Proceedings of Machine Learning Research vol 278:1–11, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

The Remaining Useful Life Prediction of Bearings Based on ICPO-TCN

Jiayu Tian 15668620711@163.COM
Dalian Jiaotong University, Shahekou District-No.794, Huanghe Road, Dalian, China

Tiantian Liang* LIANGTIANTIAN1122@163.COM
Dalian Jiaotong University, Shahekou District-No.794, Huanghe Road, Dalian, China

Zhuangzhuang Ma 1771233079@QQ.COM
Dalian Jiaotong University, Shahekou District-No.794, Huanghe Road, Dalian, China
**Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

In modern industrial equipment maintenance management, bearings, as key rotating components, perform a crucial role in maintaining the stable performance of machinery. Therefore, this paper proposes a bearing Remaining Useful Life (RUL) prediction method based on the Improved Crested Porcupine Optimizer-Time Convolutional Network (ICPO-TCN). Firstly, the improved crested porcupine optimizer (ICPO) is used to search for the best number of modes and penalty factor in VMD, enabling the selection of effective components for signal reconstruction, noise reduction, and enhanced time-frequency feature extraction. A feature dataset is then constructed by combining the selected time-domain and frequency-domain characteristics. Next, reducing the dimensionality by kernel principal component analysis (KPCA), which is then used as input for the TCN model. Finally, ICPO is again employed to optimize the convolution kernel size and learning rate of the TCN to improve RUL prediction accuracy. Experimental results demonstrate that ICPO-TCN outperforms traditional TCN and LSTM models, achieving higher prediction accuracy.

Keywords: Bearing, RUL, feature extraction, Improved Crested Porcupine Optimizer, VMD, TCN

1. Introduction

Bearing performance significantly influences equipment effectiveness and operational safety. As operating time increases, bearings experience wear and fatigue, ultimately leading to failure. Therefore, accurately predicting the RUL of bearings is essential for predictive maintenance and safe industrial operation.

Degradation feature extraction is a critical phase in predicting the RUL of bearings. Many researchers extract features from multiple domains to enhance prediction accuracy (Tang et al., 2024). Among these, time–frequency features compensate for the limitations of time-domain features, which lack frequency information, and frequency-domain features, which ignore temporal variations. By simultaneously capturing both time and frequency characteristics of the signal, time–frequency features enhance the accuracy and robustness of RUL prediction. VMD (Wang et al., 2023), an advanced signal processing method, has been widely applied for time-frequency domain feature extraction in bearing, as it effectively addresses mode aliasing and end effects. But its performance largely depends on the proper selection of parameters. Inappropriate parameter settings can prevent optimal time–frequency features from being extracted. Therefore, many scholars have carried out research on how to select appropriate VMD parameters. Luan et al. (2024) used the Sparrow search algorithm to improve VMD. Dai et al. (2024) used the Pelican algorithm to optimize VMD parameters. Li et al. (2024b) used a genetic algorithm to optimize VMD parameters . However, the above algorithms still need to be enhanced the performance.

Data-driven methods are widely used in the RUL prediction. Wang et al. (2022) used KPCA to reduced dimensionality of multi-domain features, and employed them as the input of LSTM for RUL prediction. Yao et al. (2023) employed a convolutional denoising stacked autoencoder for feature denoising and dimensionality reduction, which were used as the input of BiLSTM to estimate the bearing RUL. Li et al. (2024a) used CNN to improve feature attention, and further input it into LSTM to improve RUL prediction accuracy. Li and Jian (2024) optimized the hyperparameters of LSTM through TSA to improve the RUL prediction accuracy. However, excessive historical data increases the computational burden of LSTM and its variants, lowering prediction accuracy. In contrast, TCN replaces recurrent structures with convolutional layers, improving efficiency in processing long time series. (Cao et al., 2021). Nevertheless, the parameters of TCN are often set empirically, which cannot guarantee that the model is in an optimal state.

This paper proposes an ICPO-TCN model for RUL prediction. ICPO optimizes VMD parameters to extract time-frequency features, constructs a multi-domain degradation feature set, reduces dimensionality by KPCA, and further tunes TCN's learning rate and kernel size for final RUL prediction.

