



Article

# An End-to-End Adaptive Method for Remaining Useful Life Prediction of Rolling Bearings Using Time–Frequency Image Features

Liang Chen <sup>1,2</sup>, Hao Wang <sup>1,\*</sup>, Linshu Meng <sup>1,2</sup>, Zhenzhen Xu <sup>1</sup>, Lin Xue <sup>1</sup> and Mingfa Ren <sup>1</sup>

- <sup>1</sup> Department of Mechatronic Engineering, College of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China; zhihaofan@mail.dlut.edu.cn (L.C.); lizhaoxiang@mail.dlut.edu.cn (L.M.); whtot@mail.dlut.edu.cn (Z.X.); linxue@dlut.edu.cn (L.X.); renmf@dlut.edu.cn (M.R.)  
<sup>2</sup> Shenyang Aircraft Design & Research Institute, The Aviation Industry Corporation of China, Ltd., Shenyang 110035, China  
\* Correspondence: zhaohengyi2022@mail.dlut.edu.cn; Tel.: +86-182-9244-7341

**Abstract:** The deep learning model has attracted widespread attention in the field of rolling bearing remaining useful life (RUL) prediction due to its advantages of less reliance on prior knowledge, high accuracy, and strong generalization. However, a large number of prediction models use very complicated artificial feature extraction and selection methods to build the original input features of the deep learning model and health indicator. These approaches do not fully exploit the capabilities of deep learning models as they continue to heavily rely on prior knowledge. The accuracy of their predictions largely hinges on the quality of the input features, and the generalization of manually crafted features remains uncertain. To address these challenges, in this paper, an end-to-end prediction model for the remaining useful life of rolling bearings is proposed, which is divided into three modules. First, a short-term Fourier transform module is incorporated into the model to automatically obtain the time–frequency information of the signal. Then, the convolutional next (ConvNext) module, which is a simple and efficient pure convolutional neural network, is utilized to extract features from the spectrogram. Finally, we capture the short-term dependence and long-term dependence by two parallel channels Transformer and self-attention convolutional long short-term memory (SA-ConvLSTM), and the self-attention mechanism is employed for the adaptive prediction of the bearing’s remaining useful life. Through integration with artificial intelligence, this method proposes a high-performance solution for predicting the remaining useful life of bearings. It has minimal reliance on manual labor, stronger fitting capabilities, and can be widely used for predicting the remaining useful life of bearings.

**Keywords:** rolling bearings; remaining useful life prediction; short-term Fourier transform; model fusion



**Citation:** Chen, L.; Wang, H.; Meng, L.; Xu, Z.; Xue, L.; Ren, M. An End-to-End Adaptive Method for Remaining Useful Life Prediction of Rolling Bearings Using Time–Frequency Image Features. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 2892–2912. <https://doi.org/10.3390/make6040138>

Academic Editors: Affan Yasin, Javed Ali Khan and Lijie Wen

Received: 21 October 2024

Revised: 6 December 2024

Accepted: 9 December 2024

Published: 16 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In light of the substantial progression within the realms of industrial internet and sensor technologies, industrial condition monitoring data are growing explosively, with over 1000 EB being added every year [1]. The use of industrial big data for equipment health monitoring and remaining useful life prediction has great significance for achieving equipment predictive maintenance and improving production efficiency. Bearings are fundamental elements that are indispensable to the functionality of mechanical systems, and the decline in functionality and collapse of bearings, stemming from fatigue, abrasion, and shape alteration, can lead to economic damage as well as pose risks to human safety [2,3]. Consequently, assessing the RUL of bearings is essential for the upkeep and administration of mechanical equipment [4,5]. And forecasting their RUL is attracting increasing attention [6].

The methods for predicting the remaining useful life (RUL) of rolling bearings are typically classified into three main categories: model-dependent approaches, data-centric techniques, and combined forecasting strategies [7]. Model-based techniques encompass physical modeling and empirical modeling strategies [8]. A physical model is used to understand the mechanism of bearing degradation and establish a degradation model with physical meaning to predict the RUL of the bearing [9]. An empirical model is a degradation model summarized during the long-term operation of the equipment. The degradation of bearing performance is examined using a stochastic process model, with the RUL forecasts being performed by determining the probability distribution of the initial passage time within this model [10]. Model-based methods require a lot of prior knowledge, and the complexity of operating conditions and degradation processes makes it difficult to accurately model degradation [11,12]. Data-driven methods can make good use of the historical data of condition monitoring during machine operation and create a correlation between sensor data and residual life using deep learning or machine learning to facilitate the forecasting of the RUL. In contrast to model-based approaches, it relies less on prior knowledge and can mine the degradation information of bearings hidden in historical data by using powerful nonlinear fitting abilities [13–15]. An optimal equilibrium is struck concerning predictive precision, complexity of modeling, and overall generalization capability. The hybrid approach integrates the features of various forecasting models to estimate the RUL [16]. In recent years, massive sensor data and powerful computing power have provided good soil for the development of data-driven methods [17]. In contrast to traditional machine learning methods like support vector machines (SVMs) [18] and artificial neural networks (ANNs) [19], deep learning frameworks offer a different approach, and Gaussian process regression (GPR) [20] exhibits a reduced reliance on domain-specific feature engineering, possesses robust generalization capabilities, and can harness complex, high-dimensional data to deliver enhanced predictive capabilities. This has led to their extensive adoption in the prognostics domain for remaining useful life (RUL) estimation [8,14,21].

Frameworks and their adaptations, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and the Transformer architecture, have demonstrated exceptional results in the field of remaining useful life (RUL) prediction for roller bearings. Owing to its distinctive looped architecture, the RNN framework is adept at capturing temporal sequence characteristics, making it highly appropriate for bearing RUL forecasting. Its extensions, long short-term memory (LSTM) and gated recurrent unit (GRU), have gained extensive application in this domain. Shen et al. [22] proposed a bidirectional LSTM model based on the multi-head attention mechanism for the RUL prediction of rolling bearings, embedding a multi-head attention mechanism in an LSTM unit to weaken the impact of redundant information and improve the performance of model. Que et al. [23] used the GRU network to predict the bearing RUL to improve the information extraction ability of the model; an attention mechanism based on Dynamic Time Warping (DTW) was proposed, and the Bayesian method was used to evaluate the uncertainty of the prediction. Although fully connected RNNs have demonstrated their effectiveness in RUL prediction tasks, nonetheless, certain limitations persist; LSTM and GRU theoretically can capture infinitely long historical information, but experiments have shown that they perform poorly on long-sequence time-series [24]. In addition, they cannot be parallelized because of their cyclic structure. The temporal convolutional network (TCN) has been shown to outperform the RNN on time-series forecasting problems [25]. Chen et al. [26] designed an adversarial TCN network based on Bayesian optimization, which avoids the problem where RNNs cannot be parallelized. However, the convolution-based processing is still local, and the long-term dependency problem has not been well solved. Different from the information update and persist strategy of LSTM, the Transformer model is capable of directly forging a connection between the input data and the target output via its multi-head self-attention mechanism, thereby addressing the issue of long-range dependency effectively.

Chang et al. [27] used the Transformer model to predict the RUL of rolling bearings and achieved extremely high prediction accuracy.

The one-dimensional, manually extracted time domain features, frequency domain features, and health indicators are still used in training data without fully unleashing the potential of deep learning models. Different from time domain and frequency domain features, time–frequency domain features can observe the frequency change over time in non-stationary vibration signals, which is a good input for the RUL prediction model; while regular fully connected RNNs cannot directly process image input, CNNs and ConvLSTMs are capable of addressing this issue. Yoo Y. et al. [28] used the time–frequency image of a wavelet transform to construct a health indicator that can characterize the operating state of rolling bearings and used a CNN model to process the wavelet power spectrum, clustered bearings with different failure modes, and obtained a high-accuracy RUL prediction value. But the threshold used to determine the failure status of each bearing cluster is manually set based on experience, which makes the proposed method have limitations. Shi et al. [29] first proposed ConvLSTM for weather forecasting, which has a good ability to capture spatial–temporal correlation and is very suitable for time–frequency domain features of RUL prediction. However, the problem of long-term dependence on the LSTM model has not been well solved. Table 1 summarizes key findings and methodologies from recent studies relevant to this research.

**Table 1.** Key findings from related works.

Models	Strengths	Limits
LSTM [21]	Handling long-term dependencies, preventing gradient vanishing issues, high-capacity memory units, and good learning capabilities.	High computational complexity, long training time, strong dependence on data, inability to handle sequences with large time spans
Transformer [27]	Parallel processing capabilities, effective capture of long-distance dependencies, flexibility, and scalability.	High computational and memory requirements, dependence on training data, complex optimization and adjustment.
ConvLSTM [29]	Handling spatial and temporal information, improving prediction accuracy, and powerful feature extraction capabilities.	High computational complexity, numerous parameters, prone to overfitting, and high requirements for data quality and quantity
ConvNeXt [30]	Excellent performance, high accuracy, and high scalability.	Overimitation of ViT, hindering operator fusion, limiting efficiency and flexibility.
SA-ConvLSTM [31]	Enhancing spatiotemporal prediction accuracy, extracting global and local spatial features, fewer parameters, and higher time efficiency.	Computational complexity, dependence on data quality, parameter adjustment and optimization.
This study	High prediction accuracy and strong interpretability, suitable for handling sequences with large time spans	Higher model complexity and more parameters

In view of the above problems, it is an obvious idea to integrate different deep learning models. Therefore, this paper proposes an end-to-end adaptive rolling bearing RUL prediction model using time–frequency image features and discusses how to combine CNN, LSTM, and Transformer models. The CNN has powerful image feature extraction capabilities and is often used as a backbone feature extraction network for image classification, object detection, and other tasks. ConvNeXt [30] stands out as one of the premier choices, offering an optimal trade-off between precision and computational efficiency. Therefore, we downsample the time–frequency spectrograms after a short-term Fourier transform to reduce the computational amount of the subsequent network and extract features from spectrograms by ConvNeXt. Secondly, parallel SA-ConvLSTM and Transformer are used to mine the degradation information in the feature map. Ultimately, the channel attention mechanism is integrated to facilitate the adaptive assignment of distinct weights to the regression sequences pertaining to the SA-ConvLSTM and Transformer models, thereby synthesizing their respective proficiency in local and global feature discernment.

