

Article

Not peer-reviewed version

Remaining Useful Life Prediction of Rolling Bearings Based on Multi-scale Permutation Entropy and SSA-LSTM

[Hongju Wang](#)^{*}, [Xi Zhang](#), Mingming Ren, Tianhao Xu, Chengkai Lu, Zicheng Zhao

Posted Date: 11 September 2023

doi: 10.20944/preprints202309.0676.v1

Keywords: remaining useful life; maximum correlation kurtosis deconvolution; multi-scale permutation entropy; long short-term memory



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Remaining Useful Life Prediction of Rolling Bearings Based on Multi-Scale Permutation Entropy and SSA-LSTM

Hongju Wang *, Xi Zhang, Mingming Ren, Chengkai Lu, Tianhao Xu and Zicheng Zhao

School of Mechanical and Electrical Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China

* Correspondence: qishangyinyin@163.com

Abstract: The performance of bearings plays a pivotal role in determining the dependability and security of rotating machinery. In intricate systems demanding exceptional reliability and safety, the ability to accurately forecast fault occurrences during operation holds profound significance. Such predictions serve as invaluable guides for crafting well-considered reliability strategies and executing maintenance practices aimed at enhancing reliability. In order to ensure the reliability of bearing operation, this article investigates the application of three advanced techniques—Maximum Correlation Kurtosis Deconvolution (MCKD), Multi-Scale Permutation Entropy (MPE), and Long Short-Term Memory (LSTM) recurrent neural networks—for the prediction of the remaining useful life (RUL) of rolling bearings. Each technique's principles, methodologies, and applications are comprehensively reviewed, offering insights into their respective strengths and limitations. Case studies and experimental evaluations are presented to assess their performance in RUL prediction. Findings reveal that MCKD enhances fault signatures, MPE captures complexity, and LSTM excels in modeling temporal patterns. The root mean square error of the prediction results is 0.007. The fusion of these techniques offers a comprehensive approach to RUL prediction, leveraging their unique attributes for more accurate and reliable predictions.

Keywords: remaining useful life; maximum correlation kurtosis deconvolution; multi-scale permutation entropy; long short-term memory

1. Introduction

Rolling bearings are crucial components in various industrial systems, including machinery, automotive, aerospace, and wind turbines[1]. The reliable and efficient functioning of these systems heavily depends on the health and performance of rolling bearings. However, the degradation and failure of rolling bearings can lead to costly downtime, productivity losses, and safety risks. To mitigate these issues, the concept of remaining useful life (RUL) prediction has gained significant attention in recent years. RUL prediction aims to estimate the remaining operational lifespan of rolling bearings, enabling proactive maintenance strategies and optimizing asset management.

Traditional maintenance strategies, such as time-based or reactive maintenance, often result in inefficient resource allocation and unnecessary maintenance activities[2]. By accurately predicting the RUL of rolling bearings, maintenance activities can be planned in advance, leading to reduced downtime, optimized maintenance schedules, and cost savings. RUL prediction also enables condition-based maintenance, where maintenance actions are triggered based on the actual health condition of rolling bearings rather than arbitrary time intervals. This approach enhances reliability, minimizes the risk of catastrophic failures, and improves overall system performance.

Rolling bearings' RUL prediction has traditionally relied on statistical and data-driven methods[3]. However, recent advancements in signal processing, data analytics, and machine learning techniques have provided new opportunities to enhance the accuracy and reliability of RUL prediction. In this article, we investigate the application of three innovative techniques:

maximum correlation kurtosis deconvolution (MCKD)[4], multi-scale permutation entropy (MPE)[5], and long short-term memory (LSTM) recurrent neural network[6].

The objective of this article is to explore the potential of these techniques for rolling bearings' RUL prediction, discuss their advantages and limitations, and highlight their contributions to proactive maintenance strategies and asset management. Through a comprehensive review and analysis of existing literature and research studies, we aim to provide insights into the capabilities and practical implications of MCKD, MPE, and LSTM in the context of rolling bearings' RUL prediction.

By leveraging these advanced techniques, it is expected that rolling bearings' RUL prediction can be improved in terms of accuracy, reliability, and timeliness. This will facilitate the implementation of condition-based maintenance practices, where maintenance activities are performed based on the actual health condition of rolling bearings, optimizing resource allocation and enhancing overall system performance.

The subsequent chapters of this article will delve into the principles, methodologies, applications, and performance evaluation of MCKD, MPE, and LSTM techniques for rolling bearings' RUL prediction. Comparative analysis and discussion of the advantages and limitations of each technique will be presented, along with potential synergies and future directions for enhanced maintenance practices in rolling element systems.

2. Correlation Methods

2.1. Maximum Correlation Kurtosis Deconvolution

(MCKD) is a signal processing technique that aims to enhance the quality and resolution of signals by effectively removing noise and distortion[7]. It is particularly useful in scenarios where the signal of interest is corrupted by additive noise and is convolved with an unknown system impulse response. The mathematical formula can be expressed as follows:

$$x=h * y \quad (1)$$

where: x is the signal convoluted from various signals on the transmission path, y denotes the impulse signal, h represents the response of the y signal after passing the transmission path[8].

The core principle of MCKD is to maximize the correlation kurtosis of the deconvolved signal, which is a statistical measure of the signal's non-Gaussianity. The maximum correlation kurtosis is considered as[9] :

$$O(CK_M(T)) = \frac{\sum_{n=1}^N (\prod_{m=0}^M y(n - mT))^2}{(\sum_{n=1}^N y_n^2)^{M+1}} \quad (2)$$

By maximizing correlation kurtosis, MCKD aims to enhance the signal's sparsity and separate it from the noise and distortions.

$$\max_{f'} CK_M(T) = \max_{f'} \frac{\sum_{i=1}^N (\prod_{m=0}^M y_{i-mT})^2}{(\sum_{i=1}^N y_i^2)^{M+1}} \quad (3)$$

To obtain the maximum value of the relevant kurtosis, it is equivalent to solving the following equation, where the derivative function is 0.