2. Feature extraction and dimension reduction of bearing degradation

2.1. Time-domain and frequency-domain feature extraction

This paper evaluates time and frequency-domain features by calculating their monotonicity and trend, then derives a weighted comprehensive score, as shown in Equation (1).

$$Score = 0.5Mon + 0.5Tre \tag{1}$$

Based on the comprehensive scores, the top 7 scoring features are selected to best capture the bearing state.

2.2. Time-frequency domain feature extraction

2.2.1. CROWN PORCUPINE OPTIMIZER

The CPO combines both exploration and exploitation mechanisms (Abdel-Basset et al., 2024). The algorithm simulates the defensive behaviors of the crown porcupine in nature to avoid predators, which include four phases: visual, auditory, olfactory, and physical attacks.

In the visual phase, the porcupine raises its quills, prompting the predator to approach or retreat. Approaching enhances convergence by reducing distance, while retreating increases distance to boost global search and escape local optima:

$$\overrightarrow{x_i^{t+1}} = \overrightarrow{x_i^t} + \tau_1 \times \left| 2 \times \tau_2 \times \overrightarrow{x_{CP}^t} - y_i^t \right| \tag{2}$$

Where $\overrightarrow{x_{CP}^t}$ is the optimal solution, y_i^t is the position of the predator at iteration t, τ_1 is a random number generated from a normal distribution, τ_2 is a random value within the range [0,1].

In the auditory phase, The crown porcupine repels the predator by making a sound:

$$\overrightarrow{x_t^{t+1}} = \left(1 - \overrightarrow{U_1}\right) \times \overrightarrow{x_t^t} + \overrightarrow{U_1} \times \left(\boldsymbol{y} + \tau_3 \times \left(\overrightarrow{x_{r_1}^t} - \overrightarrow{x_{r_2}^t}\right)\right) \tag{3}$$

Where r_1 and r_2 are two random integers between $[1, N]$, τ_3 is a random number in the range [0, 1], $\overrightarrow{\mathrm{U}_1}$ is is a randomly generated binary vector in the range [0,1]. $\overrightarrow{y}$ is the position of the predator.

In the olfactory phase, to prevent the predator from approaching, the crown porcupine secretes a foul odor to drive the predator away:

$$\overrightarrow{x_t^{t+1}} = \left(1 - \overrightarrow{U_1}\right) \times \overrightarrow{x_t^t} + \overrightarrow{U_1} \times \left(\overrightarrow{x_{r_1}^t} + S_i^t \times \left(\overrightarrow{x_{r_2}^t} - \overrightarrow{x_{r_3}^t}\right) - \tau_3 \times \boldsymbol{\delta} \times \gamma_t \times S_i^t \tag{4}$$

Where r_3 is a random number in the range $[1, N]$, $\overrightarrow{x_i^t}$ is the position of the ith individual at the tth iteration, $\overrightarrow{\delta}$ is the search direction parameter, γ_t is the defense factor, S_i^t is the diffusion factor.

In the physical attacks phase, when the predator approaches, the crown porcupine uses its quills to attack the predator:

$$\overrightarrow{x_i^{t+1}} = \overrightarrow{x_{CP}^t} + (\alpha\,(1 - \tau_4) + \tau_4) \times \left(\overrightarrow{x_{r_1}^t} + \boldsymbol{\delta} \times \left(\overrightarrow{x_{CP}^t} - \overrightarrow{x_i^t}\right) - \tau_5 \times \boldsymbol{\delta} \times \gamma_t \times F_i^t \tag{5}$$

Where α is the convergence rate factor, τ_4 is in the range [0,1], F_i^t is the average force affecting the i-th predator.

2.2.2. ICPO-VMD

CPO faces slow convergence, low accuracy, and local optima. Therefore, it introduces a mixed sine and cosine strategy, best points set, adaptive T-distribution and nonlinear weighting factor to improve CPO.

