, and a RUL predictor, which are described in detail as follows.

2.1 RUL prediction framework

The goal of the RUL prediction framework of the CVHNet model proposed in this paper is to more fully and comprehensively utilize the implicit characterization information in the monitoring data and provide a more effective and reliable RUL prediction method to support equipment maintenance and decision-making.

The RUL prediction framework based on the CVHNet model is shown in Figure 1. Specifically, the monitoring data is first cleaned of noise, normalized, and sliced by window functions in the data preprocessing layer to ensure the quality and standardization of the model input data. Then, since the quality of the features directly affects the final prediction results of the model, it is necessary to perform further feature engineering on the sliced data sequence. In the dual-channel feature extraction module, GADF-CNN is used to convert the one-dimensional signal data into an image and extract its spatial features, and STFT is used to extract the time-frequency domain features to obtain information of different modalities so that the model can more comprehensively consider the characteristics of the monitoring data. Finally, the feature information extracted from the two different modal information is input, the feature information is fused, and the RUL predictor is used to perform RUL prediction.

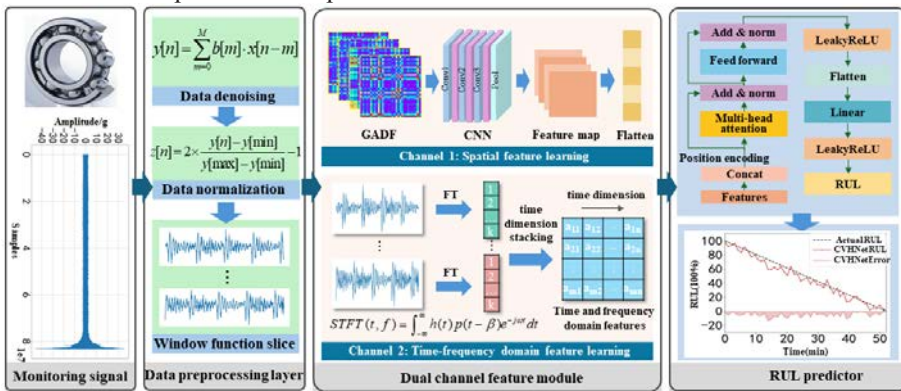


Fig. 1. RUL prediction framework based on CVHNet.

2.2 Dual-channel feature extraction across views

2.2.1. Spatial feature extraction based on GADF-CNN.

The process of using GADF-CNN to implement deep spatial feature extraction and representation learning of monitoring signals is shown in Figure 2.

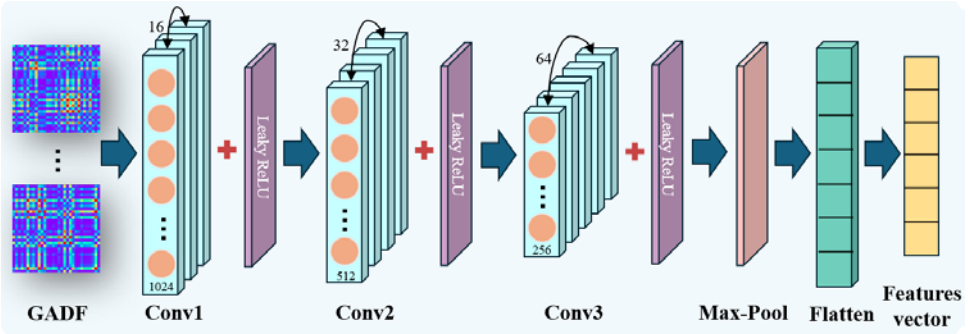


Fig. 2. Spatial feature extraction based on GADF-CNN.

GADF is an image coding method that uses the Gram Matrix to evaluate the linear correlation in a vector set for visual transformation, thereby converting continuous time series data into a two-dimensional image format and retaining the temporal dependence in the time series^[15]. From the data preprocessing layer given above, we get L time series S_i with a length of 1000. Then, we encode the value of S_i as the angle cosine α through formula (1), and encode the time node t_n as the radius R . Thus, we can get the time series S_i represented in the polar coordinate system.

$$\begin{cases} \alpha = \arccos(s_j), -1 \leq s_j \leq 1, s_j \in S_i \\ R = \frac{t_j}{N} \end{cases} \quad (1)$$

where s_j is the j th data in S_i ; t_j is the time node of s_j ; N is the constant factor of the regularized polar coordinate system space, which is used to adjust the span of the polar coordinate system.

