

Article

Research on Remaining Useful Life Prediction Method of Rolling Bearing Based on Digital Twin

Rui Zhang , Zhiqiang Zeng, Yanfeng Li , Jiahao Liu and Zhijian Wang *

School of Mechanical Engineering, North University of China, Taiyuan 030051, China

* Correspondence: wangzhijian1013@163.com

Abstract: Bearing is a key part of rotating machinery. Accurate prediction of bearing life can avoid serious failures. To address the current problem of low accuracy and poor predictability of bearing life prediction, a bearing life prediction method based on digital twins is proposed. Firstly, the vibration signals of rolling bearings are collected, and the time-domain and frequency-domain features of the actual data set are extracted to construct the feature matrix. Then unsupervised classification and feature selection are carried out by improving the self-organizing feature mapping method. Using sensitive features to construct a twin dataset framework and using the integrated learning CatBoost method to supplement the missing data sets, a complete digital twin dataset is formed. Secondly, important information is extracted through macro and micro attention mechanisms to achieve weight amplification. The life prediction of rolling bearing is realized by using fusion features. Finally, the proposed method is verified by experiments. The experimental results show that this method can predict the bearing life with a limited amount of measured data, which is superior to other prediction methods and can provide a new idea for the health prediction and management of mechanical components.

Keywords: digital twins; prediction of remaining useful life; neural network



Citation: Zhang, R.; Zeng, Z.; Li, Y.; Liu, J.; Wang, Z. Research on Remaining Useful Life Prediction Method of Rolling Bearing Based on Digital Twin. *Entropy* **2022**, *24*, 1578. <https://doi.org/10.3390/e24111578>

Academic Editors: Claude Delpha and Demba Diallo

Received: 4 October 2022

Accepted: 24 October 2022

Published: 31 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Rolling bearing is a key component in rotating machinery, which has been widely used in modern industry [1,2]. Accurate remaining useful life estimation of bearings can significantly improve the reliability of mechanical systems, which can avoid serious failures and reduce maintenance costs. In recent years, the prediction of bearing remaining useful life (RUL) based on deep learning has made great progress [3–5].

Among the methods commonly used for bearing life prediction, there are often physical models and data-based methods. Physical models need to be built taking into account the influence of the complex surrounding environment and usually exhibit a weak generalization capability. Data-driven approaches, however, avoid the need for detailed modeling of complex environments and have better generalization capabilities by building models such as statistical extrapolation from historical data, and are one of the most popular research directions in the field of health prediction recently. Data-based methods often require manual extraction of features, construction of a health indicator (HI), determination of the health stage (HS), and determination of the first predicting time (FPT) before the final life expectancy can be predicted [6–8].

Machine learning, as a typical data-driven approach, constructs approximate models to approximate the real situation and build predictive models based on real-time, historical, and relational data. On this basis, Wang et al. proposed a residual service life prediction method for rolling bearings based on PCA and multi-dimensional feature fusion, aiming at the low reliability of bearing single feature characterization [9]. The life information of rolling bearing is characterized from many aspects, and the prediction result of residual life is more accurate and reliable. Chen et al. aimed at the problem that it is difficult to predict

the bearing life under the action of a single horizontal stress [10]. A bearing life prediction method based on the failure physical reliability model is proposed to predict the bearing degradation data. Jiang et al. proposed a new dual residual attention network [11]. The hybrid extended convolutional neural network is used to learn useful features from both time and frequency directions. It provides a reliable prediction for the remaining service life of the bearing. Ahmad et al. used the adaptive prediction model based on regression to learn the evolution trend of bearing health indicators [12]. Realize accurate prediction of the remaining life of the bearing. Xu et al. proposed a hybrid model of expandable service life based on continuous monitoring and bearing condition classification [13]. The feasible parameters of bearing state quantification are evaluated, which provides an intuitive reference for the prediction of the residual life of bearings. Pan et al. proposed a two-stage prediction method for the remaining service life of bearings [14], which divided the bearing degradation into the normal stage and the degradation stage. By constructing a multivariable feedback extreme value learning machine model, the rapid prediction of the remaining useful life of the bearing is realized.

In recent years, digital twin technology has developed rapidly and has been applied in a number of practical projects [15–17]. Tao et al. proposed the concept of the digital twin workshop [18] and explored the application of the five-dimensional model in several fields in conjunction with practical applications. Xie et al. proposed an adaptive development environment for automotive systems based on digital twins [19], which overcomes the problems of long development cycles and poor scalability in the production process. Wei et al. proposed an optimal deployment strategy using digital twins to fully exploit the advantages of digital twins and perform tool life prediction in response to the shortcomings of current manufacturing systems [20]. Xia et al. proposed a fault diagnosis framework based on digital twins in response to the lack of fault data [21] and pre-trained the conditional data generated by digital twins to achieve accurate fault prediction. Liu et al. addressed time-varying error prediction and compensation for CNC machine tools [22] by establishing a heat transfer model for tool spindles and visualizing time-varying error models. The performance of digital twins in predicting the performance of machine tools was explored. In addition to this, digital twins have been widely used in some industrial production [23].

This paper combines digital twin with bearing life prediction and proposes a bearing residual life prediction method driven by macro and micro attention bi-directional long short-term memory (MMA-BiLSTM). Signal features are extracted from actual signals, a feature matrix is constructed, and feature selection is carried out by improving the self-organizing feature mapping method. The twin dataset framework is constructed by using sensitive features, and the missing data set is supplemented by the integrated learning CatBoost method to form a complete digital twin dataset. A new database is built to provide a qualitative analysis basis for the prediction of bearing residual life. The MMA-BiLSTM model is used for training to obtain the final residual life prediction results. The main contributions of this paper are as follows.

- (1) An improved self-organizing feature mapping method is proposed, which can achieve automatic extraction of sensitive features by calculating the corresponding probability density interval of feature values;
- (2) A twin data construction method is proposed to use sensitive data in the original data as a digital twin framework, and CatBoost is used to learn the remaining features and generate new digital twins;
- (3) An MMA-BiLSTM neural network is proposed to extract important information through macro and micro attention mechanisms to achieve weight amplification and improve the accuracy of remaining useful life prediction.