$$\frac{d}{df'_l} CK_M(T) = 0 \quad (l = 1, 2, \dots, L) \quad (4)$$

The final coefficients of the filter can be obtained from equations (1) to (4) and expressed in matrix form:

$$f' = \frac{\|y\|^2}{2\beta^2} (X_0 X_0^T)^{-1} \sum_{r=T}^{mT} X_r \psi_m \quad (5)$$

where:

$$\boldsymbol{\beta} = \begin{bmatrix} y_1 y_{1-T} \cdots y_{1-MT} \\ y_2 y_{2-T} \cdots y_{2-MT} \\ \vdots \\ y_N y_{N-T} \cdots y_{N-MT} \end{bmatrix} \quad (6)$$

$$\mathbf{X}_r = \begin{bmatrix} x_{1-r} & x_{2-r} & x_{3-r} & \cdots & x_{N-r} \\ 0 & x_{1-r} & x_{2-r} & \cdots & x_{N-1-r} \\ 0 & 0 & x_{1-r} & \cdots & x_{N-2-r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x_{N-L-r+1} \end{bmatrix} (r = [T, \dots, mT]) \quad (7)$$

$$\boldsymbol{\psi}_m = \begin{bmatrix} y_{1-mT}^{-1} (y_1^2 y_{1-T}^2 \cdots y_{1-MT}^2) \\ y_{2-mT}^{-1} (y_2^2 y_{2-T}^2 \cdots y_{1-MT}^2) \\ \vdots \\ y_{N-mT}^{-1} (y_N^2 y_{N-T}^2 \cdots y_{N-MT}^2) \end{bmatrix} \quad (8)$$

In conclusion, the implementation process of the MCKD algorithm can be formulated as follows:

- (1) Initialize parameters such as the deconvolution period T , the number of shifts M , and the length of the filter L .
- (2) Calculate the X_0^T and $(X_0 X_0^T)^{-1}$ of the input signal x .
- (3) Compute the filtered output signal y .
- (4) Calculate $\boldsymbol{\psi}_m$ and $\boldsymbol{\beta}$ based on y .
- (5) Update the coefficients of the filter f .

If the kurtosis difference value $\Delta CK_M(T)$ between the signals before and after filtering is smaller than the threshold, end the iteration. Otherwise, repeat steps 3 to 5.

2.2. Multi-Scale Permutation Entropy

Multi-Scale Permutation Entropy (MPE) is a powerful tool used for analyzing the complexity and irregularity of time series data[10]. It is particularly useful in the field of signal processing and analysis, where it can provide valuable insights into the underlying dynamics and patterns present in the data.

The core concept of MPE lies in the analysis of the ordinal patterns or permutations that occur within a time series at different scales or resolutions. Ordinal patterns capture the relative order of the data points within a sliding window of fixed length. By examining the frequency and distribution of these ordinal patterns, MPE quantifies the complexity and information content of the time series.

The MPE methodology involves the following steps:

Signal Segmentation: The time series data is divided into non-overlapping segments or windows of fixed length. The length of the window determines the scale or resolution at which the analysis is performed.

$$y_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_i, j = 1, 2, \dots, [N/s] \quad (9)$$

where: s denotes the scale factor; $[N/s]$ represents the length of new time series.

Ordinal Pattern Generation: Within each segment, the ordinal patterns are generated by assigning a rank to each data point based on its relative position compared to other data points within the window. For example, the smallest data point is assigned rank 1, the second smallest rank 2, and so on[11].

Permutation Encoding: Each ordinal pattern is encoded into a permutation symbol, representing the order of ranks. For instance, if the ranks within a window are 3, 1, 2, the corresponding permutation symbol would be 312.

Permutation Frequency Analysis: The frequency of occurrence of each permutation symbol is calculated across all the segments at the given scale. This information reflects the distribution of ordinal patterns and provides insights into the complexity and regularity of the time series.

Entropy Calculation: The entropy is computed based on the probabilities of the permutation symbols. Entropy measures the amount of uncertainty or information content in the time series. Higher entropy values indicate higher complexity and irregularity, while lower entropy values suggest more regular and predictable patterns.

$$H_p(m) = -\sum_{r=1}^R P_r \ln P_r \quad (10)$$

The value of H_p reaches the highest when $P_r = 1/m!$. For convenience, normalization is generally accomplished[12].

$$H_p = H_p(m)/\ln(m!) \quad (11)$$

In the context of rolling bearings' RUL prediction, MPE can be utilized to analyze the vibration signals obtained from bearings. Vibration signals contain valuable information about the health condition and fault characteristics of the bearings. By applying MPE, the complexity and irregularity of these signals can be quantified, providing useful features for fault diagnosis and RUL prediction.

MPE offers several advantages in RUL prediction analysis. First, it is a non-parametric technique, meaning it does not assume any specific underlying distribution of the data. This flexibility makes it suitable for analyzing complex and non-linear dynamics commonly observed in rolling element systems.

Second, MPE is capable of capturing both short-term and long-term temporal dependencies in the data[13]. By analyzing the ordinal patterns at different scales, MPE can reveal the presence of localized or global patterns, offering a comprehensive understanding of the bearing's health condition.

Furthermore, MPE can capture subtle changes in the complexity of the vibration signals, allowing for early detection of fault initiation and progression. This early detection can lead to timely maintenance actions and improved RUL prediction accuracy.

By quantifying the complexity and irregularity of vibration signals, MPE-based features can effectively discriminate between different fault conditions and provide valuable information for estimating the remaining operational lifespan of the bearings.

2.3. Long Short-Term Memory Recurrent Neural Network

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture specifically designed to handle the challenges of learning and remembering long-term dependencies in sequential data. Unlike traditional RNNs[14], which suffer from the "vanishing gradient" problem and struggle to capture long-term dependencies, LSTMs are equipped with memory cells and gating mechanisms that enable them to selectively retain and update information over time[15].

The key components of an LSTM network include:

Memory Cell \hat{c}_t : The memory cell serves as the main building block of an LSTM. It maintains an internal state that can be updated or preserved using gating mechanisms. The memory cell enables the LSTM to learn and store information over long sequences[16].

$$\hat{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (12)$$

where: W_{xc} represents the connection weight between the input layer and the hidden layer at time t . W_{hc} denotes the connection weight between the hidden layers at time $t-1$ and t [17]. b_c and h_{t-1} respectively represent the biases of the input nodes and the previous time step's output.