After S_i is converted into a polar coordinate system, the temporal correlation within different time intervals can be identified by considering the angle difference between each point through GADF, which is defined as shown in formula (2).

$$\text{GADF} = \begin{bmatrix} \sin(\alpha_1 - \alpha_1) & \cdots & \sin(\alpha_1 - \alpha_n) \\ \sin(\alpha_2 - \alpha_1) & \cdots & \sin(\alpha_2 - \alpha_n) \\ \vdots & \ddots & \cdots \\ \sin(\alpha_n - \alpha_1) & \cdots & \sin(\alpha_n - \alpha_n) \end{bmatrix} \quad (2)$$

where α_i represents the i -th polar angle in α .

Through the above GADF transformation, SI is effectively mapped into image data. Subsequently, CNN is used to mine deep spatial features from these image data. CNN mainly integrates several key components: convolutional layer, activation layer, pooling layer and Flatten layer. By alternately stacking multiple convolutional, activation and pooling layers, a deep CNN is constructed.

The image data is first subjected to a convolution operation, and a filter is used to extract key visual features from the input image data to generate a feature map. The convolution operation formula is:

$$x_j^l = \sum_{i \in M_j} x_i^{l-1} \cdot \omega_{ij}^l + b_j^l \quad (3)$$

where x_j^l is the output of layer l , x_i^{l-1} represents the output of layer $l-1$, M_j is the feature set of layer $l-1$, ω_{ij}^l represents the convolution kernel weight, b_j^l is the bias, and \bullet represents the convolution operator.

Use ReLU as the activation function to perform nonlinear transformation on the feature map after convolution operation, as shown in formula (4):

$$y_i = f(x_j^l) = f\left(\sum_{i \in M_j} x_i^{l-1} \bullet \omega_{ij}^l + b_j^l\right) \tag{4}$$

The maximum pooling strategy is used to reduce the number of parameters in the network, thereby alleviating the risk of overfitting and improving computational efficiency. The pooling calculation is shown in formula (5):

$$P_i^{l+1} = \max_{(r-1)S+1 < t < rS} \{q_i^l(t)\} \tag{5}$$

The final output feature map of CNN is transformed into a one-dimensional array through a Flatten function.

2.2.2. Spatial feature extraction based on GADF-CNN

For a given time series $X = \{x_1, x_2, \dots, x_n\}$, STFT first uses a window function to segment the continuous time series, selects time domain samples of a fixed length, and then applies FFT to these samples to obtain a frequency distribution image at a specific time period t , namely, a spectrogram. By gradually moving the position of the window function on the time axis and repeating this process, a series of local spectra can be continuously constructed to jointly depict the global time-frequency behavior of the signal. The STFT calculation is shown in formula (6):

$$STFT(t, f) = \int_{-\infty}^{\infty} h(t) p(t - \beta) e^{-j\omega t} dt \tag{6}$$

where $h(t)$ is the time domain signal; $p(t - \beta)$ is the time window centered on β , and the window function is Hamming, and its calculation formula is as follows:

$$w(n) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \tag{7}$$

where $w(n)$ represents the n th sample point, and N is the total length of the window function.

By performing STFT on the signal, we can reveal its dynamic evolution in the time series and its inherent frequency domain characteristics. This conversion process transforms the complex data stream into an intuitive two-dimensional matrix rich in time-frequency domain information, where the horizontal axis represents the extension of time and the vertical axis maps the distribution of the spectrum. Each matrix element carries the frequency energy information unique to a specific moment.

2.3 RUL estimation based on lr-encoder

In order to overcome the problems of poor real-time performance of serial operations, poor capture of long-term dependencies, and neglect of negative components in existing research, the Transformer encoder model based on parallel operation is combined with a nonlinear fitter to construct an LR-Encoder model to achieve RUL estimation. The LR-Encoder structure is shown in Figure 3.