First, to enhance population diversity, a best points set is introduced for a more uniform distribution in the search space:

$$R_{i,j} = i \times e^j - \left\lfloor i \times e^j \right\rfloor,\ 1 \le i \le N, 1 \le j \le D \tag{6}$$

Where $R_{i,j}$ represents the ith best points set in the jth dimension, $e^{(\bullet)}$ is exponential operation, $\lfloor \bullet \rfloor$ is round down, N is the total of candidate solutions, D is the total of dimensions. Based on the generation form of the best points set, the population is initialized as follows:

$$\overrightarrow{X_{\imath,1}} = R_{i,j} \times (\boldsymbol{U} - \boldsymbol{L}) + \boldsymbol{L}, 1 \le i \le N, 1 \le j \le D \tag{7}$$

Where $\overrightarrow{X_{i,j}}$ is the population after the initialization of the best points set. And then, in the olfactory phase, the introduction of a nonlinear weight factor ω improves the positive and cosine strategy to enhance the algorithm's convergence rate, as shown in the following equation:

$$\omega = \frac{e^{\frac{t}{T_{max}}}}{e - 1} \tag{8}$$

Where t is the present iteration count, T_{max} is the maximum iteration limit. To prevent getting stuck in local optima and enhance the algorithm's convergence rate, the improved sine-cosine strategy is shown in Equation(10):

$$\overrightarrow{x_t^{t+1}} = \begin{cases} \omega \cdot \overrightarrow{x_t^t} + r_1 \cdot \sin r_2 \cdot \left| r_3 \cdot \overrightarrow{x_{CP}^t} - \overrightarrow{x_t^t} \right|, r_4 < 0.5 \\ \omega \cdot \overrightarrow{x_t^t} + r_1 \cdot \cos r_2 \cdot \left| r_3 \cdot \overrightarrow{x_{CP}^t} - \overrightarrow{x_t^t} \right|, r_4 \geq 0.5 \end{cases} \tag{9}$$

Where $\overrightarrow{x_i^{t+1}}$ is a nonlinear weight factor introduced to improve the solution of sine and cosine strategies, $\overrightarrow{x_i^t}$ is the position of the ith individual at the tth iteration. $\overrightarrow{x_{CP}^t}$ is the best solution found during the search.

Similarly, the optimization strategy with a nonlinear development phase is still prone to getting stuck in local optima. Therefore, the introduction of an adaptive t-distribution strategy to perturb individuals is beneficial for escaping local optimum solutions and accelerating the convergence rate, freedom of movement parameters $t(iter)$:

$$t\left(iter\right) = e^{\frac{iter}{T_{max}}} \tag{10}$$

Where $iter$ is the present iteration count, T_{max} is the maximum iteration limit.

The population location update process is represented as

$$\overrightarrow{X_l^{t+1}} = \begin{cases} \overrightarrow{x_l^{t+1}} + t(\text{ iter }) \cdot \overrightarrow{x_l^{t+1}}, p < 0.5 \\ \overrightarrow{x_l^{t+1}}, p \geq 0.5 \end{cases} \tag{11}$$

Where $\overrightarrow{X_i^{t+1}}$ is the position parameter after adaptive t distribution, $\overrightarrow{x_i^{t+1}}$ is the position parameters at time $t+1, p \in [0, 1]$,and it's a random number.

Therefore, the optimal K and α in VMD can be obtained adaptively by ICPO. Further, K IMF components are decomposed, and appropriate IMF components are selected as time-frequency domain features.

2.3. KPCA features dimension reduction

KPCA employs a kernel function to transform the data into a higher-dimensional feature space, where the nonlinear features of the data become linearly separable. Then, traditional PCA is employed in the transformed feature space to extract the principal components. In this way, KPCA can capture the nonlinear patterns, thus overcoming the linear limitation of traditional PCA.

RBF has a strong non-linear mapping ability and local feature, capable of handling high-dimensional and complex non-linear data, showing superior performance compared to other kernel functions. Therefore, RBF is selected as the kernel function of KPCA to transform the original data into a high-dimensional feature space, as shown in Equation (13):

$$K\left(x_i, x_j\right) = e^{\frac{\left|\left|x_i - x_j\right|\right|}{2\sigma^2}} \tag{12}$$

Where $K\left(\bullet\right)$ is a kernel function, x_i, x_j are input samples' feature vector, σ is a customized parameter to control how the similarity between data is calculated.