Input Gate i_t : The input gate determines the amount of new information to be stored in the memory cell at each time step. It takes input from the current time step and the previous hidden state and applies a sigmoid activation function to generate an input gate activation value[18].

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (13)$$

where: W_{xi} represents the connection weight between the input layer and the hidden layer at time t . W_{hi} denotes the connection weight between the hidden layers at time $t-1$ and t . b_i and h_{t-1}

respectively represent the biases of the input gate and the previous time step's output. σ denotes the sigmoid activation function.

Forget Gate f_t : The forget gate determines the extent to which previous information should be forgotten or preserved in the memory cell. It takes input from the current time step and the previous hidden state and applies a sigmoid activation function to generate a forget gate activation value[19].

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (14)$$

Output Gate: The output gate regulates the amount of information to be output from the memory cell to the next time step. It takes input from the current time step and the previous hidden state and applies a sigmoid activation function to generate an output gate activation value.

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (15)$$

Hidden State: The hidden state carries information from the memory cell and previous hidden state to the next time step. It is computed by applying a tanh activation function to the current input and the memory cell state, and then scaling it by the output gate activation value.

The structure of LSTM is shown in Figure 1. The use of these gates and memory cells in LSTMs allows the network to selectively update, forget, and output information at each time step, facilitating the learning and retention of long-term dependencies in sequential data.

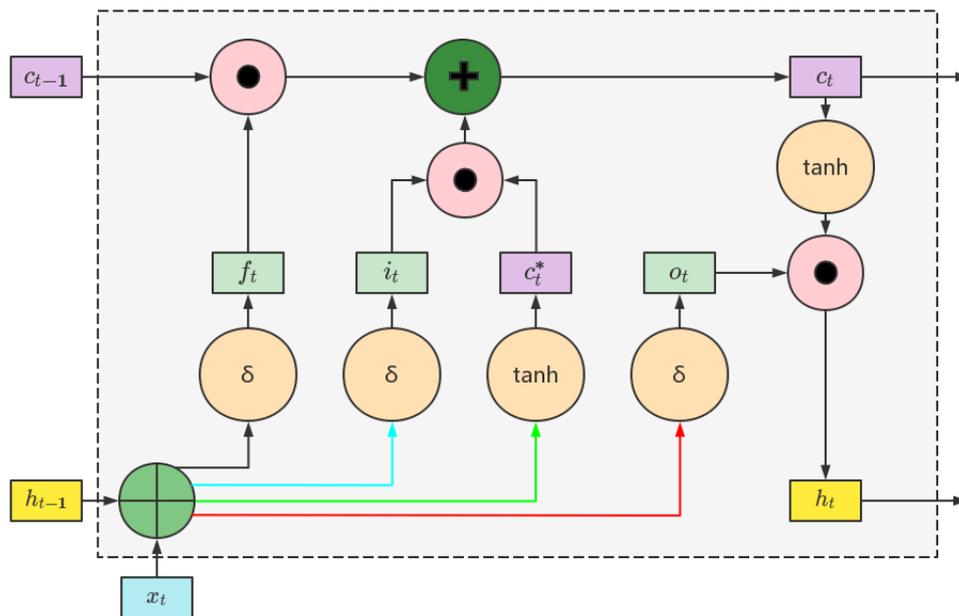


Figure 1. The structure of LSTM.

LSTM has gained significant attention in the field of rolling bearings' RUL prediction due to its ability to model complex temporal dependencies and effectively handle time-series data. By processing the vibration signals obtained from rolling bearings, LSTM networks can learn the underlying patterns and characteristics indicative of bearing health conditions.

In the context of rolling bearings' RUL prediction, LSTM networks can be utilized after data preprocessing and feature extraction:

LSTM Network Architecture: The LSTM network is constructed with input, hidden, and output layers. The input layer receives the sequence of MPE, which is fed into the LSTM layer. The hidden layer contains the LSTM units, responsible for processing and capturing the temporal dependencies in the data. The output layer generates predictions based on the learned patterns and features extracted by the LSTM layer.

Training and Optimization: The LSTM network is trained using a labeled dataset of vibration signals and corresponding RUL values[20]. The network learns to minimize the difference between

its predicted RUL values and the actual RUL values. Training involves forward propagation, backpropagation through time, and optimization algorithms such as gradient descent to update the network's weights and biases.

RUL Prediction: Once the LSTM network is trained, it can be used to predict the remaining useful life of rolling bearings[21]. Given a new sequence of vibration data, the LSTM network processes the sequence through the trained network and generates a predicted RUL value based on the learned temporal patterns and dependencies.

The flowchart of RUL prediction is shown as Figure 2.