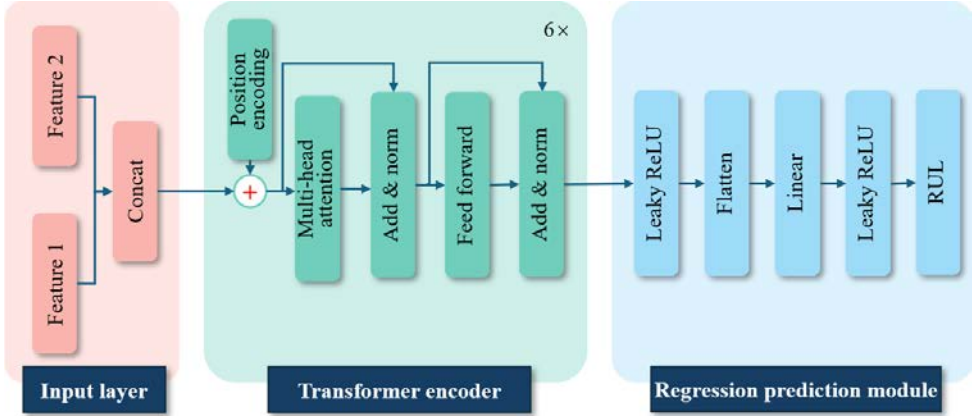


Fig. 3. RUL predictor based on LR-Encoder.

For the features extracted in the previous article, the spatial features are concatenated with the time-frequency features of the corresponding time point to form a new fusion feature vector, which represents the comprehensive features of the time point. Then, the position encoding of F is performed using formula (8).

$$\begin{cases} PE_{(pos,2i)} = \sin(p / 10000^{2i/dx}) \\ PE_{(pos,2i+1)} = \cos(p / 10000^{2i/dx}) \end{cases} \quad (8)$$

where $PE_{(pos,2i)}$ and $PE_{(pos,2i+1)}$ represent the position encoding of even-dimensional and odd-dimensional features respectively; p is the position of the corresponding feature vector in the sequence; dx is the dimension of the added feature vector. F_{PE} with position encoding can be expressed as:

$$F_{PE} = F + PE \quad (9)$$

The first layer of the encoder is a multi-head attention mechanism, and the input vector is first converted into query, key, and value vectors. Then, the attention output of F_{PE} is calculated using the scaled dot product attention, and the calculation formula is as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK}{\sqrt{d}}\right)V \quad (10)$$

The multi-head attention mechanism learns the attention values in different subspaces by splicing multiple single attention mechanisms, thereby capturing richer data features and finally obtaining the attention information in all subspaces through parallel operations. After the residual connection and layer normalization are completed, it is sent to the feedforward neural network for nonlinear transformation.

In the regression prediction module, the Leaky ReLU activation function is first applied to introduce a non-zero slope, which effectively promotes the gradient flow of negative features and ensures the model's sensitivity and learning ability to the full range of data. Secondly, the Flatten layer operation is implemented to reduce the input dimension to a one-dimensional vector. Then, a fully connected Linear layer is integrated to map the previously refined feature vectors to the predicted output of RUL by learning a set of weight parameters. In order to further enhance the expressive power of the model and improve its adaptability to continuous output, the Leaky ReLU activation function is deployed again after the Linear layer to ensure that the output of the model in the prediction stage has rich nonlinear features, and finally output the RUL value of the device.

3 Results and analysis

3.1 Experimental setup

3.1.1. Dataset description.

Taking the RUL prediction of rolling bearings as an example, the effectiveness and practicality of the proposed method are demonstrated. The data set is the XJTU-SY rolling bearing accelerated life test data set^[16]. The test bearing model is the LDKUER204 rolling bearing. The data set has three working conditions, and 5 bearings are processed under each working condition. The sampling frequency is 25.6kHz, the sampling time is 1.28s, and the sampling interval is 1min. Bearing1_5, bearing2_5, and bearing3_5 are selected as the test set, and the remaining data are used as the training set for building and training the model.

3.1.2. Evaluation indicators.

Cumulative Relative Accuracy (CRA) is used as an evaluation indicator to measure the RUL prediction effect of rolling bearings. The advantage of the CRA indicator is that it can transcend the limitations of a single prediction and effectively reveal the consistency and long-term expressiveness of the model prediction trend through cumulative calculation of continuous time series data. Its calculation is shown in formula (14).

$$CRA = \sum_{i=1}^n w_i RA(T_i) \tag{11}$$

where $RA(T_i)$ is the relative prediction accuracy at time T_i , and the formula is as follows:

$$RA(T_i) = 1 - \frac{|RUL_{Act}(T_i) - RUL_{Pre}(T_i)|}{RUL_{Act}(T_i)} \tag{12}$$

$$RUL(T_k) = \inf\{r : f(r + T_k) \geq \omega | F\} \tag{13}$$

3.2 Analysis of results

In order to verify the performance of the proposed method under different working conditions, bearing1_5, bearing2_5, and bearing3_5 under different working conditions are

selected as test sets for model verification. The fitting results of the model prediction values and the actual RUL values of the three experiments are shown in Figure 4.