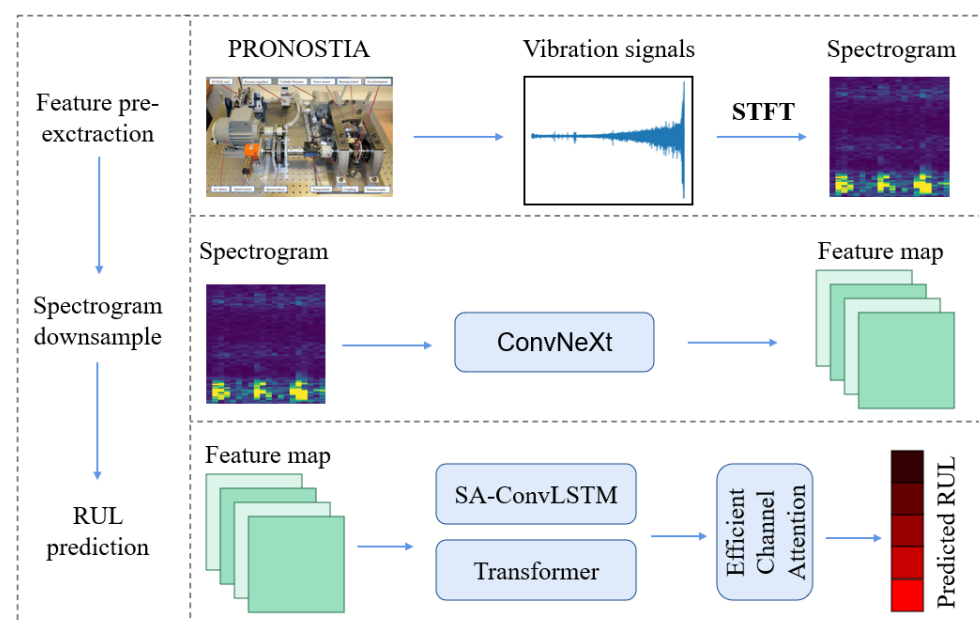
The primary advancements introduced by the methodology delineated in this study are outlined below.

1. By bypassing complex feature extraction and health indicator establishment methods, an end-to-end remaining useful life prediction system is created that yields predictions by solely inputting the raw vibration data. It eliminates the disturbance of man-made factors and takes full advantage of deep learning models;
2. We tried to explore the effective combination of CNN, LSTM, and Transformer models, and a hybrid structure of CNN, LSTM, and Transformer combining serial and parallel features is designed. Among them, CNN is used for the downsampling and feature extraction of the original time–frequency map, and parallel LSTM and Transformer are used to further learn and transform the degradation information in the feature map;
3. The efficient channel attention mechanism is introduced to quantify and learn the contribution of the features from two parallel models at each time step for RUL prediction, adaptively assign attention for the output feature channels of the two models at each moment, acquire useful information of the parallel channel, and harness the benefits of both models to augment the precision of predictions.

The subsequent chapters of this article are as follows. In Section 2, we will provide a detailed introduction to the neural network model we designed, In Section 3, the verification experiments and outcomes are explored and analyzed with the public bearing dataset, and finally, Section 4 contains the conclusion.

## 2. Methods

We proposed an end-to-end deep learning framework for predicting the RUL of rolling bearings, as shown in Figure 1. The framework is composed of five integral components: STFT, ConvNeXt, SA-ConvLSTM, Transformer, and efficient channel attention (ECA), incorporating the component dedicated to feature delineation, the time–frequency image downsampling part, and the RUL prediction part, which correspond to the three steps of the RUL prediction process, respectively. First, the unprocessed one-dimensional vibration data are converted into a two-dimensional time–frequency representation, namely a spectrogram, through the application of the short-time Fourier transform technique. Second, the original spectrogram is downsampled using ConvNeXt. Finally, the downsampled feature maps are input into parallel SA-ConvLSTM and Transformer networks, and their prediction results are adaptively fused by the ECA module to obtain the final RUL prediction value.



**Figure 1.** Procedure of the proposed approach.



### 2.1. Feature Selection and Processing

The vibration signal of rolling bearings contains rich degradation information that can be used for RUL prediction. However, the frequency components in the vibration signal are complex, and the characteristic frequencies of various faults are mixed with each other, making it extremely challenging to determine sensitive features that represent the degradation state of bearings. Features in the time domain and frequency domain can, to some degree, indicate the degradation status of bearings. But considering the non-stationary nature of signals and the need to preserve more information from the raw vibration signal, this paper integrates STFT for signal processing in the first stage of the prediction framework.

The idea of STFT comes from a sliding window, and the choice of window function and window length has a great impact on its results. To enhance spectral resolution, it is crucial to constrict the main lobe of the analysis window while maintaining that the neighboring lobes are faint and their amplitude drops rapidly, which mitigates the effects of spectral distortion. The window size has a direct impact on the resolution of the time–frequency analysis: a larger window length results in poorer time resolution, whereas a smaller window length leads to poorer frequency resolution. This study aims to establish an end-to-end model that minimizes the influence of manually processed features on the model’s prediction results. Therefore, considering all these elements, a widely used Hamming window known for its superior spectral performance is selected as the window function. For window length, the characteristic frequencies of rolling bearing faults are typically within the range of several tens to thousands of Hz, and the sampling frequency of the test bench is 25.6 kHz (2560 samples/0.1 s); considering the low fault frequencies and abrupt failures of the bearing, the window length  $L$  was set to 102 to balance the frequency and time resolution, ensuring that the data dimensions of both were consistent after transformation. An overlap of 50% was commonly adopted. Figure 2 displays the STFT outcomes, signifying that with the degradation of bearing health, the energy in the low-frequency part of the spectrum increases significantly and becomes more concentrated. Additionally, the results of transformations using different windows and lengths are depicted in Table 2, demonstrating that the parameters chosen in this study meet the necessary requirements.

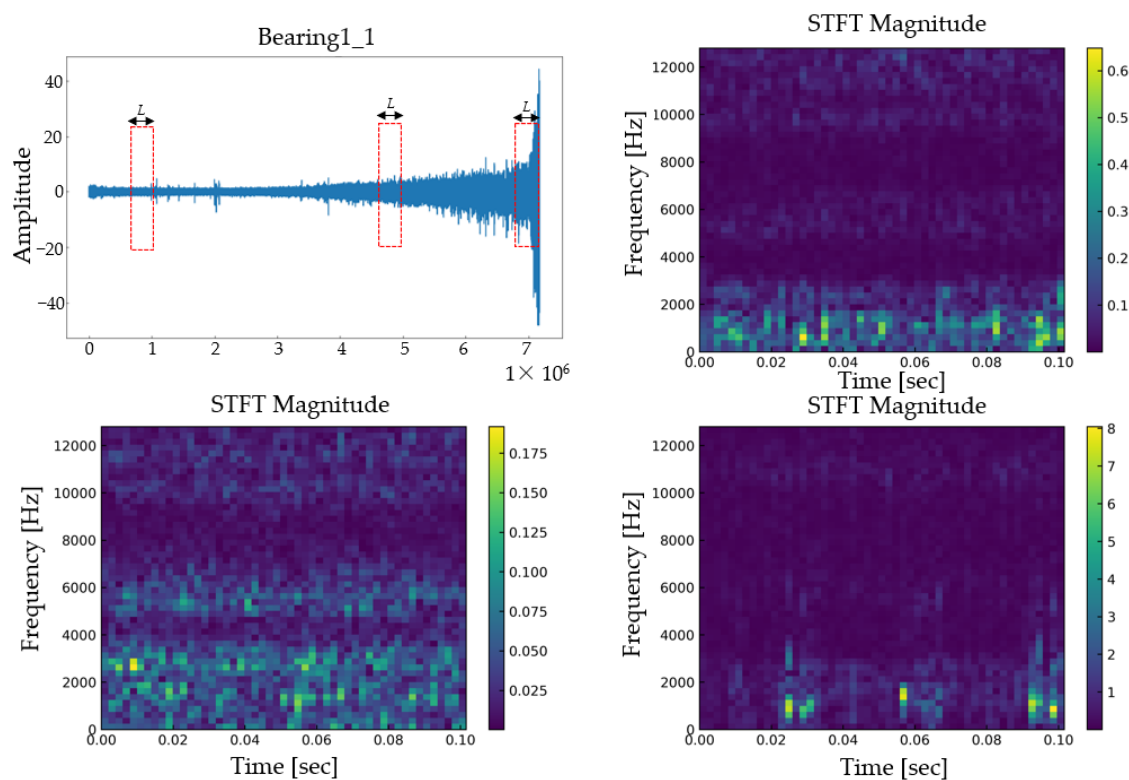
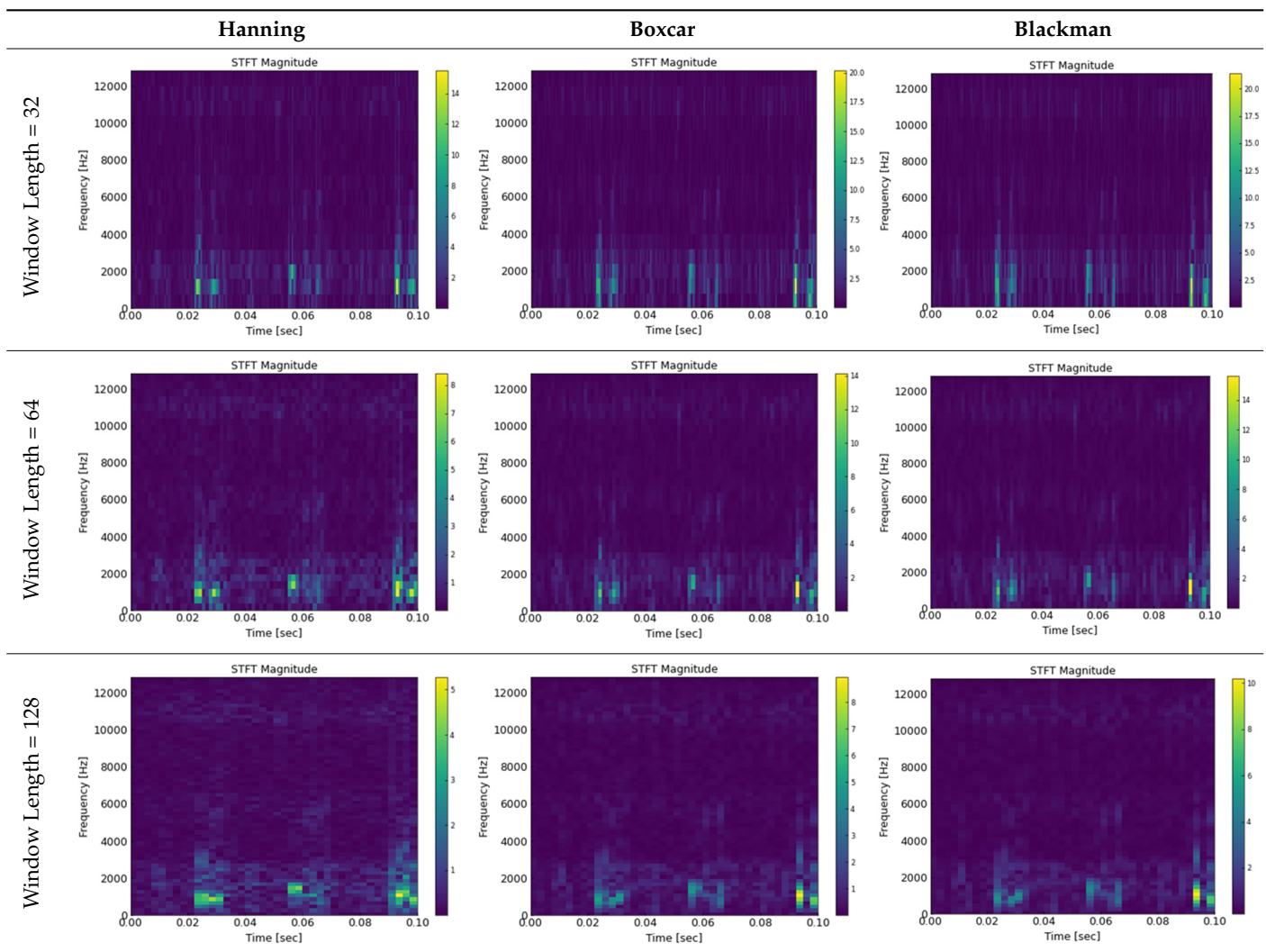