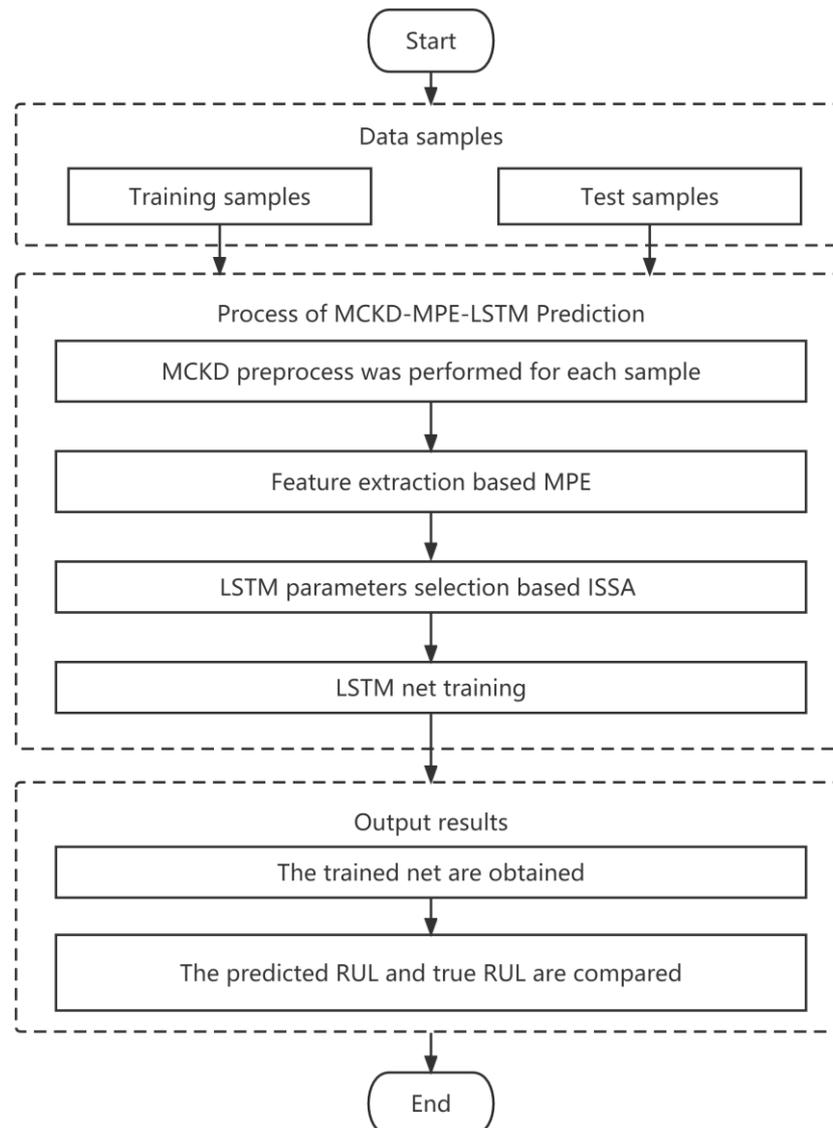


Figure 2. Flowchart of the prediction process.

3. Experiments and Results

3.1. Experimental Platform

To facilitate a comprehensive analysis of the acquired findings, the experimental dataset featuring LDK UER204 rolling element bearings from the XJTU-SY bearing was employed.

Figure 3 provides a visual depiction of the rolling bearings testbed, a sophisticated assembly comprising essential components such as an alternating current (AC) motor, motor speed controller, support shaft, heavy-duty rolling bearings serving as support bearings, and a hydraulic loading system, among others[22]. This experimental platform stands equipped to execute accelerated

degradation tests on bearings across varied operational scenarios, simultaneously capturing comprehensive run-to-failure data. The parameters for data acquisition were configured with a sampling frequency of 25.6 kHz and a sampling interval of 1 minute, respectively. This arrangement culminated in a total of 32,768 individual samples being recorded. Subsequently, the analysis focused on the horizontal vibration signals originating from the dataset designated as "bearing 3_1." The entire life cycle bearing vibration signal is shown in Figure 4.

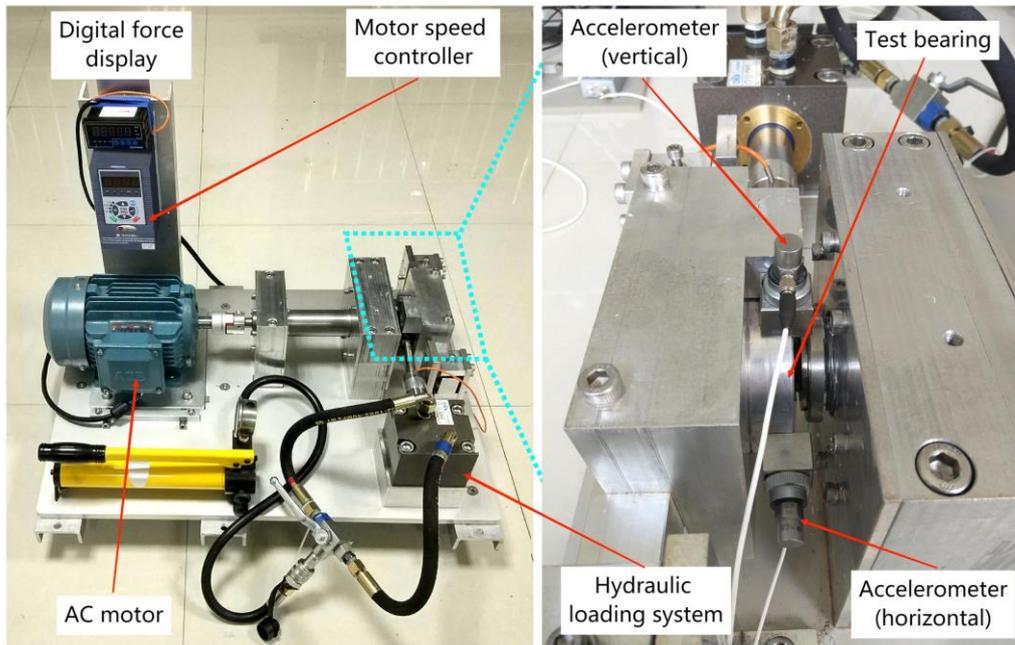


Figure 3. Bearing accelerated life test bed.

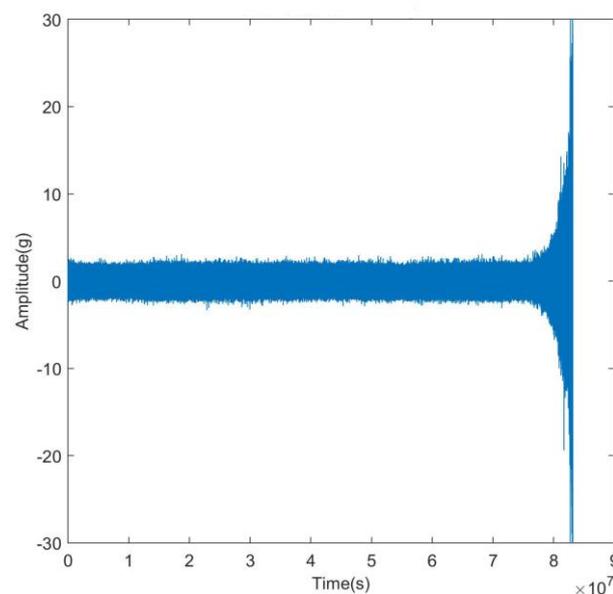


Figure 4. Horizontal vibration signal.