Section 2 introduces the relevant background and theory in detail. The digital twin method and MMA-BiLSTM network proposed in this paper are given in Section 3. Finally, the effectiveness of the proposed method in this paper is demonstrated by experiments in Section 4.

2. Related Work

2.1. Self-Organizing Mapping

Self-organizing mapping (SOM) is an unsupervised learning method that can be clustered and visualized in high dimensions. On the basis of unsupervised advantage, this method can also provide the change of feature weight in the connection layer after classification. By iteratively updating the feature weight in the network structure, you can easily observe the change in the feature weight, obtain the sensitive features of the dataset classification, and provide the classification basis. SOM converts the input data into discrete low-dimensional data, which is then represented as active points in local areas or networks. After the initialization step is completed, the following three important learning processes are competition, collaboration, and adaptation.

In SOM, each neuron of the competition layer is connected according to the input N -dimensional feature vector (x) and weight (w). The range of w is between $(0, 1)$ and is initialized with any normalized value. In the learning process, calculate the distance between the feature vector x and the weight w of all neurons. When the distance is the smallest, the neuron becomes the optimal solution, which is the process of competition.

The cooperative process is that only the optimal solution of the competitive process and its neighboring neurons learn from the provided input data. In order to form a map more sensitively for similar features in the competitive hierarchy, the "optimal" neuron determines the adjacent neurons according to a fixed function, and the corresponding weight of this neuron will be updated.

The adaptive process refers to the adaptive activation function, which makes the optimal neuron and neighboring neurons more sensitive to specific input values, and also updates the corresponding weights. Through this process, the neurons adjacent to the optimal neuron will be more adaptive than those far away. The size of adaptation is controlled by the learning rate, which decreases with the learning time, and plays a role in reducing the convergence rate of SOM.

The algorithm flow of SOM is as follows:

- (1) Initialization weight w ; set a large initial neighborhood and set the number of network cycles;
- (2) Give an eigenvector X_k : $X_k = \{X_{1k}, X_{2k}, X_{3k} \cdots X_{nk}\}$;
- (3) Calculate the distance d_{jk} between the feature vector X_k and the output neuron, when d_{jk} takes the minimum value, c is the optimal neuron, i.e., $x_k - W_c = \min_j \{d_{jk}\}$;
- (4) Update the connection weights $w_{ij}(t+1) = w_{ij}(t) + \eta(t)\{x_i - w_{ij}(t)\}$ of c and its domain nodes where $0 < \eta(t) < 1$ is a gain function that decreases with time;
- (5) Input another feature vector into the network and return to step (3) until all the feature vectors are traversed;
- (6) Return to step (2) by making $t = t + 1$, until $t = T$.

2.2. Catboost

The CatBoost algorithm is a model based on the decision tree. It does not need a large number of samples as the training data and can adapt to the training of small-scale samples and high-precision diagnosis. The CatBoost algorithm belongs to the Boosting algorithm family and is a new machine learning algorithm framework based on a gradient boosting decision tree (GBDT). The GBDT algorithm is an algorithm for regression and classification proposed by Friedman in 2000 [24], which can avoid the problem of overfitting a single decision tree due to the internal integration of multiple decision trees and the accumulation of multiple decision trees. The GBDT algorithm constructs a learner to reduce the loss along the steepest direction of the gradient at each iteration step to make up for the shortcomings of the current model.

In CatBoost, the target statistics (TS) method is usually used to process categorical features target statistic (TS) method refers to replacing category features with calculated val-

ues. Representing the i th category feature of the k th training sample as x_k^i , and representing the replacement target value as y . The expression of the TS method is:

$$\hat{x}_k^i = E(y | x^i = x_k^i) \quad (1)$$

The commonly used TS value calculation method is the Greedy TS method, which can be smoothed by using the average of the target variable y of the same category x_k^i in training samples and using a prior probability p , expressed as:

$$\hat{x}_k^i = \frac{\sum_{j=1}^n II_{\{x_j^i=x_k^i\}} \cdot y_j + ap}{\sum_{j=1}^n II_{\{x_j^i=x_k^i\}} + a} \quad (2)$$

However, due to the duplication in the use of the training set and test set, this method will lead to condition deviation and result in overfitting. Based on this situation, CatBoost uses a method to improve the category feature processing and uses the sorting principle to solve the problem of condition shift and overfitting. In the Ordered TS, a random sequence σ is generated to number the data. The training data is selected according to the sorting principle: $D_k = \{X_j : \sigma(j) < \sigma(k)\}$, and the test set uses all the data: $D_k = D$. Then the divided data set is used to calculate the Greedy TS value x_k^i with a priori probability.

For multi-dimensional feature data sets, the relationship between the actual value of most features and the prediction is often nonlinear, which brings great difficulties to the analysis of feature changes. To solve this problem, the feature intersection approach is proposed, which combines different features to form new cross-features to fit the changing relationship of data set features. If all features of the dataset are crossed, the exponential dimension will grow exponentially, thus increasing the computational complexity. Based on this problem, CatBoost adopts a greedy strategy to deal with it and does not cross features at the previous node of the gradient lifting tree. Instead, the features divided in front of the node and the features within the node are considered as two groups of features to cross, and one pot vector method is used for feature fusion. The acquisition method of cross-features is expressed by the formula:

$$\hat{y} = b + w_1 \hat{x}_1 + w_2 \hat{x}_2 + w_3 \hat{x}_3 \quad (3)$$

Where x_3 represents the sum of features of feature set x_1 and x_2 , and w represents the weight relationship between feature sets. b is a constant term, which solves the problem of too many nonlinear fitting relations and cross-features. In the process of gradient promotion, different data generate different gradient classes. If the training data is repeated a lot, it will lead to gradient boosting overfitting and skew the prediction results. Based on this problem, CatBoost uses the Ordered method to sort the data sets, which reduces the error of gradient fitting. The method to generate the base evaluator is:

$$h_t = \arg \min_{\{h \in H\}} \frac{1}{n} \sum_{k=1}^n (-g^t(x_k, y_k) - h(x_k)) \quad (4)$$

Where h_t represents the generated basis evaluator and $-g^t(x_k, y_k)$ represents the negative gradient value of the loss function in the current gradient model. The problem of gradient lifting prediction migration is solved by the Ordered method.