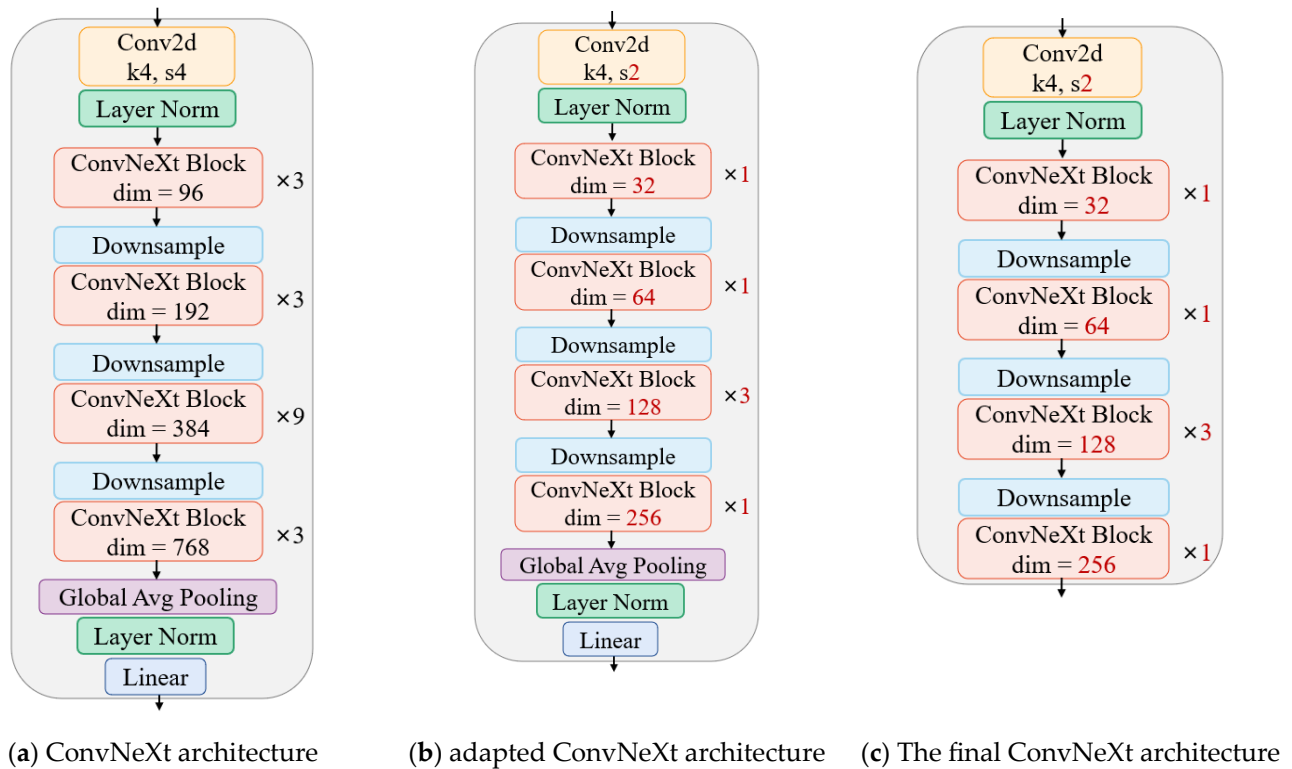
Figure 2. Degradation state of the bearing.

Table 2. Results of STFT with different parameters.



## 2.2. Downsampling

The CNN has an excellent ability to capture the local features of images. After STFT, the size of the time–frequency map is  $52 \times 52$ . Directly inputting it into the SA-ConvLSTM network will result in high computational overhead. Downsampling the original image is undoubtedly a better choice, and we can further learn the deep representation of rolling bearing fault features. ConvNeXt is a masterpiece of a CNN that has achieved excellent results in both model performance and FLOPs. Therefore, this paper uses it as the downsampling and in-depth feature extraction part of the entire framework. There are four versions of ConvNeXt: T/S/B/L, with increasing FLOPs. Considering the size of the time–frequency map in this paper, the T version with lower FLOPs was selected and modified. First, the stride of the convolution layer in the downsampling layer connected to the input image was reduced from 4 to 2. Secondly, the number of times each module in the model was stacked, and the number of channels were proportionally reduced. Figure 3b illustrates the adapted ConvNeXt architecture. At the same time, the layer used for image classification in the original model's output part was removed to facilitate the connection of the SA-ConvLSTM and Transformer modules. The ultimate design is depicted in Figure 3c.



**Figure 3.** Structures of ConNeXt.

The original time–frequency image is downsampled by ConvNeXt, reducing the image size from  $1 \times 52 \times 52$  to  $256 \times 3 \times 3$ . With a batch size of 1 and a sequence length of 6, the computational cost of this process is 34.8 M FLOPs. The computational costs of each model are shown in Table 3, which indicates that adding downsampling can significantly reduce the computational cost of the model in the prediction framework.

**Table 3.** FLOPs of models.

Models	FLOPs
ConvNeXt	34.8 M
SA-ConvLSTM	233.6 M
Transformer	15.5 M
ConvNeXt + SA-ConvLSTM	43.5 M
ConvNeXt + Transformer	14.3 M

### 2.3. RUL Prediction

The downsampled feature map can be directly used as the input of the SA-ConvLSTM module, but for the Transformer module, the original 2D feature map needs to be flattened into a 1D sequence. Here, we drew inspiration from the design of ViT [32], and the specific structure is shown in Figure 4.

In the prediction part of the framework, we adopt a parallel SA-ConvLSTM and Transformer structure to combine the advantages of both models in mining local and global degradation information for better prediction performance. The Transformer model has excellent global feature extraction ability and can capture the overall degradation trend, but its ability to extract fine-grained local features is insufficient. Shallow Transformers tend to ignore local feature information, while local abrupt information in rolling bearing vibration signals is crucial for predicting RUL. Therefore, complementary SA-ConvLSTM is added to extract local degradation information in the feature map. Figure 5 shows the specific structure of the prediction part.

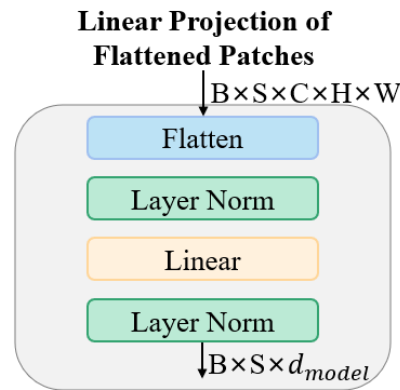


Figure 4. Structure of the linear projection of flattened patches.

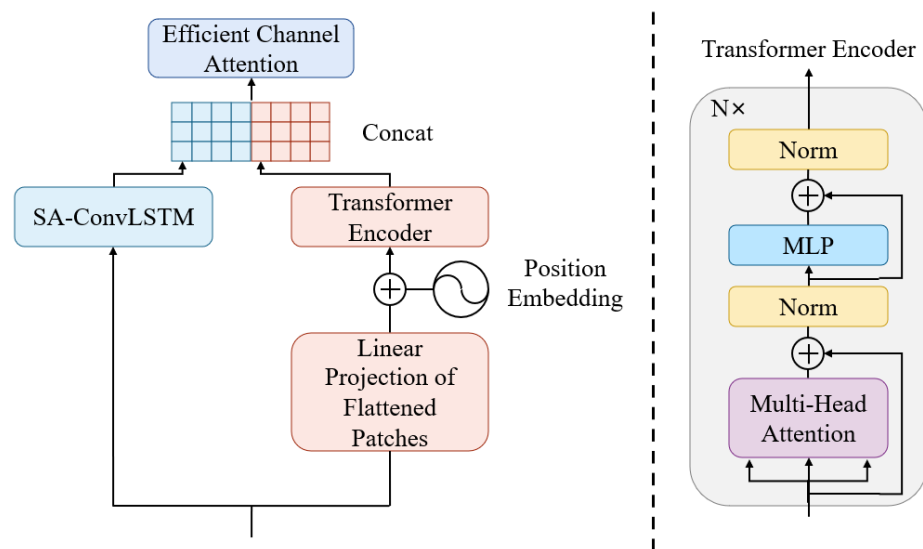


Figure 5. Structure of the prediction module.