3.2. Results and Discussion

Due to the inevitable presence of noise in the collected raw vibration signals, the fault frequencies of the bearings are embedded within other spectral components. The envelope spectrum of the raw signal is shown in Figure 5. The envelope spectrum reaches its peak around 10Hz, which is not the characteristic fault frequency of a rolling bearing.

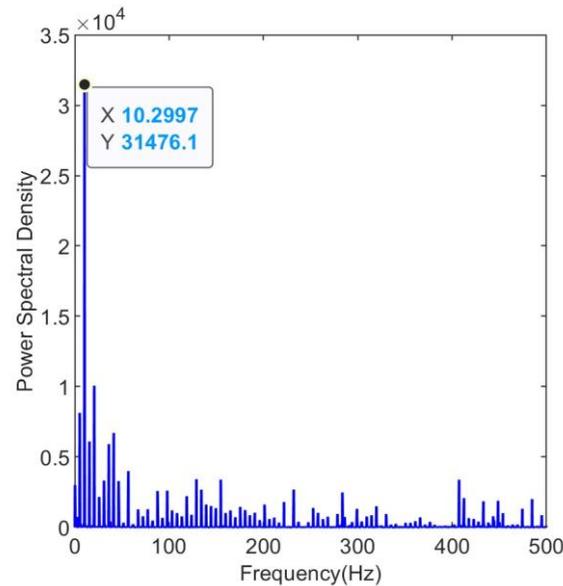


Figure 5. Envelope spectrum of the raw signal.

Perform maximum kurtosis deconvolution on the raw vibration signal. Set the filter length to 19. The maximum number of iterations is 30. The deconvolution period is set to 609, which is the ratio between the sampling frequency of 25.6kHz and the fault frequency of 42Hz. The shift number M is set to 3, indicating that three consecutive impacts are considered as a single valid impact.

The filtered envelope spectrum is shown in Figure 6. The envelope spectrum reaches its peak at 41.73Hz, which is very close to the fault frequency of 42Hz. The harmonics are still clearly visible, demonstrating the effectiveness of the maximum kurtosis deconvolution process.

After applying the maximum kurtosis deconvolution for data preprocessing, the multiscale permutation entropy is utilized for feature extraction. The purpose of this is to quantify the degradation information during the operational life of bearings.

Four parameters must be established before MPE can be used[23]: encapsulation dimension m , time series length N , scale factor s , and time delay τ . According to the reference literature, we set the encapsulation dimension m to 5. The time series length N is 3000, which can meet the criterion of $N \geq 5m$!. The time delay $\tau = 1$ here since the time delay τ has no significant impact on the outcome[24]. The scale factor s will influence the subsequent feature dimension. When the feature dimension is too small, it may not meet the requirements for RUL prediction. On the other hand, too many features can lead to the curse of dimensionality. Through multiple experiments, we have found that setting the scale factor s to 6 yields satisfactory results. Therefore, the scale factor s is set to 6 to obtain permutation entropy at each scales.

The time evolution curve of permutation entropy for six scales with bearing degradation is shown in Figure 7. It can be observed that, except for the scale factor $s=1$, the permutation entropy values of the other five scales remained relatively stable in the early stages, close to a value of 1. As the early stages of bearing faults emerge, the permutation entropy values gradually decrease. As the bearing faults become more severe, the permutation entropy values decrease to a new steady state.

The reason why the permutation entropy values decrease as the bearing degrades is that the magnitude of permutation entropy represents the level of disorder in the information. When the bearing is in a healthy state, the vibration signal tends to be more random. As the bearing gradually develops faults, periodic impacts occur due to cyclic collisions at the faulty region. These periodic impacts introduce a more regular pattern in the vibration signal compared to random vibrations. As a result, the permutation entropy values decrease with the bearing degrades.

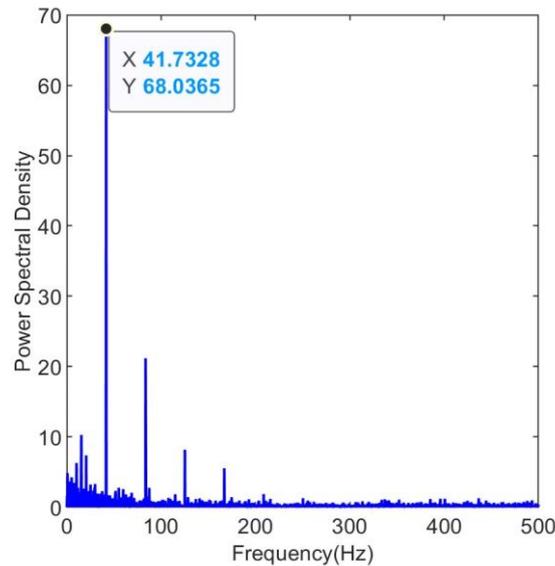


Figure 6. Envelope spectrum of the filtered signal.

Furthermore, the permutation entropy with a scale factor $s=1$ cannot effectively capture the degradation process of the bearing[25], while the permutation entropy with other scale factors can better accomplish this representation. This is because vibration signals can exhibit rapid variations within continuous time scales, while showing more stable trends over longer time scales. Single-scale permutation entropy might not capture these multi-scale characteristic changes, as it focuses solely on patterns within continuous time scales. Permutation entropy at other scales achieves a coarser representation of the signal across different time scales, mitigating the impact of short-term fluctuations in the signal. This aids in extracting the overall trends present in the signal, resulting in a more comprehensive and accurate feature extraction.

The other permutation entropy values apart from whose scale factor $s=1$ are used as feature vectors to input into the subsequent Long Short-Term Memory (LSTM) neural network model[26]. To make more efficient use of computational resources while improving the accuracy of the prediction model, we have truncated the samples to exclude the initial stable operating period. This allows the model to focus more on the later degradation phase, where the critical information for prediction lies.

To employ LSTM for remaining useful life prediction, several hyperparameters need to be set in advance, which include the number of neurons in the hidden layer, the maximum number of epochs, the initial learning rate[27]. Different parameter settings can indeed have a significant impact on the final results obtained. It's crucial to carefully tune these parameters to ensure the best performance and meaningful predictions in specific use case. Experimenting with various parameter combinations and evaluating their effects on the model's performance is an essential step in optimizing the prediction accuracy.