CatBoost uses a symmetric tree structure in the decision tree structure. The advantage of the symmetric tree structure is that it is not easy to overfit, and it is much faster than the gradient lifting algorithm, such as XGBoost. In addition, the CatBoost algorithm can realize multiple graphics processing unit (GPU) operations. The distributed learning tree enables CatBoost to perform parallel computing, thus improving its overall computing speed.

2.3. BiLSTM

LSTM network is an improvement in recurrent neural network (RNN). The key to the LSTM network is the cell state. The information in the cell state is updated and deleted through the forgetting gate, update gate, and output gate. The structure of LSTM is shown in Figure 1. The following is the representation of the three gates of LSTM:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \tag{7}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{8}$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t \times \tanh(C_t) \tag{10}$$

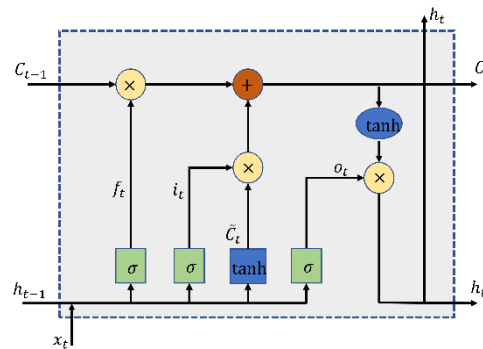


Figure 1. Structure of LSTM.

Because the next moment prediction output of unidirectional LSTM is only affected by the previous multiple time inputs, in many cases, the prediction will be affected by the previous and subsequent multiple time inputs at the same time. In order to fully extract the correlation between before and after features and obtain better prediction results. The BiLSTM network was introduced to calculate the front and back information from two opposite directions (forward network output is \vec{h}_t , backward network output is \overleftarrow{h}_t):

$$\vec{h}_t = \text{LSTM}(x_t, h_{t-1}) \tag{11}$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, h_{t+1}) \tag{12}$$

Finally, output the comprehensive result of the result stack of the forward network layer and the backward network layer.

3. Prediction of Bearing RUL Based on Digital Twin

Aiming at the complex working environment of special mechanical equipment, data collection is difficult, and the amount of data is small. This paper presents a method for predicting the remaining service life of small sample bearings based on data twin driving. First of all, feature extraction is carried out on the actual data to form a high-dimensional feature dataset. Then, an improved self-organizing feature mapping method (ISOFM) is used to select features, calculate the numerical probability density intervals of features corresponding to sensitive features, determine the optimal number of sensitive features, and form a feature framework. The feature framework is combined with existing data to form an interactive dataset with missing data, and CatBoost integrated learning algorithm is introduced. The missing eigenvalues are taken as the feature learning objectives of CatBoost, respectively, and their regression operation characteristics are used to complement the

interactive dataset, thus forming a complete twin dataset. Finally, the macro and micro attention mechanisms are combined with BiLSTM to form MMA-BiLSTM. The weight of MMA-BiLSTM is amplified in the whole time dimension and each time dimension to realize the residual life prediction of bearings.

3.1. ISOFM

The classical SOM method needs to preset the number of nodes in the output layer. Therefore, it is necessary to improve this method to make it have the ability to adaptively select the number of nodes in the output layer. In this paper, a generation-by-generation node processing method is proposed to remove nonsensitive features in the neighborhood during SOM’s selection of sensitive features. Both the number of output nodes of SOM can be adaptively obtained, and the nonsensitive feature removal strategy can also improve the feature selection efficiency of SOM. The proposed ISOFM method mainly selects nodes by introducing a learning rate α and a relative removal rate parameter β . Set the Euclid distance and the threshold value of the feature node weight to improve. The improved weight update formula is:

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t+1}) \tag{13}$$

$$W_j(t+1) = \lambda [W_j(t) + \alpha(t)(X_i - W_j(t))] \tag{14}$$

$$\lambda = \begin{cases} 0, & \frac{(1-\beta)[\max(d_j) - \min(d_j)]^2}{4\left(d_j - \sum_{j=1}^m d_j/J\right)^2} < 1; \\ 1, & \frac{(1-\beta)[\max(d_j) - \min(d_j)]^2}{4\left(d_j - \sum_{j=1}^m d_j/J\right)^2} \geq 1 \end{cases} \tag{15}$$

3.2. MMA-BiLSTM

The derivation of BiLSTM based on macro and micro attention mechanisms is as follows: macro and micro attention mechanisms refer to the operation of the attention mechanism on the whole time dimension of input data and data on each time dimension. Specifically, firstly, the data matrix generated by digital twins is processed, and its macro and micro attention coefficients are calculated using MMA. In the prediction process, the input dataset of the whole time dimension is $X_t = [x_1, x_2, \dots, x_t]$. Where $x_{t'} = [x_{t',1}, x_{t',2}, \dots, x_{t',n}]^T$ represents the input data at time t' , and the macro attention mechanism processes the data in the whole time dimension through the attention mechanism; the micro attention mechanism is to use the attention mechanism to process input data $x_{t'}$ in each time dimension [25].

The formula for calculating macro and micro attention coefficients:

$$\chi_{t'} = \frac{\exp(S(\bar{x}_{t'}, q_M))}{\sum_{j=1}^t \exp(S(\bar{x}_j, q_M))} \tag{16}$$

$$\alpha_{t',i} = \frac{\exp(s(x_{t',i}, q_{t',m}))}{\sum_{j=1}^n \exp(s(x_{t',j}, q_{t',m})) + \sum_{p=1}^m \exp(s(h_{t'-1,p}, q_{t',m}))} \tag{17}$$

Where $\alpha_{t',i}$ is the attention coefficient of input data in the micro attention mechanism. $\chi_{t'}$ is the macro attention coefficient obtained in the whole time dimension. $\bar{x}_{t'}$ is the mean value of $x_{t'}$, $x_{t',j}$ is the j element in the input data $x_{t'} = [x_{t',1}, x_{t',2}, \dots, x_{t',n}]^T$ at the time of t , t is the dimension of the input dataset $X_t = [x_1, x_2, \dots, x_t]$, \bar{x}_j is the mean value of the j vector in the input dataset X_t , and q is the query vector. In the MMA-BiLSTM

network training process, set the macro level query vector q_M and the micro level query vector q_m ; the relevant scoring function is calculated as follows:

$$S(\bar{x}_j, q_M) = \frac{\bar{x}_j q_M}{\sqrt{t}} \tag{18}$$

$$s(x_{t',j}, q_{t',m}) = \frac{x_{t',j}^T q_{t',m}}{\sqrt{n+m}} \tag{19}$$

Where n is the dimension of input data $\mathbf{x}_{t'} = [x_{t',1} \ x_{t',2} \ \dots \ x_{t',n}]^T$ at time t .