To better integrate the prediction results of the two modules, they were concatenated. All time steps were processed in a loop. The ECA [33] module was used to adaptively assign weights to the prediction results at each time step. The ECA module uses a 1D convolution with weight sharing for inter-channel information exchange, realizing channel attention extraction. It does not significantly increase the FLOPs and can effectively improve the performance of the fusion model. The specific structure is shown in Figure 6.

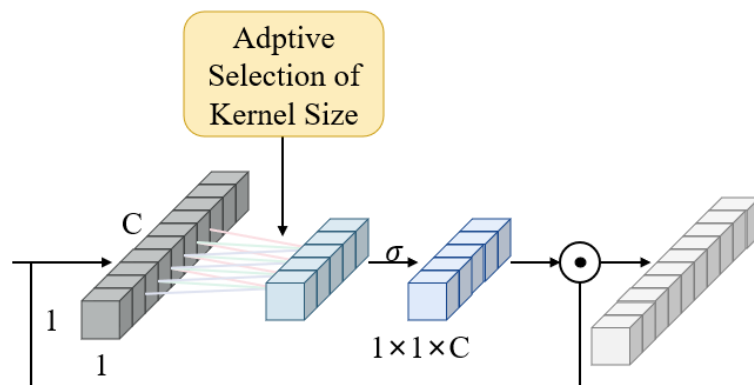


Figure 6. Structure of the ECA module.

The main process of the proposed prediction framework has been introduced. Finally, the training samples of the model were constructed based on a sliding window. The raw data were collected every 10 s, with 2560 data points being collected each time, taking 0.1 s. Therefore, using 10 s as a time step, the raw data were first reshaped, and then training samples were constructed with a sliding window of 1. The training samples have labels indicating the percentage of RUL of rolling bearings at the current time step with respect to the total life.

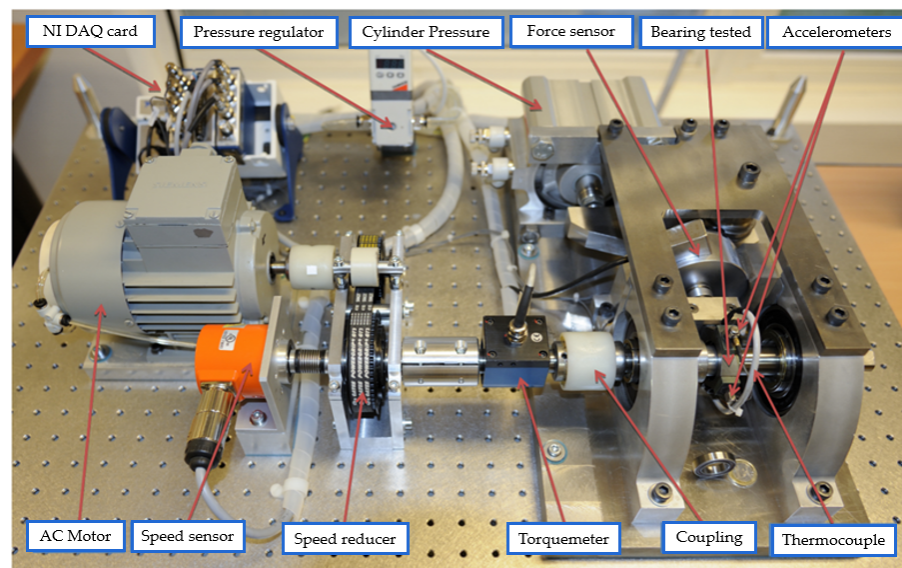
In engineering practice, the lifespan of different bearings, or the same bearing under different operating conditions, can vary significantly. Directly using the actual remaining useful life (RUL) of the bearing as the label makes it difficult to ensure consistency in the dataset labels. This paper employs a normalized remaining useful life label, with the specific formula being:

$$RUL_{norm} = 1 - \frac{RUL_t}{T}$$

In the formula,  $T$  represents the total operating time of the bearing,  $RUL$  denotes the actual remaining useful life at the current moment, and NRUL represents the normalized remaining useful life, ranging from 0 to 1.

### 3. Materials

In this paper, data from the IEEE PHM2012 Challenge dataset were used to verify the validity of the proposed method, as shown in Figure 7. The data were collected by the FEMTO-ST Institute on a rolling bearing test bench for accelerated life testing in PRONOSTIA [34]. Two DYTRAN accelerators were used to collect vibration signals in both the horizontal and vertical axis of each bearing, with a sampling frequency of 25.6 kHz, sampling duration of 0.1 s, and sampling interval of 10 s.



**Figure 7.** Overview of PRONOSTIA.

The dataset contains 6 training data of full lifecycles and 11 truncated testing data under three operating conditions, as shown in Table 4. The rolling bearing was subjected to a horizontal load during the test. Previous studies by Soualhi [35], Singleton [36], and others have shown that useful degradation information contained in vertical vibration signals is minimal compared with horizontal vibration signals. Therefore, this paper finally uses vibration signal data in the horizontal direction of the bearing for experimental verification.



**Table 4.** PHM2012 dataset.

Condition	Load	Rotation Speed	Train	Test
1	4000 N	1800 rpm	Bearing 1_1 Bearing 1_2	Bearing 1_3 Bearing 1_4 Bearing 1_5 Bearing 1_6 Bearing 1_7
2	4200 N	1650 rpm	Bearing 2_1 Bearing 2_2	Bearing 2_3 Bearing 2_4 Bearing 2_5 Bearing 2_6 Bearing 2_7
3	5000 N	1500 rpm	Bearing 3_1 Bearing 3_2	Bearing 3_3

### 3.1. Evaluation Metrics

To precisely assess the functionality and practical utility of a model, the root mean square deviation (*RMSE*), mean absolute error (*MAE*), and *SCORE* are used as evaluation metrics for the model, where *RMSE* and *MAE* are used to assess the model's fitting performance, with the following calculation formulas:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (actRUL_i - preRUL_i)^2}, \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |actRUL_i - preRUL_i|, \quad (2)$$

The above two evaluation metrics treat early and late predictions equally. In the formula, '*actRUL*' refers to the actual remaining useful life, and '*preRUL*' refers to the predicted remaining useful life of the bearing. But in reality, early predictions are more valuable as they can provide correct maintenance decision-making information. To address this, the PHM 2012 Challenge dataset provides a *SCORE* metric that accurately evaluates the actual performance of the model, applying different degrees of punishment to early and late predictions, as shown in Equation (3).

$$A_i = \begin{cases} \exp^{-\ln(0.5) \cdot (Er_i/5)} & \text{if } Er_i \leq 0 \\ \exp^{+\ln(0.5) \cdot (Er_i/20)} & \text{if } Er_i > 0 \end{cases} \quad (3)$$

where  $E_i = \frac{actRUL_i - preRUL_i}{actRUL_i} \times 100\%$  is the error between the true value and predicted value of the remaining life of the bearing and  $A_i$  is the final score.

### 3.2. Controlled Experiment

All experiments in this paper are conducted on a Windows 10 64-bit operating system computer with an NVIDIA 3060 GPU and Intel i5-10400F CPU, using Python (3.8.10) as the programming language, PyTorch (1.9.0) as the deep learning framework, CUDA version 11.3, and NumPy version 1.19.5.

To further investigate the integration of long short-term memory (LSTM) and Transformer models, we designed six controlled experiments, divided into multiple groups with different sequence lengths to verify the performance of LSTM and Transformer in capturing short-term and long-term dependencies, as shown in detail in Table 5. ConvNeXt (b) and ConvNeXt (c) are the b and c models corresponding to Table 3, respectively. As the regular fully connected LSTM cannot directly process downsampled two-dimensional feature maps, an additional linear layer is required. Adding the linear layer directly to the downsampling part and adding it to the LSTM part are essentially the same. Therefore, ConvNeXt (b) is directly used as the downsampling part of Experiment A.

**Table 5.** Setting of the comparative experiments.

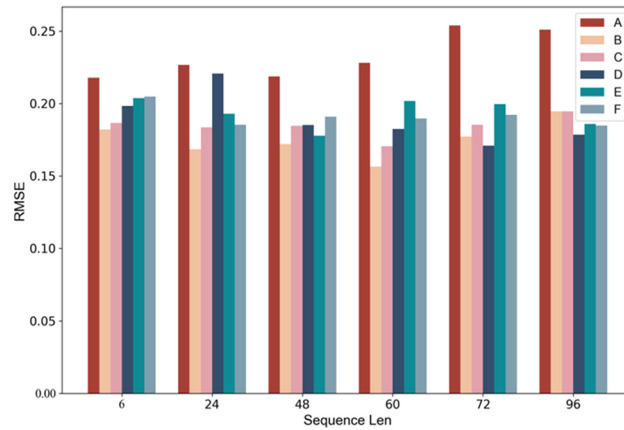
Experiment	Setting
A	ConvNeXt(b) + LSTM
B	ConvNeXt(c) + ConvLSTM
C	ConvNeXt(c) + SA-ConvLSTM
D	ConvNeXt(c) + Transformer
E	Fusion + ECA
F	Fusion + Linear

Experiment A uses downsampled feature maps from ConvNeXt as the input for a regular LSTM. Experiments B and C replace the LSTM in the prediction part of Experiment A with ConvLSTM and SA-ConvLSTM, respectively, for comparison. In Experiment D, ConvNeXt is used as the feature extraction part and Transformer is used as the final prediction part. Experiment E proposes a fusion model that uses the ECA module to fuse the regression sequences output by SA-ConvLSTM and Transformer. Experiment F simply fuses the two outputs using a linear layer. For all control experiments, hyperparameters relating to the training process are kept as consistent as possible to reduce experimental error, and the model architecture parameters are set to commonly used values, as shown in Table 6. The data from seven bearings operating under condition one were utilized for training and validation purposes.