Applying the improved sparrow search algorithm mentioned in the reference literature[28] for parameter optimization is a valuable approach. This algorithm can assist in finding optimal or near-optimal parameter settings by simulating the search behavior of sparrows and their interactions within an optimization space. It's essential to define the objective function and then use the algorithm to iteratively search for parameter combinations that yield the best results. Selecting the root mean square error (RMSE) as the objective function is a common and appropriate choice. RMSE is a widely used metric in machine learning and prediction tasks to quantify the difference between predicted and actual values, making it suitable for evaluating the performance of remaining useful life prediction model[29]. The goal of parameter optimization process would be to minimize the RMSE to achieve accurate and reliable predictions.

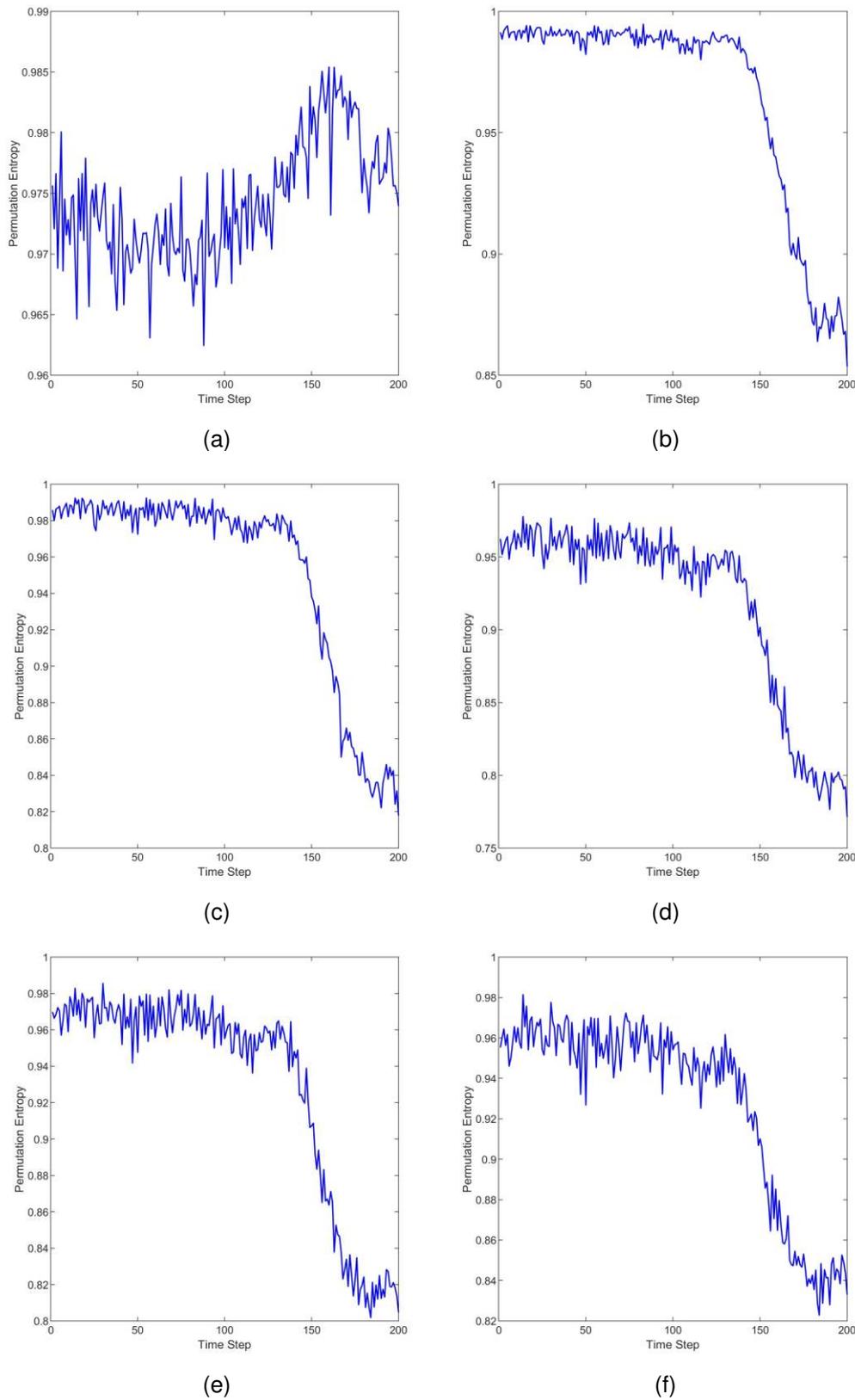


Figure 7. Time evolution curve of permutation entropy. (a) $s=1$ (b) $s=2$ (c) $s=3$ (d) $s=4$ (e) $s=5$ (f) $s=6$.

From Figure 8, it can be seen that the objective function value rapidly decreases during the first to the second iteration and converges by the fourth iteration. The parameter combination obtained through the Sparrow Search Optimization algorithm is as follows: 218, 203, 0.01, which correspond to the number of hidden units, maximum training epochs, and initial learning rate, respectively.

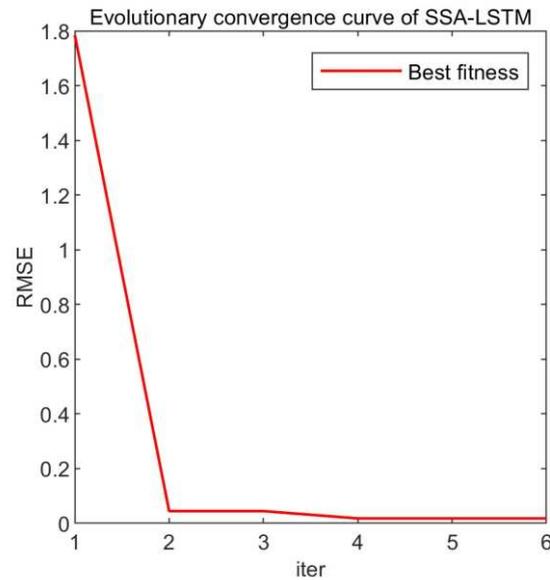


Figure 8. Evolutionary convergence curve of SSA-LSTM.