According to the corresponding macro and micro attention coefficients, the associated input data weights and recursive data weights are magnified at multiple levels.

$$\mathbf{w}'_{t',ix} = (1 + \chi_{t'}) \times \mathbf{w}_{t',ix} (\alpha_{t'} + 1) \tag{20}$$

$$\mathbf{w}'_{t',ox} = (1 + \chi_{t'}) \times \mathbf{w}_{t',ox} (\alpha_{t'} + 1) \tag{21}$$

$$\mathbf{w}'_{t',fx} = (1 + \chi_{t'}) \times \mathbf{w}_{t',fx} (\alpha_{t'} + 1) \tag{22}$$

Wherein, $\mathbf{w}'_{t',ix}$ represents the weight between the input data of the BiLSTM neural network and the input gate in the hidden layer, $\mathbf{w}'_{t',ox}$ represents the weight between the input data of the BiLSTM neural network and the output gate in the hidden layer, $\mathbf{w}'_{t',fx}$ represents the weight between the input data of the LSTM neural network and the forgetting gate in the hidden layer, $\mathbf{w}'_{t',ix}$ represents the weight between the input data of the MMA-BiLSTM neural network and the input gate in the hidden layer, $\mathbf{w}'_{t',ox}$ represents the weight between the input data of MMA-BiLSTM neural network and the output gate in the hidden layer, and $\mathbf{w}'_{t',fx}$ represents the weight between the input data of MMA-BiLSTM neural network and the forgetting gate in the hidden layer.

According to the amplification of input data weight and recursive number weight, the corresponding calculation results are obtained:

$$f_{t'} = \sigma(w'_{t',f}[x_{t'}, h_{t'-1}] + b_f) \tag{23}$$

$$i_{t'} = \sigma(w'_{t',i}[x_{t'}, h_{t'-1}] + b_i) \tag{24}$$

$$\tilde{c}_{t'} = \tanh(w_{t',c}[x_{t'}, h_{t'-1}] + b_c) \tag{25}$$

$$c_{t'} = f_{t'} \times c_{t'-1} + i_{t'} \times \tilde{c}_{t'} \tag{26}$$

$$o_{t'} = \sigma(w'_{t',o}[x_{t'}, h_{t'-1}] + b_o) \tag{27}$$

$$h_{t'} = o_{t'} \times \tanh(c_{t'}) \tag{28}$$

$$\vec{h}_{t'} = \text{LSTM}(x_{t'}, \vec{h}_{t'-1}) \tag{29}$$

$$\overleftarrow{h}_{t'} = \text{LSTM}(x_{t'}, \overleftarrow{h}_{t'+1}) \tag{30}$$

$$F_{t'} = g\left(W_{t',h_y} \vec{h}_{t'} + W_{t',h_y} \overleftarrow{h}_{t'} + b_y\right) \tag{31}$$

Wherein, σ is sigmoid activation function, g is linear activation function, b_i is MMA-BiLSTM hidden layer input gate offset term, b_f is MMA-BiLSTM hidden layer forgetting gate offset term, b_c is MMA-BiLSTM hidden layer storage cell unit offset term, b_o is MMA-BiLSTM hidden layer output gate offset term, b_y is MMA-BiLSTM output layer offset term, $i_{t'}$ is the input gate output at t' time, $f_{t'}$ is the forgetting gate output at t' time, $c_{t'}$ is the storage cell unit output at t' time, and $F_{t'}$ is the output layer output at t' time.

The bearing vibration data are collected separately as samples, and the data samples are twin expanded. Finally, different machine learning methods are used to compare the

prediction accuracy between the original sample and the interactive dataset. The specific process and structure of the proposed method can be shown in Figure 2. The specific steps of the proposed method are shown as follows:

- (1) Set up a test platform to collect vibration signals of bearings from normal operation to fault status;
- (2) Extraction of time-domain and frequency-domain features of vibration signals from the original signal;
- (3) Use ISOFM to determine the number of sensitive features and select features from the acquired feature data set, and extract the main features in the feature set that can determine the signal category;
- (4) The probability density distribution models of sensitive features in feature data sets are constructed, respectively; determine the feature frame and the selection range of its feature values;
- (5) The feature data frame generated is combined with the feature data set extracted from the initial samples interactively, and the data at the nonsensitive features are represented by missing values;
- (6) The CatBoost regression algorithm is used to fill in the missing values in the interactive dataset containing missing values. Sorted according to importance, the missing value is used as the prediction target to fill the characteristic value. During the filling process, the missing values of other features are filled with the feature mean value;
- (7) An interactive dataset with a complete data structure is obtained, i.e., a twin feature dataset that expresses vibration signal fault information obtained from a small amount of data. The dataset is normalized to fit the health indicators of the bearing;
- (8) Use the first k health indicators of the bearing as network input to predict the health value at moment $k + 1$;
- (9) Repeat step 8 a certain number of times, and when these output values are less than 0, the inverse normalization of the sampled points results in RUL.