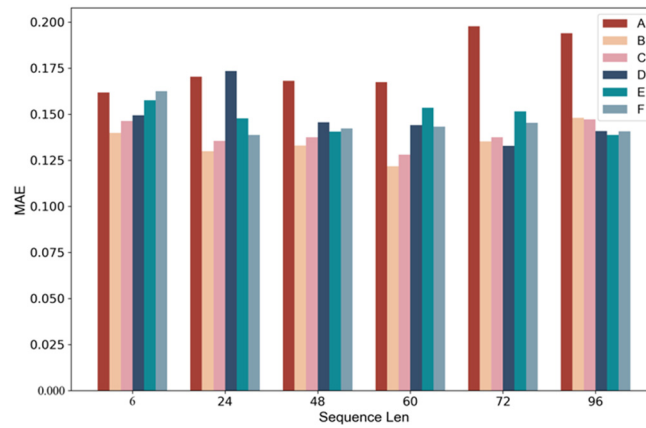
**Table 6.** Hyperparameters of models.

Model	Hyperparameters of Training	Title 3
ConvNeXt		Depths = [1, 1, 3, 1] Dims = [32, 64, 128, 256]
LSTM		Hidden size = 64 Num layer = 2 Bidirectional = False
ConvLSTM	Epoch = 10 Batch size = 4 Learning rate = 0.0001 Loss function = mse Optimizer = AdamW	Hidden channel = 64 Num layer = 2 Bidirectional = False
SA-ConvLSTM	Learning rate decay = step-based Patience = 1	Hidden channel = 64 Attn hidden channel = 64 Num layer = 2 Bidirectional = False
Transformer		Model dim = 512 Num head = 8 Dim head = 64 Mlp dim = 1024 Depth = 2

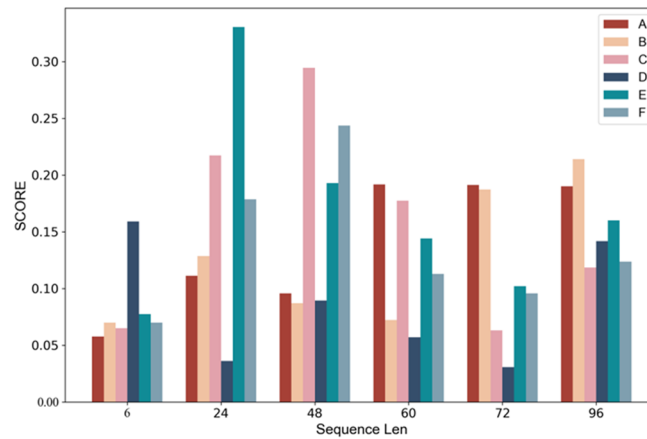
The sequence length markedly influences the model's predictive outcomes. Longer input sequences may help to better capture the long-term degradation trend of the bearings but may also increase the computational cost during model training. Additionally, different models may have varying optimal sequence lengths for achieving the best performance. Therefore, we conducted experiments using sequence lengths of 6, 24, 48, 60, 72, and 96 in the controlled experiment, as shown in Figure 8, to better understand these models.



(a) RMSE of models with different sequence lengths.



(b) MAE of models with different sequence lengths.



(c) SCORE of models with different sequence lengths.

**Figure 8.** Performance of models with different sequence lengths.

The RMSE and MAE serve as indicators of the model’s approximation capabilities to a degree. From the experimental results in the first two figures, it can be concluded that all models except for the original LSTM model have very similar RMSE and MAE metrics. Therefore, it is necessary to understand the characteristics of different deep learning models to better apply them in this task. The ConvLSTM and SA-ConvLSTM models perform well in short-sequence time-series, while the Transformer model performs better in longer sequences. Therefore, by combining the two, the fused model’s two metrics perform

differently at different sequence lengths but are overall comparable with a single model, with a slight increase compared with the best values of the two individual models. In addition, from the last figure, it can be seen that the RMSE and MAE metrics cannot fully measure the performance of the model in practical applications, and there is no significant correlation with the SCORE metric. Different models achieve their optimal SCORE metric at different sequence lengths, but the ECA module can combine the advantages of the two parallel models well, and its SCORE metric surpasses all models at a sequence length of 24.

Building upon the previous section, we optimize the hyperparameters of the fusion model using grid search and compared our proposed method with multiple advanced RUL prediction methods to validate its effectiveness. Based on the hyperparameter selection in related papers and empirical knowledge, we determined the hyperparameters that affect network performance and their search range, forming a hyperparameter space. We traversed all hyperparameter combinations in the space and trained the network model with the raw vibration signal as the input and the RUL labels as the output. The best combination of hyperparameters identified in the search space is presented in Table 7.

**Table 7.** Hyperparameters of the selected models.

Block	Hyperparameters	Band	Optimal
ConvNeXt	Depths		[1, 1, 3, 1]
	Dims		[32, 64, 128, 256]
SA-ConvLSTM	Hidden channel	[32, 64]	64
	Attn hidden channel	[32, 64]	64
	Num layer	[1, 2]	2
Transformer	Model dim	[256, 512]	256
	Num head	[4, 8]	4
	Mlp dim	[128, 1024]	1024
	Depth	[1, 2]	2
Training	Batch size	[4, 8]	4
	Learning rate	[0.0001, 0.001]	0.0001

The results of training and testing using the hyperparameters in this group are shown in Table 8 for the two metrics, RMSE and MAE. Although not all individual bearing metrics are at their minimum values, they are generally at a lower level, with mean RMSE and MAE achieving the best levels among the comparison methods, indicating a good fitting performance of the model.

**Table 8.** Comparison of different methods.

Methods	Proposed		Method [37]		Method [38]		Method [39]	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Bearing 1_3	0.0749	0.0585	0.1489	0.1303	0.0705	0.0563	0.0873	0.0655
Bearing 1_4	0.1186	0.1012	0.0824	0.0721	0.1689	0.1443	None	None
Bearing 1_5	0.1217	0.0961	0.1983	0.1735	0.3467	0.2522	0.2368	0.1902
Bearing 1_6	0.2776	0.2453	0.1990	0.1741	0.3089	0.2333	0.5387	0.4628
Bearing 1_7	0.0627	0.0479	0.2120	0.1855	0.3455	0.2479	0.2126	0.1775
Ave	0.1639	0.1191	0.1681	0.1471	0.2481	0.1868	0.2688	0.2240

Compared with other bearings, In the test dataset, Bearings 1\_5 and 1\_6 produced larger errors in all methods, so an analysis was performed on the raw training and testing data as well as prediction results. The seven bearing data for operating condition one are shown in Figures 9 and 10. Bearings 1\_1 and 1\_2 are used as training data. The amplitude of Bearing 1\_1 slowly increases when degradation begins, while the amplitude of Bearing 1\_2 slightly increases at the beginning of degradation and then rapidly increases as it approaches failure, ending the service life of the bearing. The two bearings have

significantly different failure modes. Bearing 1\_1 has a longer service life, so the training data for this failure mode were more sufficient, but the short service life of Bearing 1\_2 resulted in a shortage of training data for this failure mode.

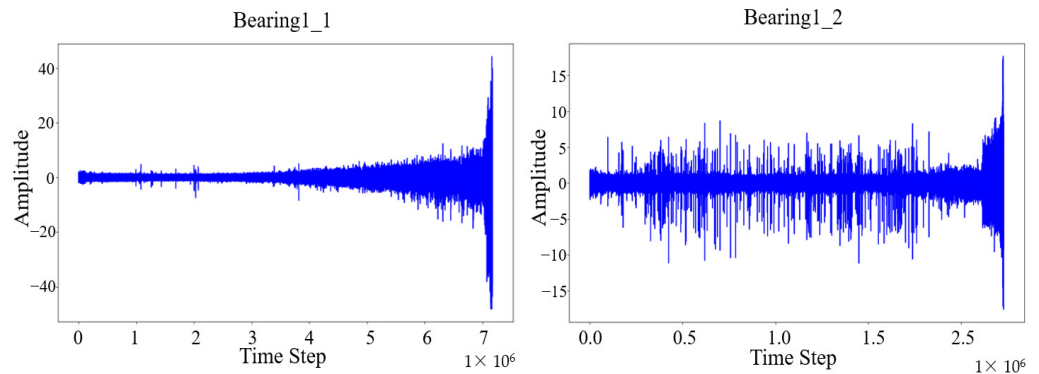


Figure 9. Training set of bearings.