And then, the traditional PCA method is applied to calculate the principal components of the features.

Finally, according to the calculated principal components, the most important components are selected to retain, so as to achieve the feature dimensionality reduction.

3. The prediction model of ICPO-TCN

3.1. TCN

As a one-dimensional fully convolutional network, TCN combines the causal convolution structure suitable for time series and the dilated convolution and residual module suitable for historical data storage, so that TCN can effectively capture local and global dependencies. The dilated convolution in the TCN residual block can learn more historical data. See Figure 1 for the structure of the dilated convolution.

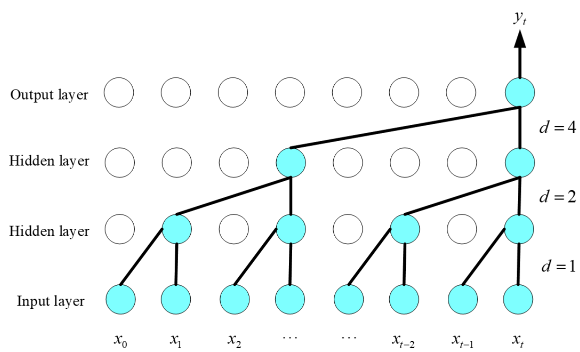


Figure 1: Dilated convolutional structures.

The dilated convolution in the residual block of TCN uses convolution kernels of different sizes to realize feature learning at different scales. For time series x, the output of TCN for feature extraction is shown in Equation (14).

$$G(s) = (x * f_d)(s) = \sum_{i=0}^{k-1} f(i)\, x_{s-di} \tag{13}$$

Where $*$ represents the convolution operation, k represents the convolution kernel's size; d is the expansion factor, $f(i)$ is the ith element of the convolution kernel.

To guarantee that no information is discarded during the training process of TCN, residual connections are introduced based on the causal dilated convolution, whose structure is shown in Figure 2.

3.2. ICPO-TCN

To enhance the prediction performance of TCN, optimized the hyperparameters of TCN by ICPO, which are the number of convolution kernels and the learning rate. The steps are as follows:

Step 1: Initialize the ICPO parameters, including the number of populations, the upper limit of iterations, and the search range of hyperparameters. Meanwhile, the hyperparameters of the TCN are used as the position of the crown porcupine.

Step 2: The fitness value of the target population is calculated according to the objective function, and all the crown porcupine positions are updated, and the updated positions are used as the new hyperparameters of TCN.

Step 3: The final position of the golden boar was taken as the best fitness value X_{best}.

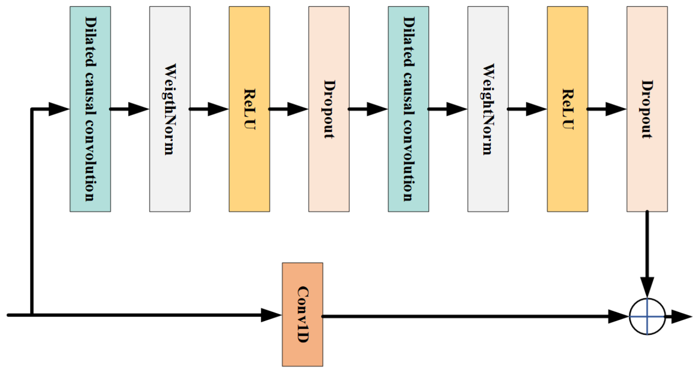


Figure 2: TCN residual structure.

Step 4: Repeat operations 2) and 3) until the iteration limit is reached.

Step 5: Outputs the best hyperparameters for TCN.

3.3. The overall flow of RUL prediction

A RUL prediction method based on ICPO-TCN is proposed. The structure of the proposed method is illustrated in Figure 3.