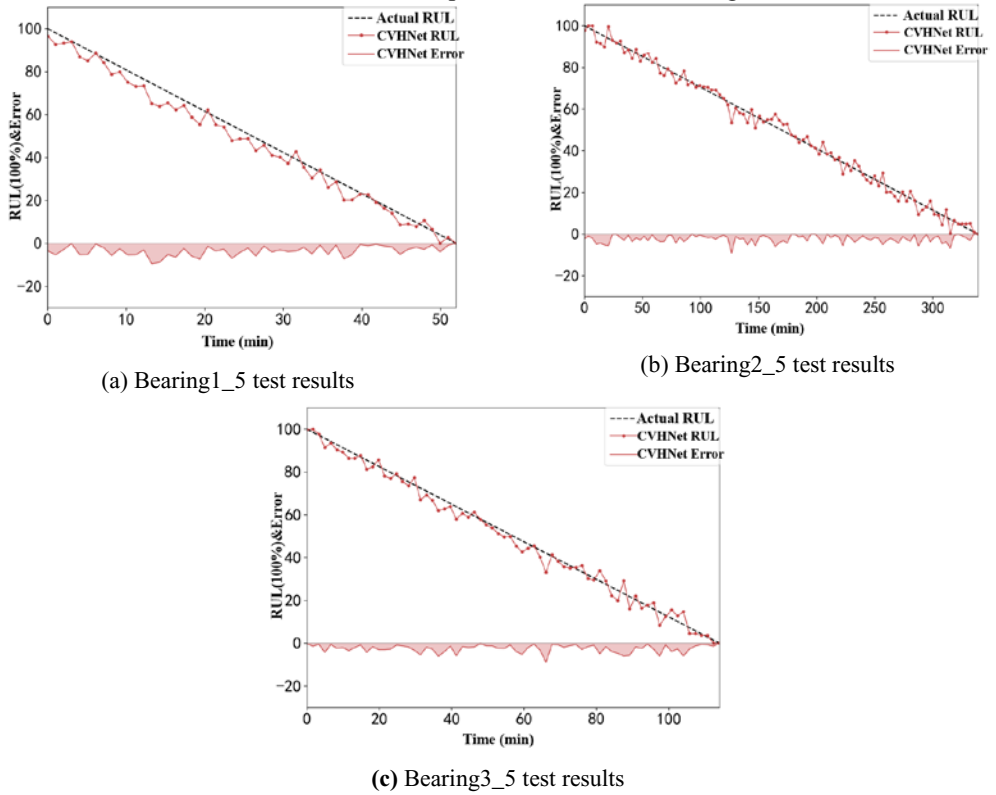


Fig. 4. Model test results under different working conditions.

As can be seen from Figure 4, the model test results fit the degradation trend of rolling bearings well. Through the verification of three independent tests, it is observed that the absolute error between the model's predicted value and the target value is maintained at a low level, and the fluctuation range is relatively stable, indicating that the output result is consistent with the target value, and the model calculation result is highly credible as a predicted value.

Table 1. Evaluation of CVHNet model prediction performance.

Model	CRA			
	bearing1_5	bearing2_5	bearing3_5	Average
CVHNet	0.9396	0.9014	0.9176	0.9195

As shown in Table 1, the CRA of RUL prediction of the CVHNet model on the three test sets of bearing1_5, bearing2_5, and bearing3_5 are 0.9396, 0.9014, and 0.9176, respectively. The results of the three test experiments all exceeded the cumulative relative accuracy of 90%, and the average CRA reached 0.9195, highlighting the superior performance of the constructed model in simulating the degradation process of rolling bearings, which strongly illustrates the high precision and generalization ability of the CVHNet model in the prediction of RUL of rolling bearings.

3.3 Comparative method analysis

To further verify the effectiveness and superiority of the CVHNet model proposed in this paper, this paper uses three comparison schemes, Transformer, LSTM, and DBN, to conduct comparative experiments with the CVHNet model.