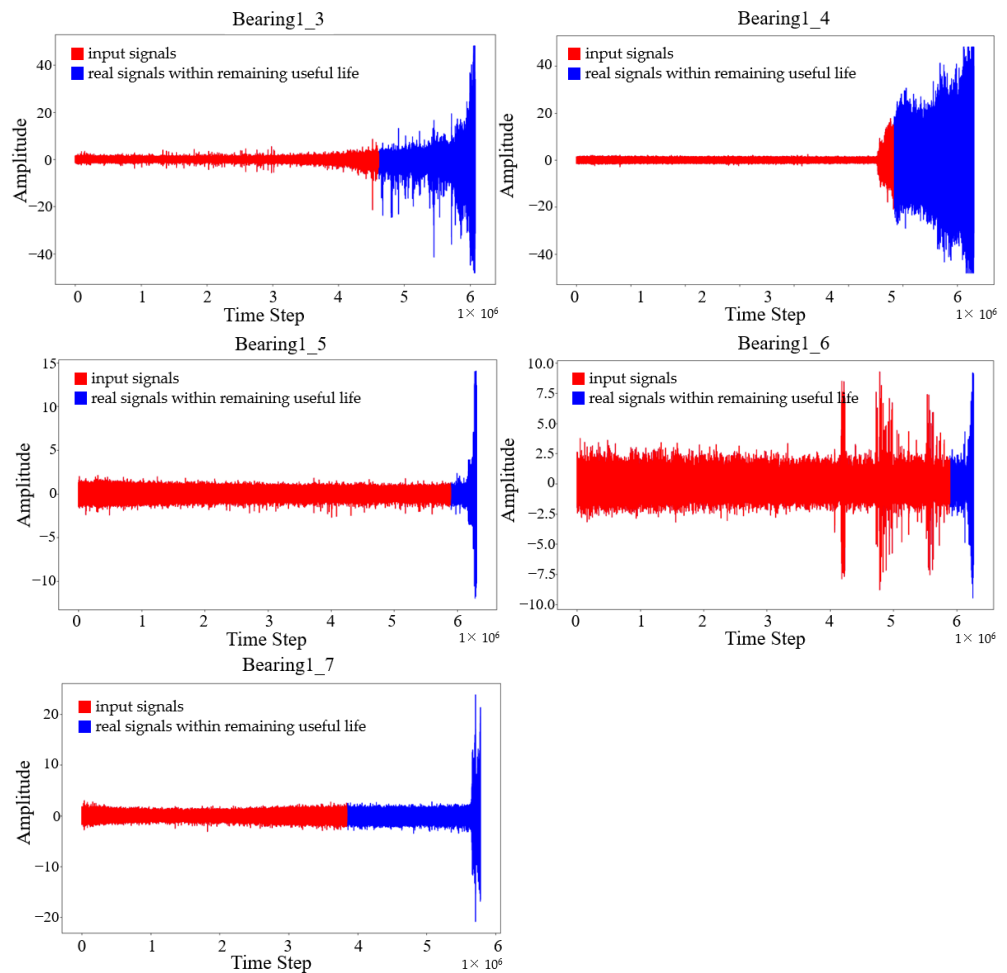
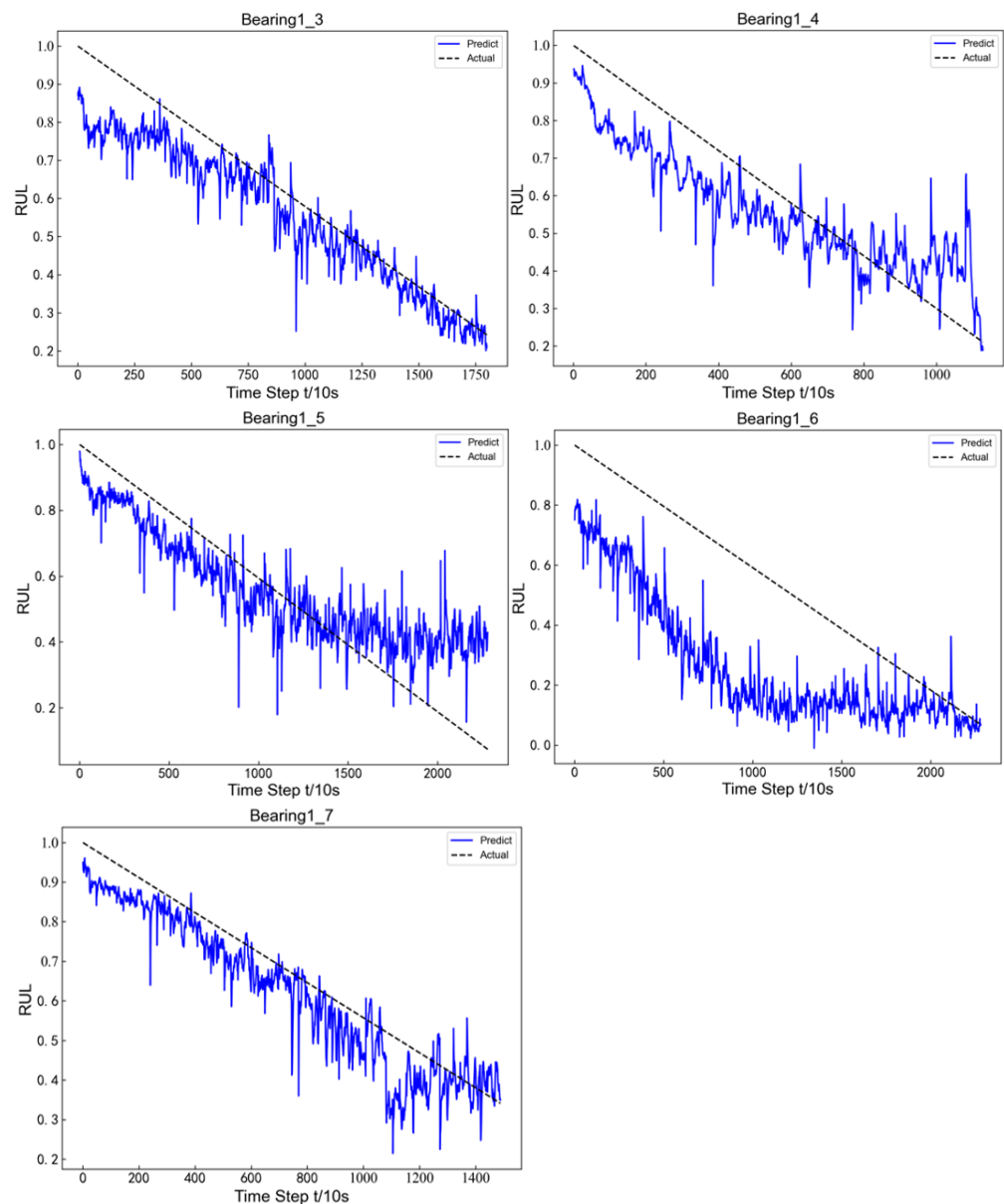


Figure 10. Test set of bearings.

Figure 10 illustrates the vibration data for the complete lifespan of the bearings under examination, with the blue segment representing the full dataset and the red segment indicating the abbreviated test data required for the model input. Observations suggest that their failure patterns are largely consistent with the training data, where Bearing 1\_3 and Bearing 1\_4 have similar failure patterns to Bearing 1\_1 in the training set, and the vibration amplitudes during failure are also very close. Bearing 1\_5, Bearing 1\_6, and Bearing 1\_7 have failure patterns similar to Bearing 1\_2 in the training set.



The model's prediction results are shown in Figure 11. Among them, the degradation trend prediction of Bearing 1\_3 is the most accurate, and it fits the real degradation curve well in the second half of the entire degradation cycle. The other four bearings all showed some prediction errors in the second half of the degradation cycle. In Figure 10, the last part of the truncated test data of Bearing 1\_4 is a more obvious representation of the degradation state, so the model's predicted value quickly approached the real degradation curve at the end of the predicted time in Figure 11. This indicates that the model can accurately capture the degradation state information contained in the vibration data. However, Bearings 1\_5, 1\_6, and 1\_7 did not show a significant downward trend in the second half of the degradation compared with the first half, which may be related to insufficient training data for this failure mode. Nevertheless, the model still captured useful degradation information, Bearings 1\_6 and 1\_7 displayed a sudden drop in RUL as they approached failure, particularly for Bearing 1\_7, and the RUL of Bearing 1\_5 exhibited significant fluctuations. These findings indicate that, in the case of abrupt failure modes, the model can still provide relatively accurate early warning information and predict the RUL of the bearings effectively.



**Figure 11.** RUL prediction of bearings.

PCA analysis is a commonly used method for feature dimensionality reduction. It linearly projects the original sample points in high-dimensional space into low-dimensional space through a projection matrix while preserving the features that contribute the most to the variance in the data to reduce information loss and extract the main features of high-dimensional data. Linear projection maps the data to the most suitable projection direction through matrix transformation. We used the PCA method to reduce the dimensionality of the merged features from two parallel channels, and the visualization result is shown in Figure 12.

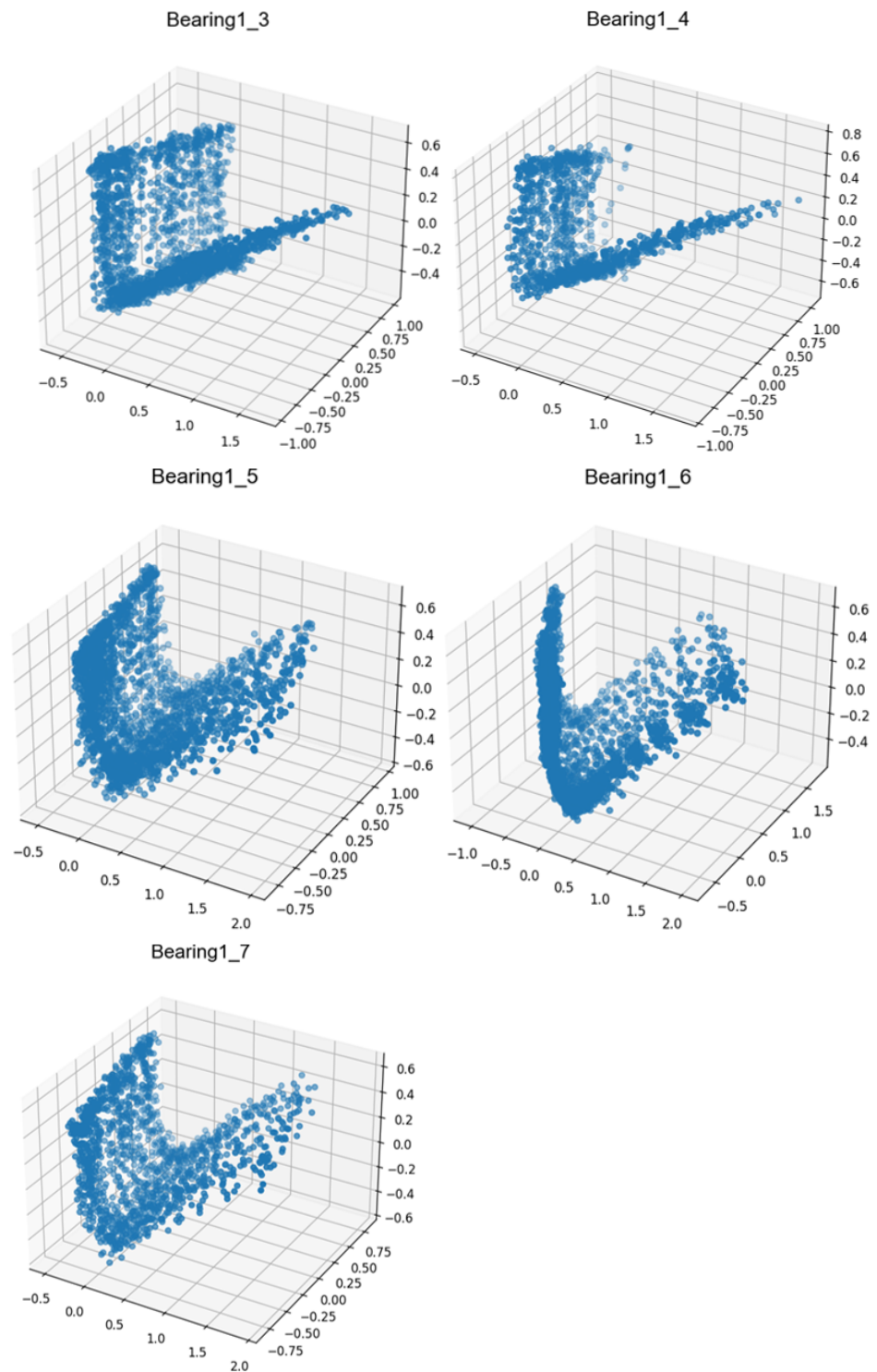


Figure 12. Visualization of the features extracted by the fusion model.