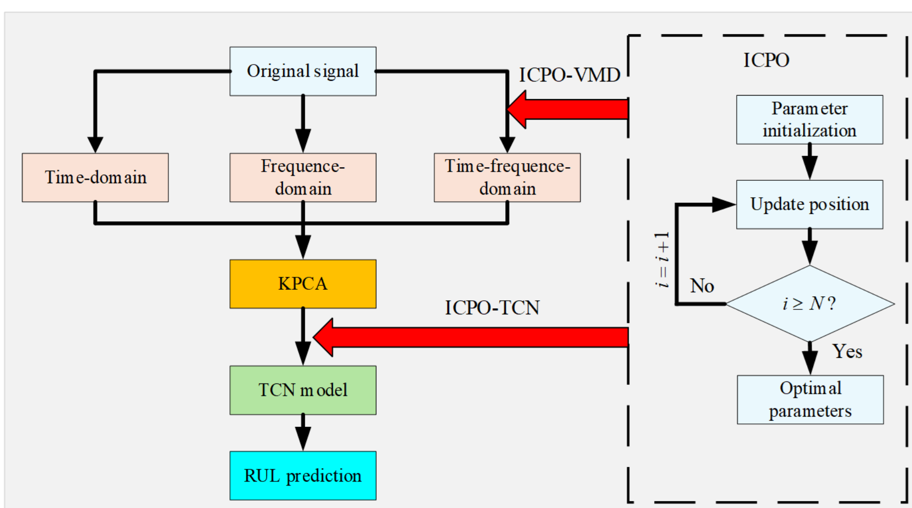


Figure 3: RUL prediction process.

Firstly, the time and frequency domain features extracted from the original signal and the time-frequency domain features extracted from ICPO-VMD are extracted; then, the reduced dimension features are obtained through KPCA; further, the TCN hyperparameters are optimized by ICPO. Finally, the RUL is predicted using the TCN model.

4. Experiment results and analysis

This section uses the IEEE PHM 2012 data (Ma and Mao, 2020) and selects the bearing 1-5 under working condition 1 (4000N and 1800r/min) for experimental verification.

4.1. Feature extraction and analysis

From Figure 4, based on the comprehensive scores, the top 7 time-domain and frequency-domain features are selected as the degradation features of the bearing, including: root mean square, kurtosis, skewness, maximum value, root mean square frequency, skewness frequency, and maximum frequency.

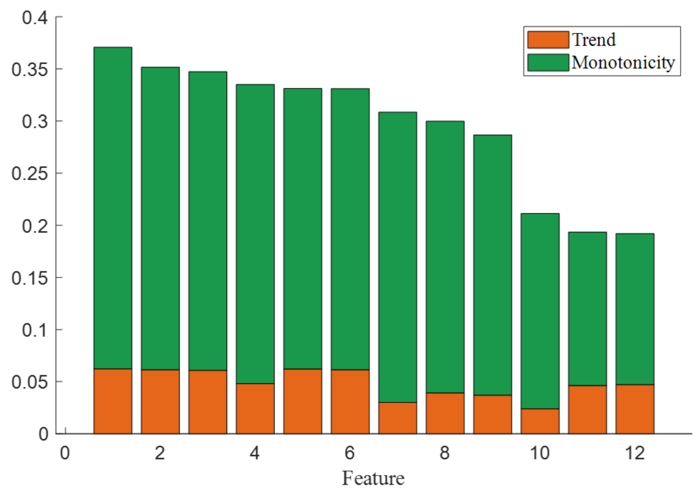


Figure 4: The comprehensive scores of time and frequency-features.

In the time-frequency-domain feature screening, to obtain the best K and α in VMD and obtain the best noise reduction signal. sample entropy, as the objective function. The range of K and α are given as follows: $K \in [[3, 10]$, $\alpha \in [500, 2000]$, ICPO is used to optimize the parameters, and the iterative results, as shown in Figure 5.

Figure 5 shows that the fitness value of ICPO-VMD is smaller than that of CPO-VMD and SSA-VMD, and it converges quickly. The iterative result of ICPO-VMD is $K = 3$ and $\alpha = 1843$. Therefore, IMF1 and IM2 are selected as time-frequency domain features to reconstruct IMF signals of ICPO-VMD, CPO-VMD and SSA-VMD, as shown in Figure 6.