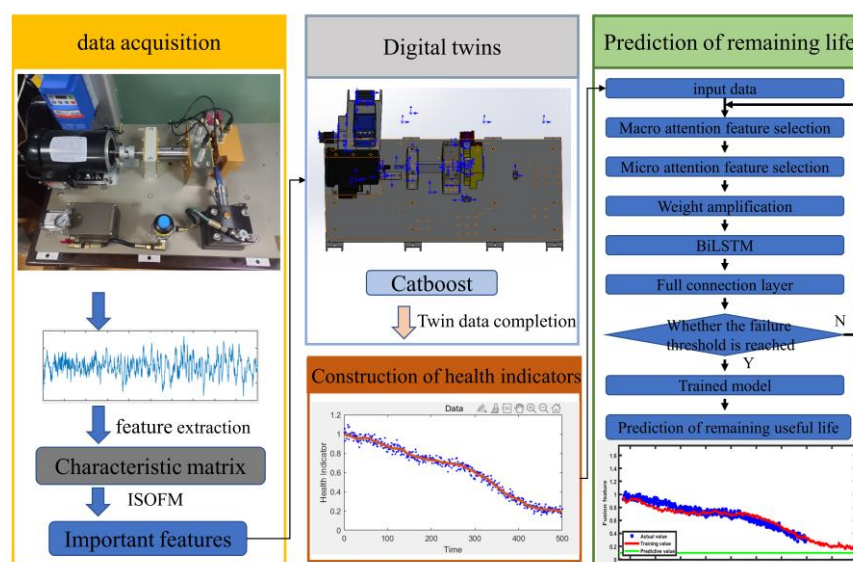


Figure 2. Flowchart of the proposed method.

4. Experimental Validation

4.1. Experimental Description

The experimental design was carried out according to the research idea shown in Figure 2, and tests were conducted on a full life-bearing fatigue test machine. The main structure of the platform is the motor, supporting bearing housing, vibration sensor, hydraulic resistor, coupling, and other mechanisms. The experimental setup is shown in Figure 3. The tests were carried out at different rotational speed conditions, and the test was

set up with a sampling frequency of 25.6 kHz, a sampling interval of 1 min, and a duration of 1.28 s per sample. The bearing vibration signal is shown in Figure 4. The experimental data are described in Table 1.

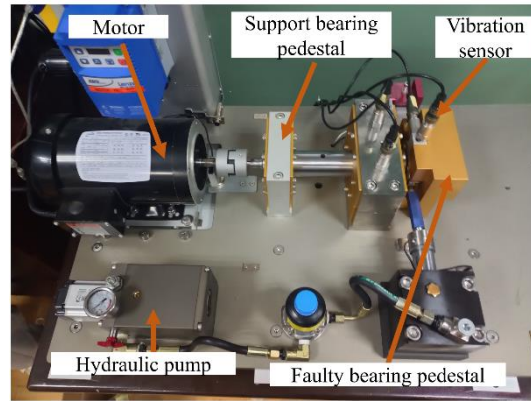


Figure 3. Bearing life test bench.

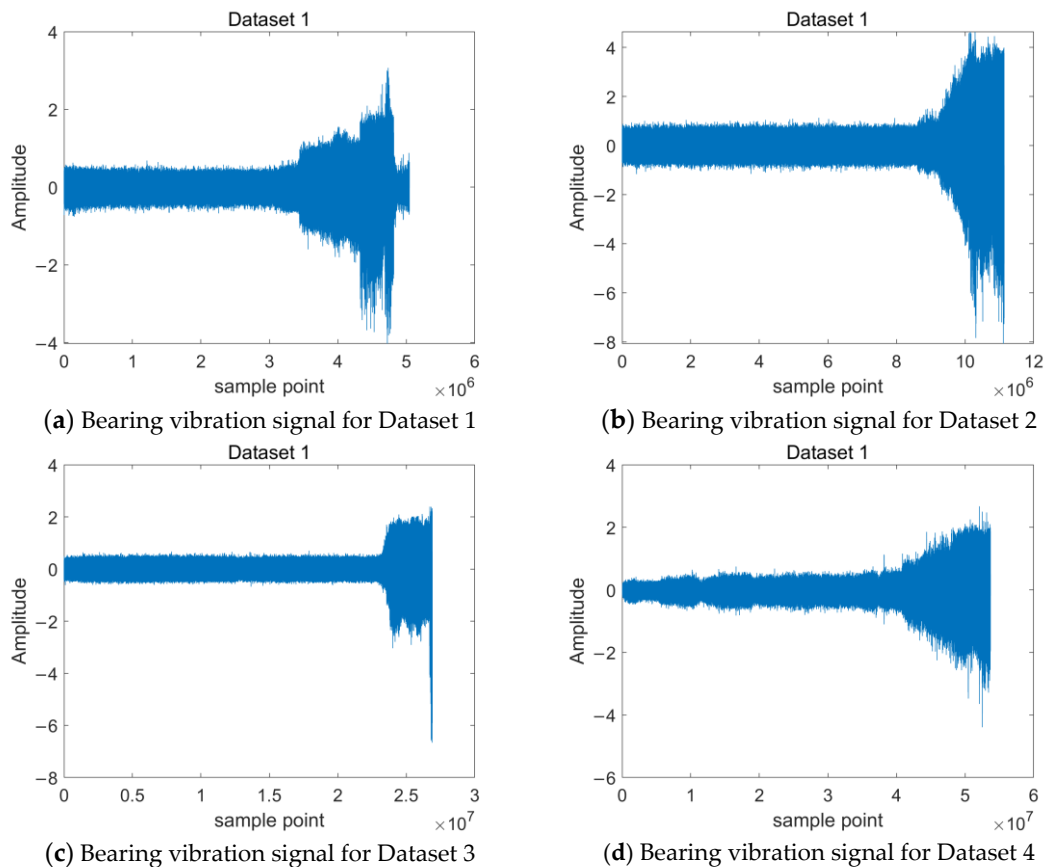


Figure 4. Bearing vibration signal.