The remaining useful life prediction results obtained by inputting the feature matrix composed of multiscale permutation entropy into the Long Short-Term Memory neural network are shown in Figure 9. The root mean square error of the prediction results is 0.007. The results indicate that there is a slight drift between the predicted remaining useful life and the actual remaining useful life at the beginning and end, showing a minor endpoint effect. Around time steps 60 and 80, the predicted remaining useful life values are lower than the actual remaining useful life values. The rest of the prediction results are very close to the true values, which validates the effectiveness of the proposed model.

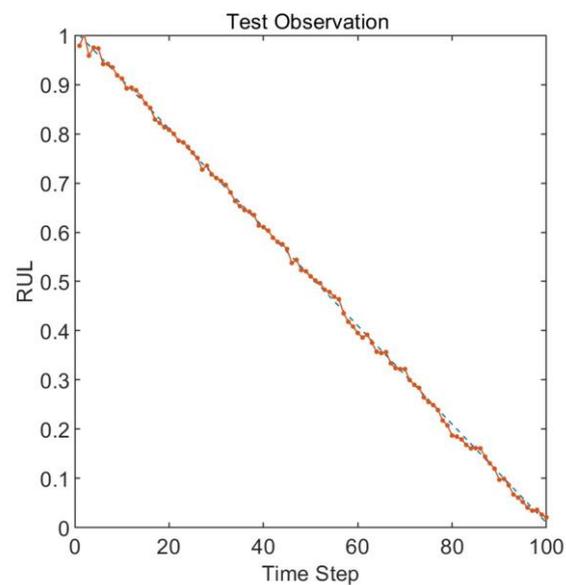


Figure 9. RUL prediction results with optimized parameters.

The prediction results obtained using the default initial values of 50, 50, and 0.1 for the Long Short-Term Memory neural network are shown in Figure 10. The root mean square error of the prediction results is 0.017. The root mean square error decreased by 58.8% after parameter optimization compared to the default settings. Compared to the optimized model, the results obtained with the default settings show a more pronounced endpoint effect, with a higher degree of deviation from the actual values at time steps 40, 60, and 80.

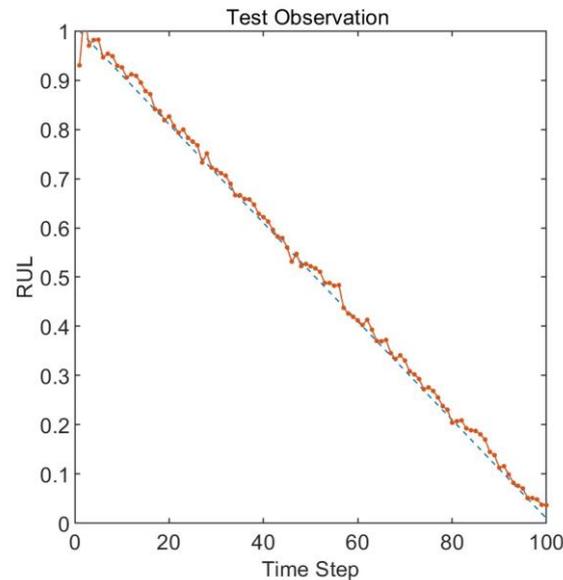


Figure 10. RUL prediction results with default parameters.

4. Conclusions

In this research, we explored the application of three different techniques for the prediction of remaining useful life of rolling bearings: Maximum Correlation Kurtosis Deconvolution (MCKD), Multi-Scale Permutation Entropy (MPE), and Long Short-Term Memory (LSTM) recurrent neural network. We presented a comprehensive review of each technique, including their underlying principles, methodologies, and applications. The main conclusions can be summarized as follows:

The MCKD technique utilizes kurtosis-based deconvolution to enhance fault signatures and improve fault detection and severity estimation. MPE quantifies the complexity and irregularity of vibration signals at multiple scales, providing valuable insights into fault conditions. LSTM networks excel at capturing long-term dependencies and modeling complex temporal dynamics in sequential data, making them suitable for RUL prediction. The integration of MCKD, MPE, and LSTM techniques offers a comprehensive approach to RUL prediction, harnessing the strengths of each technique to provide more accurate and reliable predictions.

The findings from this research pave the way for future advancements in rolling element maintenance and asset management. Further research can focus on exploring advanced fusion techniques, incorporating additional sensor modalities, and investigating novel feature selection methods. Additionally, the development of user-friendly software tools and frameworks can facilitate the practical implementation of these techniques in industrial settings.

In conclusion, the integration of MCKD, MPE, and LSTM techniques for rolling bearings' RUL prediction holds great promise in improving maintenance practices and asset management strategies. The comprehensive analysis and accurate prediction of RUL enable timely maintenance interventions, optimizing the performance, reliability, and longevity of rolling bearings. By embracing these techniques, industries can achieve significant cost savings, minimize downtime, and enhance overall operational efficiency.

Author Contributions: X.Z. designed the main idea and revised the manuscript. H.W. designed the main methods and experiments, and wrote the paper. M.R., T.X., and C.L. participated in the design of the algorithm and carried out the relevant simulations. C.L. and Z.Z. participated in the data acquisition and field inspection. M.R. also helped to revise the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (Grant No. 52121003).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: <http://biaowang.tech/xjtu-sy-bearing-datasets>, accessed on 6 September 2023.