From Figure 12, it can be clearly seen that Bearings 1\_3 and 1\_4 have very similar distribution patterns in the projection space. Bearings 1\_5, 1\_6, and 1\_7 also have similar distributions in space, but there is a slight difference between the distribution pattern of Bearing 1\_6 and that of Bearings 1\_5 and 1\_7. This is consistent with the observation pattern of the raw vibration signals. To further illustrate the capabilities of SA-ConvLSTM and Transformer in capturing local and global features, using PCA on the features extracted by both models, the result is shown in Table 9. It can be observed that, for all bearings, the features extracted by Transformer exhibit a largely consistent distribution pattern, while the SA-ConvLSTM shows significant variations. However, the results obtained from fusion model align well with the failure modes observed in the raw data. This demonstrates that the proposed model effectively integrates the strengths of both approaches in capturing local and global features.

**Table 9.** Visualization of the features extracted by SA-ConvLSTM and Transformer.

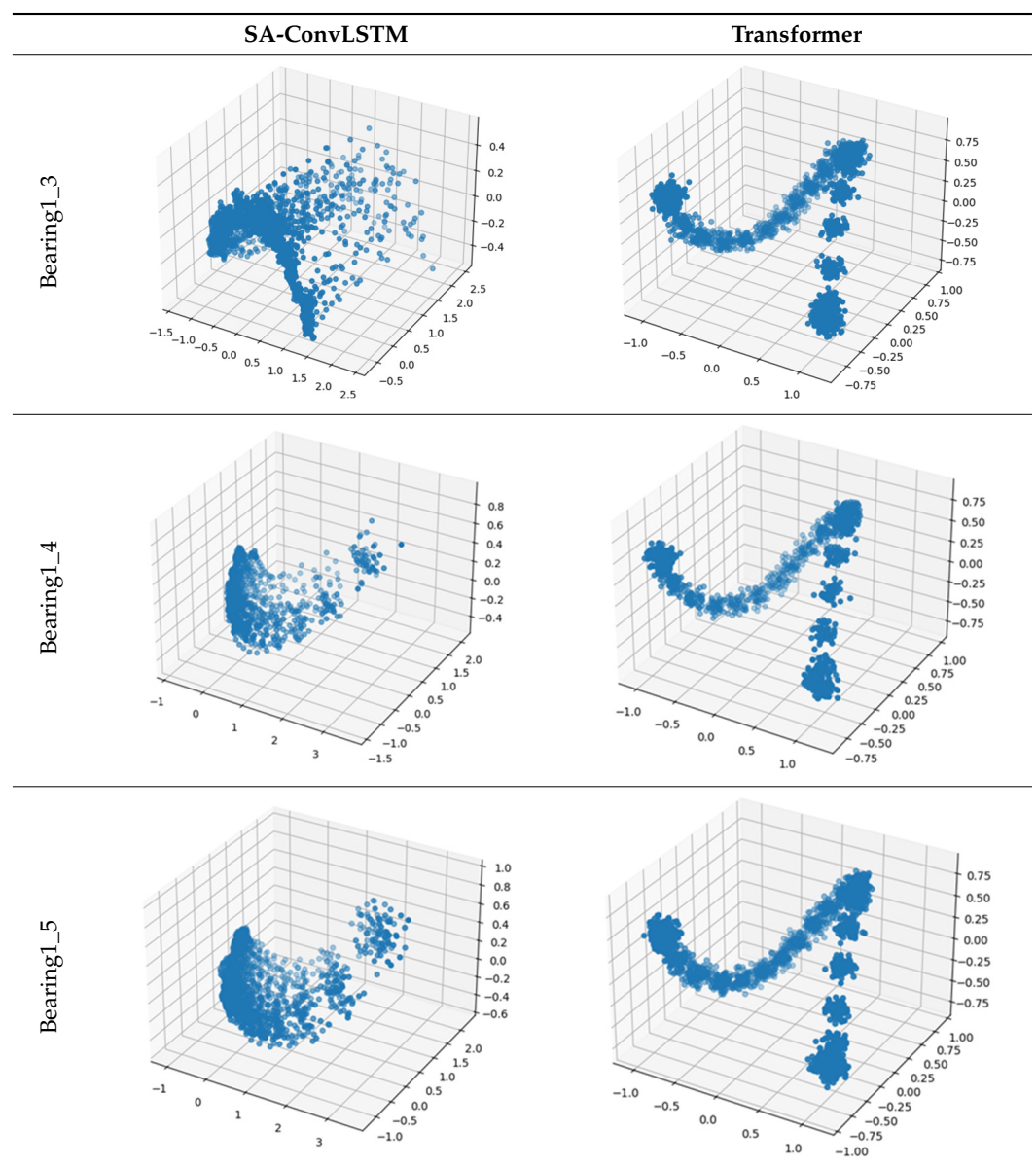
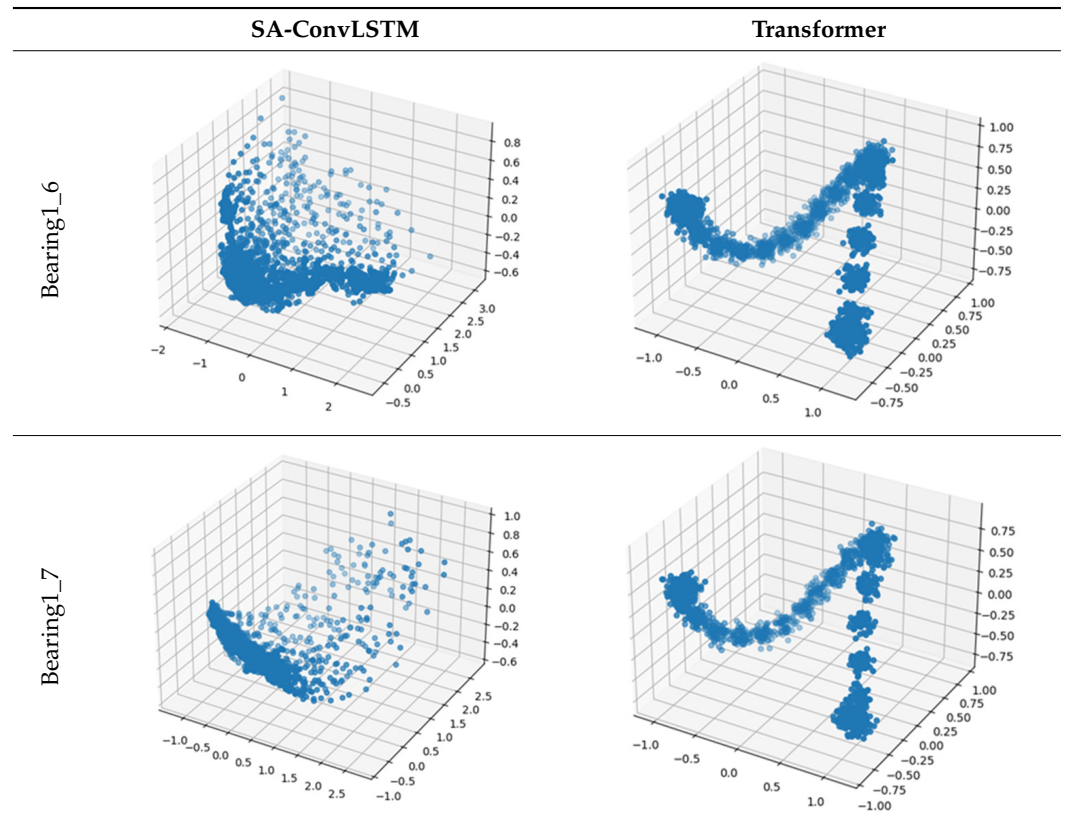


Table 9. Cont.



Finally, the proposed model was compared with the prediction results of several recent models, as shown in Table 10. Paper [37] used multiple manual features in the time domain, frequency domain, and time–frequency domain for RUL prediction. Paper [40] used ensemble empirical mode decomposition, PCA, and the local outlier factor algorithm to divide the bearing’s life degradation interval. Then, the HI obtained from the SE-ConvLSTM model was predicted using particle filtering and the double exponential model. Paper [31] used wavelet transform to extract time–frequency features and constructed training and testing datasets. The obtained HI was smoothed using the exponential moving average algorithm. It can be seen that almost all comparative methods require complex manual processing in the RUL prediction process [41]. In contrast, we have built an end-to-end prediction model directly from raw signals to RUL and achieved good results on most bearings. Additionally, the presence of the ECA module enhances the scalability of the model and allows for the fusion of deep learning models with different characteristics.

Table 10. Comparison results with recent methods.