Table 1. Description of experimental data.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Load (kg)	500	1000	1000	1000
Speed (rpm)	1200	2100	2100	2100
Time (min)	154	340	821	1639

Corresponding to the data sets in Table 1, the signal features for the different operating conditions were obtained by pre-processing. These include 13 time-domain features and 16 frequency-domain features, which are combined into a feature matrix. The details are as follows: (1) maximum value, (2) minimum value, (3) median, (4) mean, (5) peak difference, (6) mean of absolute values, (7) variance, (8) standard deviation, (9) cliffness, (10) skewness, (11) root mean square, (12) impulse factor, (13) margin factor, (14) amplitude maximum, (15) amplitude minimum, (16) amplitude median, (17) amplitude mean, (18) amplitude peak difference, (19) amplitude peak threshold, threshold of 75% of the amplitude peak difference, (20) amplitude peak, (21) amplitude peak corresponding frequency, (22) frequency center of gravity, (23) mean square frequency (24) frequency variance, (25) frequency standard deviation, (26) short time power spectral density, (27) spectral entropy, (28) fundamental frequency, and (29) resonance peak. The initial feature dataset was processed using ISOFM. Five groups of sensitive features were finally retained adaptively by the algorithm, and the feature importance is shown in Figure 5. In the experiment, the learning rate and removal rate are set to $a = 0.1$ and $b = 0.1$, respectively, and the number of iterations of the algorithm is set to 100.

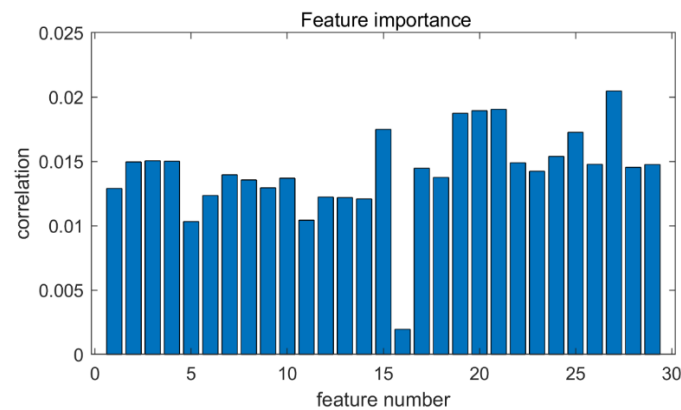


Figure 5. Feature importance diagram.

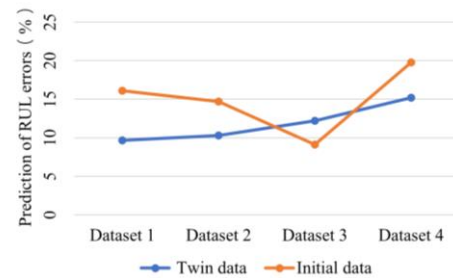
In constructing the digital twin data feature framework, the optimal number of sensitive features is first determined, and the feature framework is formed. The feature framework is combined with existing data to form an interactive dataset with missing data. The CatBoost integrated learning algorithm is introduced, and the missing feature values are used as the feature learning targets for CatBoost, respectively. The interaction dataset is complemented by using its regression operation properties to form a complete twin dataset. In this paper, the rolling bearing vibration signals are used as samples, the data samples are twinned and expanded, and finally, the prediction errors of the original samples are compared with the interaction dataset.

4.2. Comparison of Digital Twin Data with Initial Data

After generating the digital twin interaction dataset, the initial data and the digital twin-generated data were used separately for bearing life prediction. The dataset was validated for prediction by an LSTM network, and the first 80% of the normalized sample points were used as the training set for predicting the RUL. The current sample points and true RULs for Dataset 1, Dataset 2, Dataset 3, and Dataset 4 are shown in Table 2. The average error results obtained from multiple cross-validations of the different data are placed in Figure 6 for comparison.

Table 2. Comparison of digital twin data with initial data.

Data Set	Current Sample Point	Real RUL	Prediction of RUL Errors (%)	
			Initial Data	Twin Data
Dataset 1	123	31	16.12	9.67
Dataset 2	272	68	14.7	10.29
Dataset 3	657	164	9.14	12.19
Dataset 4	1311	328	19.8	15.2

**Figure 6.** Comparison of prediction errors between initial data and twin data.

It can be seen from Figure 6 that the prediction errors for the Dataset 1, Dataset 2, and Dataset 4 twin data are smaller than the original data. In Dataset 3, the original data prediction error is smaller, and the two are closer. The experimental results validate the effectiveness of the twin dataset. The twin dataset produced by constructing new feature data and fusing it with the initial dataset clearly has better predictive power than the initial dataset.

4.3. MMA-BiLSTM

To improve the prediction performance of LSTM networks. The macro-microscopic attention mechanism is combined with BiLSTM to propose MMA-BiLSTM. The method is compared with BiLSTM, LSTM, and GRU for experiments. The results are presented in Tables 3–6. where dataset a-b represents the ath dataset and the bth experiment. The experimental data are in seconds. The number of input units and output units in MMA-BiLSTM are set to 32 and 1, respectively, and the learning rate is set to 0.01. In this research, the number of hidden layer units is set to 128. the initialization method of the neural network uses standard initialization. The MMA-BiLSTN predictions for different data sets are shown in Figure 7. The mean absolute error (MAE) and root mean square error (RMSE) were used to evaluate the prediction effect. They are defined as follows.

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m (rul_i - rul_i^{\hat{}})^2 \quad (32)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (rul_i - rul_i^{\hat{}})^2} \quad (33)$$

From Tables 3–6, it can be concluded that compared with BiLSTM, LSTM, GRU, and MMA-BiLSTM has a smaller prediction error. This shows the advantages of data generated by digital twins. In Figure 8, MAE and RMSE predicted by the MMA-BiLSTM method are smaller than those predicted by other methods. It shows the suitable performance of MMA-BiLSTM in RUL prediction.