From Table 1, compared with SSA-VMD and CPO-VMD, ICPO-VMD improves the SNR by 2.31% and 0.56%, respectively, and R2 by 3.71% and 2.23%, respectively. Therefore, ICPO-VMD has better denoising effect and fitting degree than the other two methods.

Table 1: Reconstruction performance evaluation of different algorithms to optimize VMD.

Methods	SNR(dB)	R2
SSA-VMD	15.914	0.9343
CPO-VMD	16.189	0.9477
ICPO-VMD	16.281	0.9689

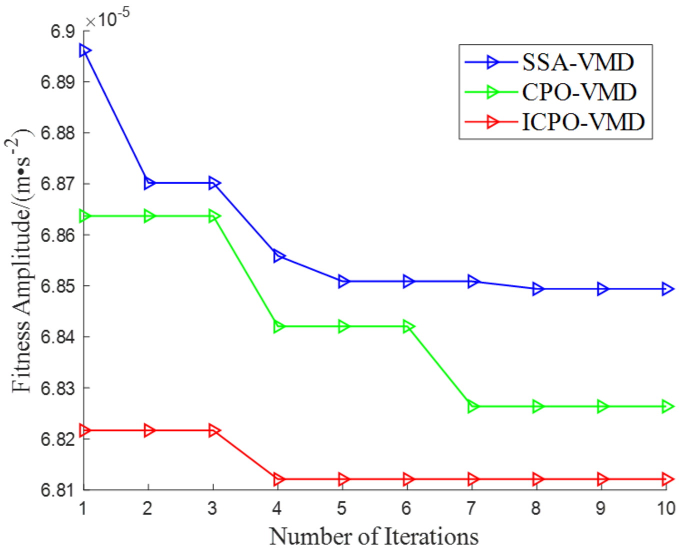


Figure 5: Iterative results of fitness functions of different optimized VMDs.

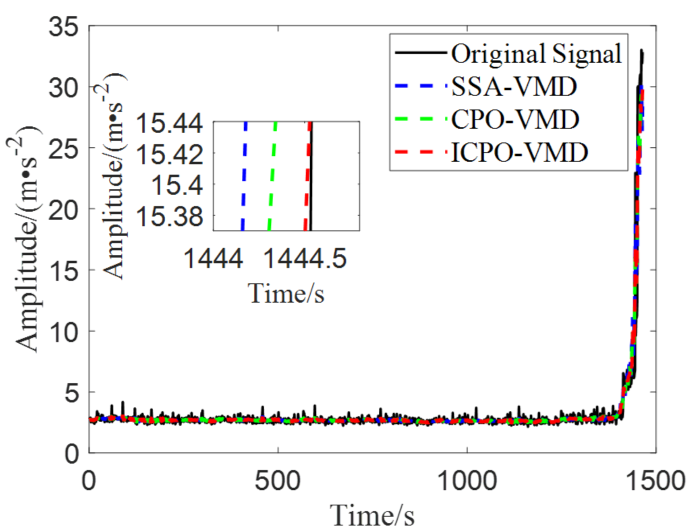


Figure 6: Reconstruction results of VMD optimized by different algorithms.

4.2. Results of feature dimensionality reduction

RBF is used as the kernel function of KPCA to calculate the contribution rate of each feature, are shown in Figure 7. From Figure 7, the contribution rate of the first kernel principal component and the second kernel principal component exceeds 90% after that, so the above features are selected as the kernel principal component variables to be used as the input of the TCN model.

The linear kernel function and the radial basis function are selected to compare the dimensionality reduction performance, and the dimensionality reduction results are shown in Figure 8. Cosine similarity is used to evaluate the performance of dimension reduction as shown in Equation (14).