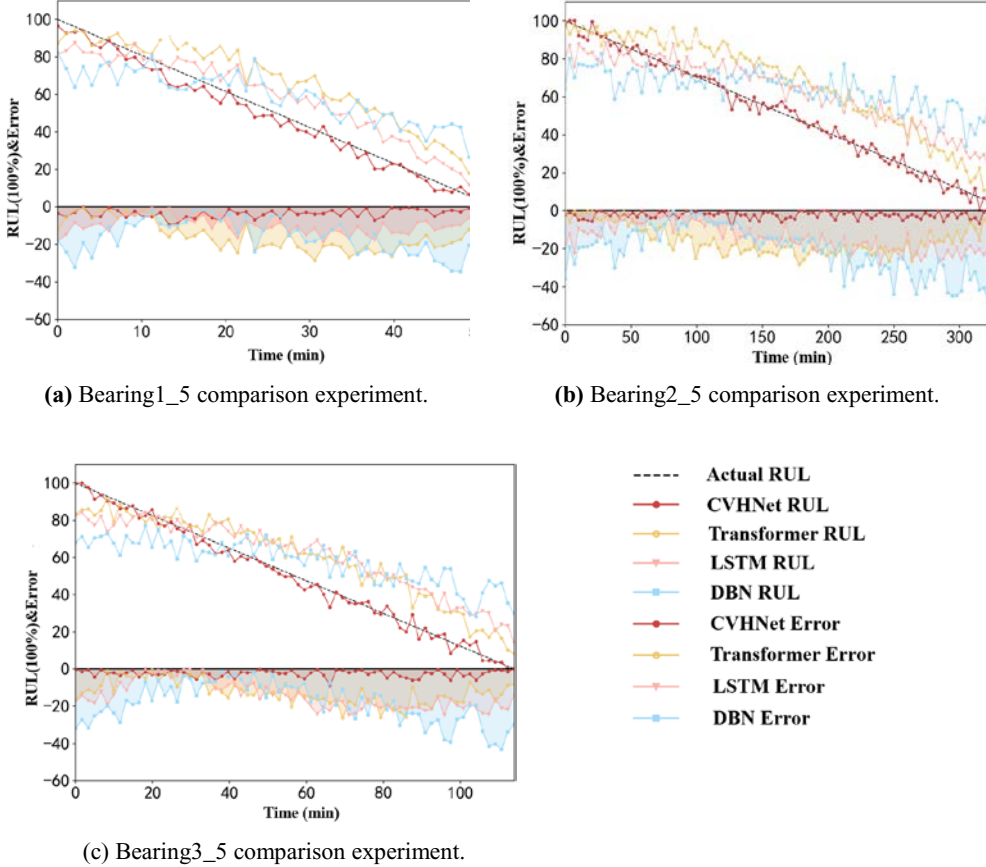


Fig. 5. Comparative experiments under different working conditions.

The RUL prediction model of the comparison scheme is constructed through the training set data, and the test experiment is carried out. The prediction fitting curves and absolute errors of the comparison scheme on bearing1_5, bearing2_5, and bearing3_5 are shown in Figure 8. The yellow curve represents the Transformer prediction result, the pink curve represents the LSTM prediction result, and the blue curve represents the DBN prediction result. The error legend corresponds to the prediction curve. It can be seen that the three comparison schemes have different prediction accuracy in predicting the RUL of rolling bearings compared with the CVHNet model in this paper. In order to more intuitively and scientifically evaluate the prediction performance of the four schemes, their CRAs are further calculated and compared, as shown in Table 2.

As shown in Table 2, the prediction results of the DBN scheme in the three tests were the worst, which were 0.7905, 7291, and 0.7456 respectively, and the average CRA was only 0.755. For the industrial-grade RUL prediction task, its accuracy was insufficient; the average prediction results of the Transformer scheme and the LSTM scheme were close, which were 0.8275 and 0.8223 respectively. It is worth noting that on bearing1_5, the prediction performance of LSTM was better than that of Transformer, while on bearing1_5

and bearing3_9, the prediction results of Transformer were better, indicating that Transformer has more advantages in dealing with tasks such as rolling bearings with a relatively long degradation process. The prediction accuracy of these two comparison schemes is improved compared with the DBN scheme, but is lower than the 0.9195 of the CVHNet model in this paper. By comparing the experimental results, the CVHNet model proposed in this paper is more suitable for this learning task, with higher recognition accuracy and more superiority.