Acknowledgments: The study was approved by the China University of Mining and Technology (Beijing). The authors would like to thank the editor and the anonymous reviewers for their constructive comments and suggestions. The authors also would like to thank Essentialink (www.essentialink.cn) for the help in academic language standardization.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. R. Shi, B. Wang, Z. Wang, J. Liu, X. Feng, and L. Dong, "Research on Fault Diagnosis of Rolling Bearings Based on Variational Mode Decomposition Improved by the Niche Genetic Algorithm," *Entropy*, vol. 24, no. 6, 2022.
2. X. Zhang, J. Kang, L. Hao, L. Cai, and J. Zhao, "Bearing fault diagnosis and degradation analysis based on improved empirical mode decomposition and maximum correlated kurtosis deconvolution," *Journal of Vibroengineering*, vol. 17, no. 1, pp. 243-260, 2015.
3. W. Heng, N. Guangxian, C. Jinhai, and Q. Jiangming, "Research on Rolling Bearing State Health Monitoring and Life Prediction Based on PCA and Internet of Things with Multi-sensor," *Measurement*, vol. 157, no. 4, p. 107657, 2020.
4. J. Kang, X. Zhang, H. Teng, and J. Zhao, "Application of maximum correlated Kurtosis deconvolution on bearing fault detection and degradation analysis," 2014.
5. Morabito *et al.*, "Entropy, Vol. 14, Pages 1186-1202: Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG," 2012.
6. A. Akandeh and F. M. Salem, "Simplified Long Short-term Memory Recurrent Neural Networks: part III," 2017.
7. Y. Zhang, Y. Lv, and M. Ge, "Time-frequency analysis via complementary ensemble adaptive local iterative filtering and enhanced maximum correlation kurtosis deconvolution for wind turbine fault diagnosis," *Energy Reports*, 2021.
8. Z. Li and L. I. Bao-Wan, "Roller Bearing Fault Diagnosis Method Based on Iterative Filtering and Maximum Correlation Kurtosis Deconvolution," *Modular Machine Tool & Automatic Manufacturing Technique*, 2019.
9. W. Sun, YuqingCao, XinchengChen, BinqiangFeng, WeiChen, Leiqing, "A two-stage method for bearing fault detection using graph similarity evaluation," *Measurement*, vol. 165, no. 1, 2020.
10. C. H. A, T. W. B, R. G. C, Z. J. B, R. M. B, and H. Q. D, "Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm – Extreme learning machine - ScienceDirect," *Measurement*, 2020.
11. G. Tang, X. Wang, and Y. He, "A Novel Method of Fault Diagnosis for Rolling Bearing Based on Dual Tree Complex Wavelet Packet Transform and Improved Multiscale Permutation Entropy," *Mathematical Problems in Engineering*, 2016, (2016-5-9), vol. 2016, no. pt.5, pp. 1-13, 2016.
12. J. Zheng, J. Cheng, and Y. Yang, "Multiscale Permutation Entropy Based Rolling Bearing Fault Diagnosis," *Shock and Vibration*, 2014, (2014-3-2), vol. 2014, no. pt.1, pp. 1-8, 2014.
13. Z. Shenghan, Q. Silin, C. Wenbing, X. Yiyong, and C. Yang, "A Novel Bearing Multi-Fault Diagnosis Approach Based on Weighted Permutation Entropy and an Improved SVM Ensemble Classifier," *Sensors*, vol. 18, no. 6, p. 1934, 2018.
14. M. A. Zaytar and C. E. Amrani, "Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks," *International Journal of Computer Applications*, vol. 143, no. 11, pp. 7-11, 2016.
15. X. G. Zhou, Jing, "Tool remaining useful life prediction method based on LSTM under variable working conditions," *The International Journal of Advanced Manufacturing Technology*, vol. 104, no. 9a12, 2019.
16. T. Song, C. Liu, R. Wu, Y. Jin, and D. Jiang, "A hierarchical scheme for remaining useful life prediction with long short-term memory networks," *Neurocomputing*, no. May 28, p. 487, 2022.
17. W. Mao, J. He, J. Tang, and Y. Li, "Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network," *Advances in Mechanical Engineering*, vol. 10, no. 12, 2018.
18. A. Elsheikh, S. Yacout, and M. S. Ouali, "Bidirectional handshaking LSTM for remaining useful life prediction," *Neurocomputing*, vol. 323, no. JAN.5, pp. 148-156, 2019.

19. C. Chen, J. Shi, N. Lu, Z. H. Zhu, and B. Jiang, "Data-driven predictive maintenance strategy considering the uncertainty in remaining useful life prediction," *Neurocomputing*, no. Jul.14, p. 494, 2022.
20. T. Xu, X. U. Wei-Xiao, T. Ji-Wen, and W. Yan-Song, "Prediction for remaining useful life of rolling bearings based on Long Short-Term Memory," *Journal of Machine Design*, 2019.
21. J. Morgenroth, K. Kalenchuk, L. Moreau-Verlaan, M. A. Perras, and U. T. Khan, "A novel long-short term memory network approach for stress model updating for excavations in high stress environments," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 17, no. 1, pp. 196-216, 2023.
22. B. Wang, Y. Lei, N. Li, and T. Yan, "Deep separable convolutional network for remaining useful life prediction of machinery," *Mechanical Systems and Signal Processing*, vol. 134, 2019.
23. Wu *et al.*, "Entropy, Vol. 14, Pages 1343-1356: Bearing Fault Diagnosis Based on Multiscale Permutation Entropy and Support Vector Machine," 2012.
24. Y. Zhang, Y. Lv, and M. Ge, "A Rolling Bearing Fault Classification Scheme Based on k-Optimized Adaptive Local Iterative Filtering and Improved Multiscale Permutation Entropy," *Entropy (Basel, Switzerland)*, vol. 23, no. 2, 2021.
25. W. Zhang and J. Zhou, "Fault Diagnosis for Rolling Element Bearings Based on Feature Space Reconstruction and Multiscale Permutation Entropy," *Entropy*, vol. 21, no. 5, p. 519, 2019.
26. P. X. Zhu, Min, "A Joint Long Short-Term Memory and AdaBoost regression approach with application to remaining useful life estimation," *Measurement*, vol. 170, no. 1, 2021.
27. S. Xiang, Y. Qin, C. Zhu, Y. Wang, and H. Chen, "Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction," *Engineering Applications of Artificial Intelligence*, vol. 91, pp. 103587-, 2020.
28. X. Zhang, H. Wang, X. Li, S. Gao, K. Guo, and Y. Wei, "Fault Diagnosis of Mine Ventilator Bearing Based on Improved Variational Mode Decomposition and Density Peak Clustering," *Machines*, vol. 11, no. 1, 2022.
29. K. Abdelli, H. Griesser, and S. Pachnicke, "A Hybrid CNN-LSTM Approach for Laser Remaining Useful Life Prediction," 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.