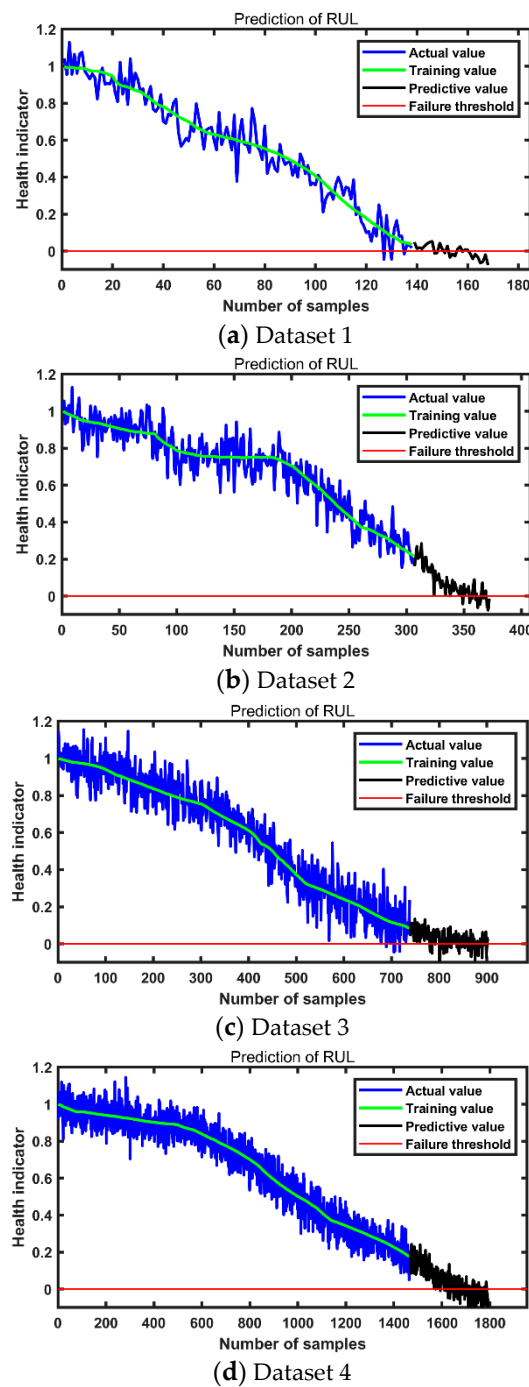


Figure 7. RUL of bearings predicted by MMA-BiLSTN.

Table 3. RUL prediction results and comparison for Dataset 1.

	MMA-BiLSTM	BiLSTM	LSTM	GRU
Dataset 1-1	145	164	158	231
Dataset 1-2	143	141	137	164
Dataset 1-3	152	139	156	169
Dataset 1-4	146	153	122	152
MAE	7.50	9.75	13.75	26.0
RMSE	8.21	11.12	18.25	39.55

Table 4. RUL prediction results and comparison for Dataset 2.

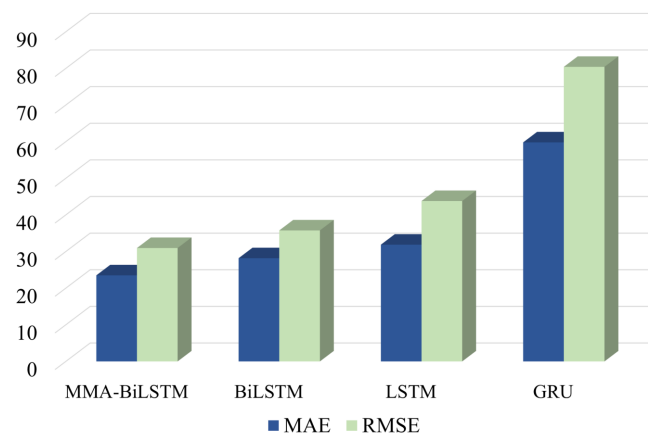
	MMA-BiLSTM	BiLSTM	LSTM	GRU
Dataset 2-1	327	357	364	390
Dataset 2-2	365	384	332	367
Dataset 2-3	350	363	359	471
Dataset 2-4	333	308	326	361
MAE	13.75	29.0	16.25	57.25
RMSE	15.35	30.73	17.29	72.16

Table 5. RUL prediction results and comparison for Dataset 3.

	MMA-BiLSTM	BiLSTM	LSTM	GRU
Dataset 3-1	820	827	780	941
Dataset 3-2	768	835	773	893
Dataset 3-3	795	758	826	754
Dataset 3-4	832	780	841	807
MAE	22.75	31.0	28.50	68.25
RMSE	30.03	38.75	33.21	77.89

Table 6. RUL prediction results and comparison for Dataset 4.

	MMA-BiLSTM	BiLSTM	LSTM	GRU
Dataset 4-1	1712	1679	1726	1738
Dataset 4-2	1691	1648	1694	1704
Dataset 4-3	1652	1712	1689	1670
Dataset 4-4	1678	1707	1746	1695
MAE	44.25	47.5	74.75	62.75
RMSE	49.30	53.93	78.33	67.31

**Figure 8.** Comparison of MAE and RMSE predicted by different models.

The prediction performance of the methods proposed in this paper has been improved, and the predicted values of RUL for bearings in different working conditions are closer to the actual values than other methods, which can provide an effective way to predict bearing life.

5. Conclusions

For the bearing life prediction problem, this paper proposes a bearing life prediction method combining digital twin and MMA-BiLSTM network. Firstly, an extracted sensitive feature matrix is constructed to build the digital twin framework; the data set is supplemented by the integrated learning CatBoost method for missing data to form a complete

digital twin data set. The MMA-BiLSTM network is proposed for life prediction. Finally, the accuracy of the proposed approach was verified by building a bearing life prediction test bench. The method can be further extended and applied to other condition parameters of gearboxes to provide data closer to the true value for predicting the RUL of machinery.

Author Contributions: Conceptualization, R.Z. and J.L.; methodology, Z.Z.; software, J.L.; validation, R.Z., Z.W. and Z.Z.; formal analysis, R.Z.; investigation, Z.W.; resources, J.L.; data curation, R.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Y.L.; visualization, Z.W.; project administration, Z.W.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Foundation of Shanxi Province [Grant No. 201601D102025], National Natural Science Foundation of China [Grant No. 52275139, 51905496], Opening Project of Shanxi Key Laboratory of Advanced Manufacturing Technology [Grant No. XJZZ201902, XJZZ202102], Shanxi Basic Research Program [Grant No. 202103021224205, 202103021223195, 202102020101015, 202103021224199], Patent transformation special plan project [Grant No. 202201051], Central government guides local special projects [Grant No. YDZJSX2022A029].

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the anonymous reviewers for their invaluable and constructive comments.