Table 2. Evaluation of CVHNet model prediction performance.

Models	CRA			
	bearing1_5	bearing2_5	bearing3_5	Average
CVHNet	0.9396	0.9014	0.9176	0.9195
Transformer	0.8172	0.8453	0.8201	0.8275
LSTM	0.8547	0.7959	0.8164	0.8223
DBN	0.7905	0.7291	0.7456	0.7550

4 Conclusion

This paper innovatively proposes a RUL prediction framework based on the cross-view hybrid network model (CVHNet), which is divided into three parts: data preprocessing layer, dual-channel feature extraction module and RUL predictor. The proposed method fully considers the correlation and complementarity between different data views, uses short-time Fourier transform, Gram difference field and convolutional neural network technology to extract key features of different spaces, and combines Transformer encoder and nonlinear Regressors make up for the shortcomings of single feature predictive modeling in processing complex data. Experimental results show that the prediction accuracy of the CVHNet model is better than the traditional LSTM, Transformer and DBN models under various working conditions, demonstrating its significant advantages in multi-view feature extraction and integration. Especially when dealing with long-term series data and complex spatial structure information in the bearing degradation process, the prediction effect of the proposed model is particularly outstanding. Future work can further explore multi-source data integration, online prediction real-time and cross-working condition transfer learning to achieve high performance and generalizable intelligent maintenance.

References

1. Omri N, Al Masry Z, Mairot N, et al. Industrial data management strategy towards an SME-oriented PHM[J]. *Journal of Manufacturing Systems*, 2020, 56: 23-36.
2. Cheng C, Ma G, Zhang Y, et al. A deep learning-based remaining useful life prediction approach for bearings[J]. *IEEE/ASME transactions on mechatronics*, 2020, 25(3): 1243-1254.
3. Zhao H, Liu H, Jin Y, et al. Feature extraction for data-driven remaining useful life prediction of rolling bearings[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-10.

4. Mi J, Liu L, Zhuang Y, et al. A synthetic feature processing method for remaining useful life prediction of rolling bearings[J]. *IEEE Transactions on Reliability*, 2022, 72(1): 125-136.
5. Zhu J, Chen N, Peng W. Estimation of bearing remaining useful life based on multiscale convolutional neural network[J]. *IEEE Transactions on Industrial Electronics*, 2018, 66(4): 3208-3216.
6. He J, Wu C, Luo W, et al. Remaining Useful Life Prediction and Uncertainty Quantification for Bearings Based on Cascaded Multi-scale Convolutional Neural Network[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023.
7. Ma P, Li G, Zhang H, et al. Prediction of Remaining Useful Life of Rolling Bearings Based on Multiscale Efficient Channel Attention CNN and Bidirectional GRU[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 1-13.
8. Wen J H, Wu R S, Li S Y, et al. Bearing residual life prediction method based on DRSN and optimized BiLSTM[J]. *Computer Integrated Manufacturing System*, 2024, 30(5): 1877-1888.
9. Li W, Zhang L C, Wu C H, et al. A data-driven approach to RUL prediction of tools[J]. *Advances in Manufacturing*, 2024, 12(1): 6-18.
10. Magadán L, Granda J C, Suárez F J. Robust prediction of remaining useful lifetime of bearings using deep learning[J]. *Engineering Applications of Artificial Intelligence*, 2024, 130: 107690.
11. Liu J, Yang Z, Xie J, et al. A Feature Fusion-Based Method for Remaining Useful Life Prediction of Rolling Bearings[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023.
12. Que Z, Jin X, Xu Z. Remaining useful life prediction for bearings based on a gated recurrent unit[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-11.
13. Huang C G, Huang H Z, Li Y F, et al. A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing[J]. *Journal of Manufacturing Systems*, 2021, 61: 757-772.
14. Zhu Q, Xiong Q, Yang Z, et al. RGCNU: Recurrent Graph Convolutional Network With Uncertainty Estimation for Remaining Useful Life Prediction[J]. *IEEE/CAA Journal of Automatica Sinica*, 2023, 10(7): 1640-1642.
15. Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline[C]//2017 International joint conference on neural networks (IJCNN). *IEEE*, 2017: 1578-1585.
16. Wang B, Lei Y, Li N, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. *IEEE Transactions on Reliability*, 2018, 69(1): 401-412.