Methods	Proposed		Method [37]		Method [40]		Method [41]	
	Score	Error	Score	Error	Score	Error	Score	Error
Bearing 1_3	0.66	12.20	0.22	43.31	0.31	33.86	0.85	4.53
Bearing 1_4	0.70	10.23	0.50	−5.04	0.14	−14.16	0.10	−16.26
Bearing 1_5	0.00	−476.90	0.52	18.75	0.56	16.77	0.05	−22.36
Bearing 1_6	0.98	0.48	0.16	−13.08	0.56	16.44	0.67	11.64
Bearing 1_7	0.71	−2.44	0.24	40.64	0.77	7.40	0.73	9.24
Ave		0.61		0.33		0.47		0.48

#### 4. Conclusions

Deep learning approaches have garnered a succession of outstanding outcomes in the realm of remaining useful life (RUL) forecasting, but the predictive performance of the

models greatly depends on expert knowledge and signal processing algorithms. Therefore, this paper proposed an end-to-end adaptive RUL prediction method for rolling bearings based on time–frequency image features to weaken the interference caused by human factors in the process of using deep learning models to predict the RUL. We used STFT to transform the raw vibration signal into the time–frequency domain while retaining early weak degradation signals and degradation mutation information in the vibration data as much as possible. In addition, we explored effective fusion methods for different deep learning models and introduced efficient channel attention mechanisms to adaptively assign attention to different channel features. Experimental results on the PRONOSTIA platform dataset show that the proposed model can effectively capture the state changes in the rolling bearing degradation process and demonstrate good predictive performance. By building a digital twin model of the bearing, the operating state of the bearing can be simulated in real time, and potential failure risks can be predicted. The digital twin technology provides an accurate and efficient method for bearing fault diagnosis, which helps to detect hidden faults in advance, reduce maintenance costs, and improve equipment operation efficiency.

However, during the analysis of the model’s prediction results, we identified some areas for improvement. Our future work will focus on incorporating automatic classification of bearing degradation patterns into the model in order to improve the predictive accuracy of the model by applying different treatment methods for different degradation patterns. Additionally, as there are significant fluctuations in the RUL predictions for rolling bearings, reasonable confidence intervals and confidence levels can also be combined into the model.

**Author Contributions:** Conceptualization, L.C. and H.W.; methodology, L.C. and H.W.; software, H.W.; validation, H.W., L.M., Z.X. and L.X.; formal analysis, L.C., H.W., L.M. and L.X.; investigation, M.R., L.C., H.W. and L.X.; resources, L.C. and H.W.; data curation, L.C. and H.W.; writing—original draft preparation, H.W.; writing—review and editing, L.C., L.M., Z.X. and L.X.; visualization, H.W. and Z.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Defense Industrial Technology Development Program of China (No. JCKY2021205B003).

**Data Availability Statement:** These data have been provided by the co-author of this manuscript, Hao Wang. Please contact him for any inquiries regarding their use.

**Conflicts of Interest:** The author, Hao Wang, studied at Dalian University of Technology. Authors Liang Chen and Linshu Meng were employed by The Aviation Industry Corporation of China, Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

## References

1. Huang, C.; Huang, H.Z.; Li, Y.F.; Peng, W.W. A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing. *J. Manuf. Syst.* **2021**, *61*, 757–772. [[CrossRef](#)]
2. Ghods, A.; Lee, H.H. Probabilistic frequency-domain discrete wavelet transform for better detection of bearing faults in induction motors. *Neurocomputing* **2016**, *188*, 206–216. [[CrossRef](#)]
3. Chen, D.L.; Qin, Y.; Wang, Y.; Zhou, J.H. Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction. *ISA Trans.* **2021**, *114*, 44–56. [[CrossRef](#)] [[PubMed](#)]
4. Uckun, S.; Goebel, K.; Lucas, P.J.F. Standardizing research methods for prognostics. In Proceedings of the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–10.
5. Glowacz, A. Acoustic fault analysis of three commutator motors. *Mech. Syst. Signal Process.* **2019**, *133*, 106226. [[CrossRef](#)]
6. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [[CrossRef](#)]
7. Zhu, J.; Chen, N.; Shen, C. A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. *Mech. Syst. Signal Process.* **2020**, *139*, 106602. [[CrossRef](#)]
8. Yang, X.; Zheng, Y.; Zhang, Y.; Wong, D.S.; Yang, W. Bearing Remaining Useful Life Prediction Based on Regression Shapaleet and Graph Neural Network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3505712. [[CrossRef](#)]



9. Cubillo, A.; Perinpanayagam, S.; Esperon-Miguez, M. A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. *Adv. Mech. Eng.* **2016**, *8*, 1687814016664660. [[CrossRef](#)]
10. Si, X.S.; Wang, W.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation—a review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [[CrossRef](#)]
11. Hu, T.; Guo, Y.; Gu, L.; Zhou, Y.; Zhang, Z.; Zhou, Z. Remaining useful life prediction of bearings under different working conditions using a deep feature disentanglement based transfer learning method. *Reliab. Eng. Syst. Saf.* **2022**, *219*, 108265. [[CrossRef](#)]
12. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors* **2017**, *17*, 273. [[CrossRef](#)] [[PubMed](#)]
13. Cao, Y.; Ding, Y.; Jia, M.; Tian, R. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107813. [[CrossRef](#)]
14. Chen, Y.; Peng, G.; Zhu, Z.; Li, S. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Appl. Soft Comput.* **2020**, *86*, 105919. [[CrossRef](#)]
15. Suh, S.; Lukowicz, P.; Lee, Y.O. Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks. *Knowl.-Based Syst.* **2022**, *237*, 107866. [[CrossRef](#)]
16. Liao, L.; Köttig, F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Trans. Reliab.* **2014**, *63*, 191–207. [[CrossRef](#)]
17. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [[CrossRef](#)]
18. Yan, M.; Wang, X.; Wang, B.; Chang, M.; Muhammad, I. Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. *ISA Trans.* **2020**, *98*, 471–482. [[CrossRef](#)]
19. Ali, J.B.; Chebel-Morello, B.; Saidi, L.; Malinowski, S.; Fnaiech, F. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mech. Syst. Signal Process.* **2015**, *56*, 150–172.
20. Aye, S.A.; Heyns, P.S. An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission. *Mech. Syst. Signal Process.* **2017**, *84*, 485–498. [[CrossRef](#)]
21. Wang, W.; Lei, Y.; Yan, T.; Li, N.; Nandi, A. Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification. *J. Dyn. Monit. Diagn.* **2022**, *1*, 2–8. [[CrossRef](#)]
22. Shen, Y.; Tang, B.; Li, B.; Tan, Q.; Wu, Y. Remaining useful life prediction of rolling bearing based on multi-head attention embedded Bi-LSTM network. *Measurement* **2022**, *202*, 111803. [[CrossRef](#)]
23. Que, Z.; Jin, X.; Xu, Z. Remaining useful life prediction for bearings based on a gated recurrent unit. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3511411. [[CrossRef](#)]
24. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*, 2018; arXiv:1803.01271.
25. Hewage, P.; Behera, A.; Trovati, M.; Pereira, E.; Ghahremani, M.; Palmieri, F.; Liu, Y. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Comput.* **2020**, *24*, 16453–16482. [[CrossRef](#)]
26. Chen, Q.; Liu, Y.B.; Ge, M.F.; Liu, J.; Wang, L. A Novel Bayesian-Optimization-Based Adversarial TCN for RUL Prediction of Bearings. *IEEE Sens. J.* **2022**, *22*, 20968–20977. [[CrossRef](#)]
27. Chang, Y.; Li, F.; Chen, J.; Liu, Y.; Li, Z. Efficient temporal flow Transformer accompanied with multi-head probsparse self-attention mechanism for remaining useful life prognostics. *Reliab. Eng. Syst. Saf.* **2022**, *226*, 108701. [[CrossRef](#)]
28. Yoo, Y.; Baek, J.G. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. *Appl. Sci.* **2018**, *8*, 1102. [[CrossRef](#)]
29. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.c. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
30. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
31. Huang, D.; Yu, G.; Zhang, J.; Tang, J.; Su, J. An Accurate Prediction Algorithm of RUL for Bearings: Time-Frequency Analysis Based on MRCNN. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 2222802. [[CrossRef](#)]
32. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Houlsby, N. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv*, 2020; arXiv:2010.11929.
33. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
34. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebel-Morello, B.; Zerhouni, N.; Varnier, C. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In Proceedings of the IEEE International Conference on Prognostics and Health Management, PHM'12, Denver, CO, USA, 18–21 June 2012; IEEE Catalog Number: CPF12PHM-CDR; pp. 1–8.
35. Soualhi, A.; Medjaher, K.; Zerhouni, N. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Trans. Instrum. Meas.* **2014**, *64*, 52–62. [[CrossRef](#)]

36. Singleton, R.K.; Strangas, E.G.; Aviyente, S. Extended Kalman filtering for remaining-useful-life estimation of bearings. *IEEE Trans. Ind. Electron.* **2014**, *62*, 1781–1790. [[CrossRef](#)]
37. Rathore, M.S.; Harsha, S.P. An attention-based stacked BiLSTM framework for predicting remaining useful life of rolling bearings. *Appl. Soft Comput.* **2022**, *131*, 109765. [[CrossRef](#)]
38. Jiang, C.; Liu, X.; Liu, Y.; Xie, M.; Liang, C.; Wang, Q. A Method for Predicting the Remaining Life of Rolling Bearings Based on Multi-Scale Feature Extraction and Attention Mechanism. *Electronics* **2022**, *11*, 3616. [[CrossRef](#)]
39. Jiang, G.J.; Yang, J.S.; Cheng, T.C.; Sun, H.H. Remaining useful life prediction of rolling bearings based on Bayesian neural network and uncertainty quantification. *Qual. Reliab. Eng. Int.* **2023**, *39*, 1756–1774. [[CrossRef](#)]
40. Yang, S.; Liu, Y.; Liao, Y.; Su, K. A New Method of Bearing Remaining Useful Life Based on Life Evolution and SE-ConvLSTM Neural Network. *Machines* **2022**, *10*, 639. [[CrossRef](#)]
41. Zhong, D.; Liu, N.; Yang, L.; Lin, L.; Chen, H. Self-Attention Convolutional Long Short-Term Memory for Short-Term Arctic Sea Ice Motion Prediction Using Advanced Microwave Scanning Radiometer Earth Observing System 36.5 GHz Data. *Remote Sens.* **2023**, *15*, 5437. [[CrossRef](#)]

**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.