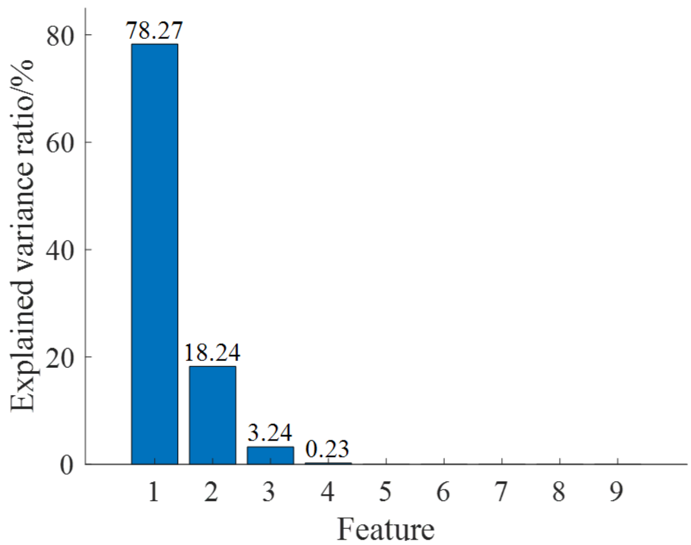
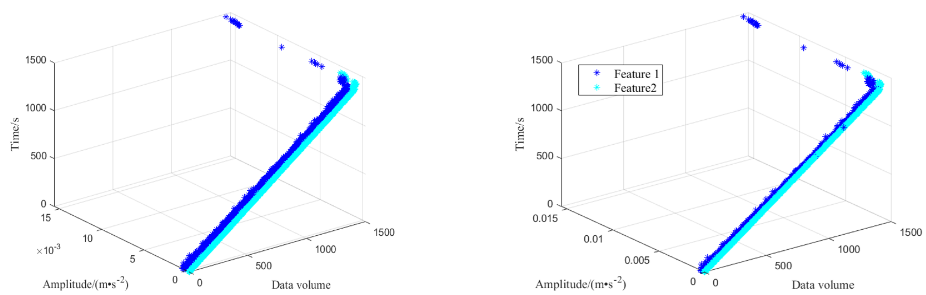


Figure 7: KPCA principal component contribution rate.

The smaller the similarity, the better the effect of dimension reduction.

$$S = \frac{x \bullet y}{|x| \bullet |y|} \tag{14}$$

Where x and y are two feature vectors



(a) RBF-KPCA

(b) Linear Kernel-KPCA

Figure 8: Dimensionality reduction results using different kernel functions.

Table 2: Different kernel's performance evaluation.

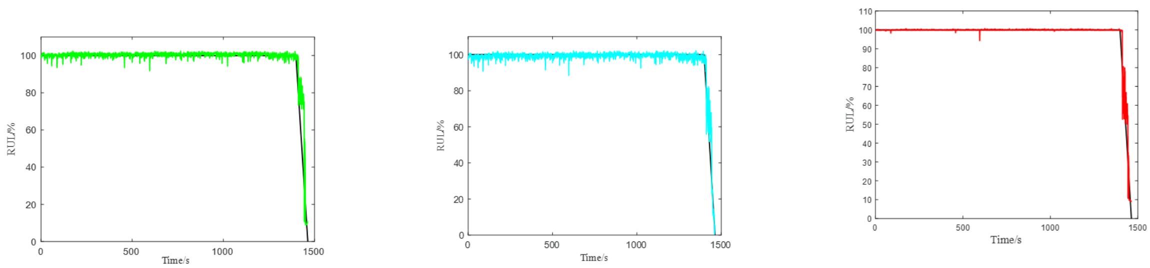
Methods	Cosine similarity
RBF	0.2452
Linear Kernel	0.9854

The similarity of the RBF kernel function is much lower than that of the linear kernel function, so the superiority of the radial basis kernel function is verified.

4.3. Results and analysis of bearing RUL prediction

ICPO optimizes TCN parameters to a kernel size of 3 and a learning rate of 0.0032. Among them, the number of convolution kernel layers of 32, and the bitch size of 32, which is consistent with TCN and LSTM for comparison, with results and performance evaluation indexes in Figure 9 and Table 3.