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

References

- Li, X.; Li, X.; Ma, H. Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. *Mech. Syst. Signal. Process.* **2020**, *143*, 106825. [[CrossRef](#)]
- Yu, K.; Lin, T.R.; Ma, H.; Li, X.; Li, X. A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. *Mech. Syst. Signal. Process.* **2021**, *146*, 107043. [[CrossRef](#)]
- Wang, Z.J.; He, X.X.; Yang, B.; Li, N.P. Subdomain Adaptation Transfer Learning Network for Fault Diagnosis of Roller Bearings. *IEEE Trans. Ind. Electron.* **2022**, *69*, 8430–8439. [[CrossRef](#)]
- Wang, Z.J.; Zhou, J.; Du, W.H.; Lei, Y.G.; Wang, J.Y. Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution. *Mech. Syst. Sig. Process.* **2022**, *162*, 108018. [[CrossRef](#)]
- He, X.X.; Wang, Z.J.; Li, Y.F.; Khazhina, S.; Du, W.H.; Wang, J.Y.; Wang, W.Z. Joint decision-making of parallel machine scheduling restricted in job-machine release time and preventive maintenance with remaining useful life constraints. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108429. [[CrossRef](#)]
- Xu, L.; Pennacchi, P.; Chatterton, S. A new method for the estimation of bearing health state and remaining useful life based on the moving average cross-correlation of power spectral density. *Mech. Syst. Signal. Process.* **2020**, *139*, 106617. [[CrossRef](#)]
- Rezamand, M.; Kordestani, M.; Orchard, M.E.; Carriveau, R.; Ting, D.S.K.; Saif, M. Improved Remaining Useful Life Estimation of Wind Turbine Drivetrain Bearings Under Varying Operating Conditions. *IEEE Trans. Ind. Inf.* **2021**, *17*, 1742–1752. [[CrossRef](#)]
- Qin, Y.; Chen, D.; Xiang, S.; Zhu, C. Gated Dual Attention Unit Neural Networks for Remaining Useful Life Prediction of Rolling Bearings. *IEEE Trans. Ind. Inf.* **2021**, *17*, 6438–6447. [[CrossRef](#)]
- Wang, H.; Ni, G.; Chen, J.; Qu, J. Research on rolling bearing state health monitoring and life prediction based on PCA and Internet of things with multi-sensor. *Measurement* **2020**, *157*, 107657. [[CrossRef](#)]
- Chen, C.; Li, B.; Guo, J.; Liu, Z.; Qi, B.; Hua, C. Bearing life prediction method based on the improved FIDES reliability model. *Reliab. Eng. Syst. Saf.* **2022**, *227*, 108746. [[CrossRef](#)]
- Jiang, G.; Zhou, W.; Chen, Q.; He, Q.; Xie, P. Dual residual attention network for remaining useful life prediction of bearings. *Measurement* **2022**, *199*, 111424. [[CrossRef](#)]
- Ahmad, W.; Khan, S.A.; Kim, J.-M. A Hybrid Prognostics Technique for Rolling Element Bearings Using Adaptive Predictive Models. *IEEE Trans. Ind. Electron.* **2018**, *65*, 1577–1584. [[CrossRef](#)]
- Xu, G.; Hou, D.; Qi, H.; Bo, L. High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life. *Mech. Syst. Signal. Pract.* **2021**, *146*, 107050. [[CrossRef](#)]
- Pan, Z.; Meng, Z.; Chen, Z.; Gao, W.; Shi, Y. A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings. *Mech. Syst. Signal. Pract.* **2020**, *144*, 106899. [[CrossRef](#)]
- Moghadam, F.K.; Nejad, A.R. Online condition monitoring of floating wind turbines drivetrain by means of digital twin. *Mech. Syst. Signal. Pract.* **2022**, *162*, 108087. [[CrossRef](#)]
- Fahim, M.; Sharma, V.; Cao, T.V.; Canberk, B.; Duong, T.Q. Machine Learning-Based Digital Twin for Predictive Modeling in Wind Turbines. *IEEE Access* **2022**, *10*, 14184–14194. [[CrossRef](#)]

17. Xu, X.; Shen, B.; Ding, S.; Srivastava, G.; Bilal, M.; Khosravi, M.R.; Menon, V.G.; Jan, M.A.; Wang, M. Service Offloading With Deep Q-Network for Digital Twinning-Empowered Internet of Vehicles in Edge Computing. *IEEE Trans. Ind. Inf.* **2022**, *18*, 1414–1423. [[CrossRef](#)]
18. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inf.* **2019**, *15*, 2405–2415. [[CrossRef](#)]
19. Xie, G.; Yang, K.; Xu, C.; Li, R.; Hu, S. Digital Twinning Based Adaptive Development Environment for Automotive Cyber-Physical Systems. *IEEE Trans. Ind. Inf.* **2022**, *18*, 1387–1396. [[CrossRef](#)]
20. Wei, Y.; Hu, T.; Wang, Y.; Wei, S.; Luo, W. Implementation strategy of physical entity for manufacturing system digital twin. *Robot. Com-Int. Manuf.* **2022**, *73*, 102259. [[CrossRef](#)]
21. Xia, M.; Shao, H.; Williams, D.; Lu, S.; Shu, L.; de Silva, C.W. Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107938. [[CrossRef](#)]
22. Liu, K.; Song, L.; Han, W.; Cui, Y.; Wang, Y. Time-Varying Error Prediction and Compensation for Movement Axis of CNC Machine Tool Based on Digital Twin. *IEEE Trans. Ind. Inf.* **2022**, *18*, 109–118. [[CrossRef](#)]
23. Rassolkin, A.; Rjabtsov, V.; Kuts, V.; Vaimann, T.; Kallaste, A.; Asad, B.; Partyshev, A. Interface Development for Digital Twin of an Electric Motor Based on Empirical Performance Model. *IEEE Access* **2022**, *10*, 15635–15643. [[CrossRef](#)]
24. Friedman, J.H. Greedy Function Approximation_ A Gradient Boosting Machine. *Ann. Math. Stat.* **2001**, *29*, 1189–1232.
25. Qin, Y.; Xiang, S.; Chai, Y.; Chen, H. Macroscopic–Microscopic Attention in LSTM Networks Based on Fusion Features for Gear Remaining Life Prediction. *IEEE Trans. Ind. Electron.* **2020**, *67*, 10865–10875. [[CrossRef](#)]