The analysis in Table 3 shows that compared with LSTM and TCN, the MAE of ICPO-TCN is reduced by 62.13% and 58.25%. The RMSE is reduced by 42.93% and 17.17%. Therefore, ICPO-TCN has high prediction accuracy.



(a) LSTM (b) TCN (c) ICPO-TCN

Figure 9: RUL prediction of different models using bearing1-5.

Table 3: Different models' performance evaluation.

Methods	MAE	RMSE
LSTM	1.6414	4.5805
TCN	1.4888	3.1557
ICPO-TCN	0.6216	2.6138

5. Conclusion

This paper proposes an ICPO-TCN model for the RUL prediction of bearings. The ICPO is used to adaptively determine the optimal parameters of VMD, enabling more effective denoising and extraction of time-frequency features. These features are then combined with extracted time and frequency-domain features to construct a comprehensive degradation representation. The ICPO is further employed to optimize the hyperparameters of the TCN, thereby enhancing its prediction performance. Experimental results demonstrate that, compared to the TCN and LSTM models, the ICPO-TCN outperforms the bearing RUL prediction task and has higher prediction accuracy.

References

M. Abdel-Basset, R. Mohamed, and M. Abouhawwash. Crested porcupine optimizer: A new nature-inspired metaheuristic. *Knowledge-Based Systems*, 284:111257, 2024. doi: 10.1016/j.knosys.2023.111257.

Y. Cao, Y. Ding, M. Jia, et al. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. *Reliability Engineering & System Safety*, 215:107813, 2021. doi: 10.1016/j.ress.2021.107813.

X. Dai, K. Yi, F. Wang, et al. Bearing fault diagnosis based on poa-vmd with gadf-swin transformer transfer learning network. *Measurement*, 238:115328, 2024. doi: 10.1016/j.measurement.2024.115328.

F. Li, Z. Dai, L. Jiang, et al. Prediction of the remaining useful life of bearings through cnn-bi-lstm-based domain adaptation model. *Sensors*, 24(21):6906, 2024a. doi: 10.3390/s24216906.

J. Li, W. Luo, M. Bai, et al. Fault diagnosis of high-speed rolling bearing in the whole life cycle based on improved grey wolf optimizer-least squares support vector machines. *Digital Signal Processing*, 145:104345, 2024b. doi: 10.1016/j.dsp.2023.104345.

L. Li and Q. Jian. Remaining useful life prediction of wind turbine main-bearing based on lstm optimized network. *IEEE Sensors Journal*, 24(13):21143–21156, 2024. doi: 10.1109/JSEN.2024.3402660.

X. Luan, C. Zhong, F. Zhao, et al. Bearing fault damage degree identification method based on ssa-vmd and shannon entropy–exponential entropy decision. *Structural Health Monitoring*, 2024. doi: 10.1177/14759217231219710.

M. Ma and Z. Mao. Deep-convolution-based lstm network for remaining useful life prediction. *IEEE Transactions on Industrial Informatics*, 17(3):1658–1667, 2020. doi: 10.1109/TII.2020.2991796.

Y. Tang, R. Liu, C. Li, et al. Remaining useful life prediction of rolling bearings based on time convolutional network and transformer in parallel. *Measurement Science and Technology*, 35 (12):126102, 2024. doi: 10.1088/1361-6501/ad73ee.

B. Wang, Y. Guo, Z. Zhang, et al. Developing and applying oegoa-vmd algorithm for feature extraction for early fault detection in cryogenic rolling bearing. *Measurement*, 216:112908, 2023. doi: 10.1016/j.measurement.2023.112908.

Y. Wang, J. Zhao, C. Yang, et al. Remaining useful life prediction of rolling bearings based on pearson correlation-kpca multi-feature fusion. *Measurement*, 201:111572, 2022. doi: 10.1016/j.measurement.2022.111572.

X. Yao, J. Zhu, Q. Jiang, et al. Rul prediction method for rolling bearing using convolutional denoising autoencoder and bidirectional lstm. *Measurement Science and Technology*, 35(3):035111, 2023. doi: 10.1088/1361-6501/ad123c.